Technical Physics, Vol. 46, No. 2, 2001, pp. 133–138. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 1–7.

The Contributions of Polarizabilities to Four Basis Polarizations of Electromagnetic Media

E. N. Bukina and V. M. Dubovik

Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

e-mail: bukina@thsun1.jinr.ru

Received February 8, 2000

Abstract—The contributions to four basis densities of polarization distribution for an arbitrary electromagnetic medium are determined. Mixed polarizabilities given by tensors of up to the fourth rank are considered. Specific physical examples are discussed. *© 2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Electromagnetic properties of some substances (e.g., crystals [1]) are described predominantly by the distributions of dipoles. Multipole expansion of the density of such a distribution differs from expansion of the current of charges and leads to unusual interaction of dipole systems with external fields. Comprehensive analysis of such expansions is not only of academic interest. It is known [2, 3] that multipole expansion of current (charge) density within the framework of the Maxwell–Lorentz electrodynamics involves three types of basis dipoles: electric **Q**, magnetic **M**, and (polar) toroidal **T**. An electromagnetic medium without free and bound charges ($\rho \equiv 0$), as well as unclosed currents ($\mathrm{div}\mathbf{j} = 0$), is described only by magnetic and toroidal polarizations [4]. When describing the distribution of magnetization in a nucleus, Blatt and Weisskopf (see [5], where they cite earlier works) were apparently the first to introduce magnetic contributions to toroidal polarizations by the name "induced electric transverse moments." In classifying point groups of magnetic symmetry of crystals, several authors [6–9] have pointed to the fact that a toroidal dipole exhibits symmetry of a special type. Magnetic substances allowing toroidal ordering have been referred to as toroics. The references cited give examples of such magnetic crystals. Recent experimental studies of toroidal polarization have been made on magnetic piezoelectric $Ga_{2-x}Fe_xO_3$ [10] and magnetoelectric Cr_2O_3 [11].

In 1986, a family of axial toroidal moments was included in the apparatus of the electrodynamics of continuous media [12].[1] A short time later, these moments found prominent application in molecular physics. In 1991, it was demonstrated [13] that crystallization of some aromatic compounds (e.g., anthracene, phenanthrene, or pentacene) may produce an axial toroidal microdipole by ordering electric dipoles of atoms of a cyclic fragment involved in a complex molecule of a benzene ring C_6 (Fig. 1). This has made it possible to give insight into the interaction between microcrystals of aromatic compounds and an alternating magnetic field [14]. Such an extraordinary type of electromagnetic induction is known as aromagnetism.

These advances, as well as the inadequacy of the theory of the Aharonov–Bohm effect [15] and the discovery of similar effects (see, for example, [16]), necessitate a more systematic approach to multipole representation of various configurations that can arise, for example, in dipole media and to selecting parameters for the description of electromagnetic properties of quantum objects like complicated molecules.

In this work, we found complete sets of basis dipoles for dipole media of most physical importance [9]. Based on the mathematical fact of their existence, we demonstrate possible unusual responses of condensed matter to uniform and/or dynamic electric and magnetic fields. This paper is an extension of [12, 17].

CONSTRUCTION OF MULTIPOLE MOMENTS FOR SPIN (DIPOLE) MEDIA AND THEIR INTERACTION WITH EXTERNAL FIELDS

Let us consider a system of continuously distributed elementary electric dipoles that can be described by a function $\mathbf{d} : \mathbf{r} \in V$ limited in space and, in the general case, time-dependent. Assume that internal interactions in this system or applied external fields lead to the formation and ordering of variously configured aggregates, each consisting of a finite number of elementary dipoles. We will describe the properties of the aggregates by multipole parameters. The first three terms in the expansion of the energy of their interaction with external fields are

$$W = -\int_V \mathbf{d}(\mathbf{r})\mathbf{E}(\mathbf{r})d^3r = \frac{1}{6}\bar{r_0^2}\mathrm{div}\mathbf{E} - \mathbf{Q}\mathbf{E} - \mathbf{T}^{(d)}\mathrm{curl}\mathbf{E}$$

[1] Historically [8], the basis dipole of this family is sometimes designated as *G*.

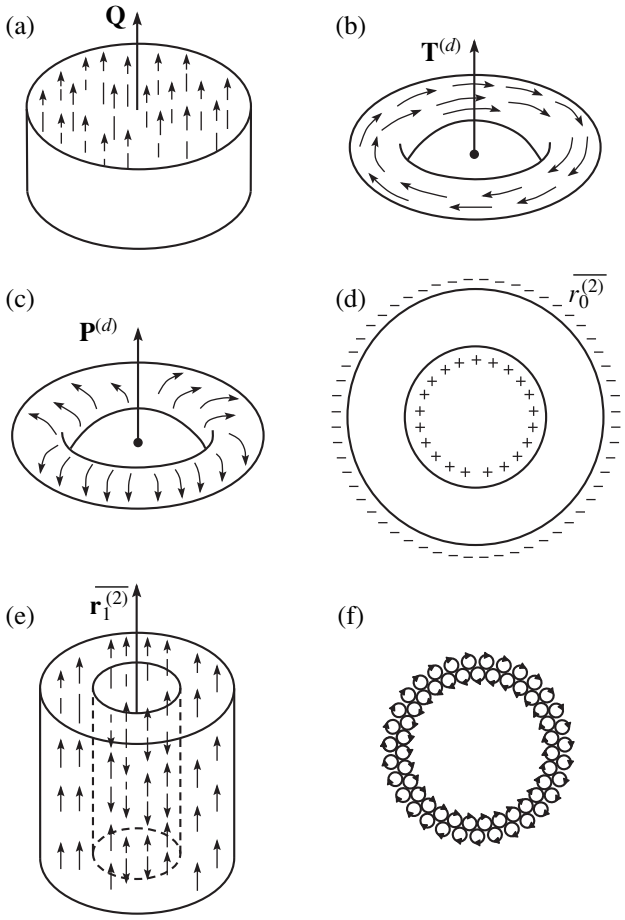

Fig. 1. (a) Model in which electric microdipoles add up to total dipole moment **Q**; (b) model of a dipole axial toroidal moment $\mathbf{T}^{(d)}$ realized in a toroid-shaped dielectric when electric microdipoles are oriented along the azimuthal directions; (c) model of a dipole poloidal moment $\mathbf{P}^{(d)}$ realized in a toroid-shaped dielectric when electric microdipoles are oriented along the meridional directions; (d) centrosymmetric polarization of a spherical dielectric layer having the symmetry of a spherical capacitor; scalar moment $\overline{r_0^{(2)}}$ is the main multipole characteristic of this object; (e) model of the longitudinal rms radius of the electric dipole moment $\overline{\mathbf{r}_1^{(2)}}$ (double cylindrical layer); (f) equatorial section of a thin-walled torus with magnetic microdipoles on the meridians; poloidal dipole moment $\mathbf{P}^{(\mu)}$ is the multipole characteristic of this object upon its contraction to a point.

$$-\mathbf{P}^{(d)}\mathrm{curl}\,\mathrm{curl}\mathbf{E}-\frac{1}{2}Q_{ij}(\nabla_i E_j+\nabla_j E_i)-T_{ij}^{(d)}\nabla_i(\mathrm{curl}\mathbf{E})_j$$
$$+\frac{1}{10}\overline{\mathbf{r}_d^{(2)}}\nabla\mathrm{div}\mathbf{E}+\ldots\longrightarrow-\mathbf{Q}\mathbf{E}-\mathbf{T}^{(d)}\mathrm{curl}\mathbf{E} \qquad (1)$$
$$+\mathbf{P}^{(d)}(\nabla\rho-\Delta\mathbf{E})+\ldots\longrightarrow-\mathbf{Q}\mathbf{E}+\mathbf{T}^{(d)}\dot{\mathbf{H}}$$
$$+\mathbf{P}^{(d)}(\dot{\mathbf{j}}-\ddot{\mathbf{E}})+\ldots,$$

where $\mathbf{Q}=\int\mathbf{d}(\mathbf{r})d^3r$ is the conventional total dipole electric moment of the system, $\mathbf{T}^{(d)}=(1/2)\int(\mathbf{r}\times\mathbf{d})d^3r$ is the dipole axial toroidal moment, and $\mathbf{P}^{(d)}=(1/10)\int[\mathbf{r}(\mathbf{r}\mathbf{d})-2r^2\mathbf{d}]d^3r$ is the dipole polar poloidal moment (see Figs. 1a–1c).

These three vectors form a basis allowing one to represent the density of distribution of a given type. In addition, (1) contains $\overline{\mathbf{r}_0^2}=\int\mathbf{r}\mathbf{E}d^3r$, a scalar moment corresponding to the time derivative of the rms charge radius in the expansion of the current (i.e., to the double spherical layer, see Fig. 1d);[2] $\overline{\mathbf{r}_0^2}=\int[2\mathbf{r}(\mathbf{r}\mathbf{E})+r^2\mathbf{E}]d^3r$, the rms radius of the longitudinal component of the dipole moment **Q**, generated by the double cylindrical layer of the dipoles (see Fig. 1e); and Q_{ij} and $T_{ij}^{(d)}$, the electric and toroidal quadrupole moments determined by conventional rules. In (1), all the interactions of the multipole characteristics with the field or its derivatives are local and written for the "center" of the system.

For the interaction energy between magnetic dipoles and an external nonuniform magnetic field, **μH** is substituted for **dE** and we have

$$W=-\int\boldsymbol{\mu}(\mathbf{r})\mathbf{H}(\mathbf{r})d^3r=-\mathbf{M}\mathbf{H}-\mathbf{T}^{(\mu)}\mathrm{curl}\mathbf{H}$$
$$-\mathbf{P}^{(\mu)}\mathrm{curl}\,\mathrm{curl}\mathbf{H}-\frac{1}{2}M_{ij}(\nabla_i H_j+\nabla_j H_i)$$
$$-T_{ij}^{(\mu)}\nabla_i(\mathrm{curl}H)_j+\ldots\longrightarrow-\mathbf{M}\mathbf{H}-\mathbf{T}^{(\mu)}\dot{\mathbf{E}} \qquad (2)$$
$$-\mathbf{P}^{(\mu)}(\mathrm{curl}\,\mathbf{j}-\ddot{\mathbf{H}})\longrightarrow-\mathbf{M}\mathbf{H}-\mathbf{T}^{(\mu)}\mathrm{curl}\mathbf{H}$$
$$+\mathbf{P}^{(\mu)}\Delta\mathbf{H}^{\mathrm{free}}+\ldots.$$

It is seen that, according to electromagnetism duality, the dipole moment **Q** in Eq. (1) is replaced by the dipole magnetic moment **M** (characterizing an equivalent current loop); $\mathbf{T}^{(d)}$, by the polar toroidal dipole $\mathbf{T}^{(\mu)}$; and the vector poloidal moment $\mathbf{P}^{(d)}$, by the dipole axial poloidal moment $\mathbf{P}^{(\mu)}$. If the magnetic moments are carried by spin particles (e.g., neutrons), then the multipole expansion of the distribution of their magnetic moments (in a nucleus or a neutron star) acquires the primary sense. Zheludev [6] was the first to predict toroidal ordering of spin magnetic moments in perovskites within the theory of condensed matter. For ultrathin films, ordering of magnetic particles into closed configurations that can be described by toroidal moments seems to be a rule rather than an exception [18, 19]. Note that exchange interactions between coplanar spins give rise to the Kosterlitz–Thouless lattice of "antitoroidally" arranged vortexes.

It is evident that expansions like (1) and (2) can be obtained for the remaining two fundamental (toroidal)

[2] In the specific case of zero inner radius, such a configuration is sometimes called "hedgehog."

dipole distributions $\boldsymbol{\tau}(\mathbf{r})$ and $\mathbf{g}(\mathbf{r})$. It is unlikely that the resulting high-order vortexes are realized in crystals. They may form in exotic media where the elementary carriers initially possess, for example, toroidal dipole moments. Recent experimental data indicate that neutrinos are apparently Majorano particles forming dense media (10^{15} cm^{-3}). It is known [20, 21] that neutrinos have only toroidal dipole moments. That coherent excitations in such media, in particular, hadron gravitational excitations, are a possibility was first indicated in [22].

VECTOR PARAMETERS AND NONLINEAR MEDIA

The above discussion is related to so-called linear electrodynamics. In its terms, elementary objects (e.g., dipoles) are controlled by a field or a current of a given type upon contact interaction so that the parameters of the objects remain unchanged. One can also consider more complex objects consisting of dipoles: quadrupoles, octupoles, etc. An example of associated macroscopic constructions is two or three current loops that are closely spaced or form a single circuit. Consider a wire ring shaped like a propeller. In each of the loops, the electric current generates a magnetic field and the magnetic properties can approximately be described by magnetic moments. The energy of local interactions between the two loops and an external magnetic field of arbitrary configuration is then given by

$$W_L = \mu_i H_i\{1\} + \mu_k H_k\{2\}. \quad (3)$$

Here, "locality" means that both terms in the above formula are written for the centers of the loops. In addition to the vectors μ_i and μ_k, which relate to the different loops and are specified at different points {1} and {2} in the space, we introduce a quadrupole moment of the system as a symmetric tensor μ_{ik} of the second rank. As in the previous section, the parameter μ_{ik} takes into account the contribution to the energy from the gradient of the external magnetic field (i.e., due to field nonuniformity between the centers of the loops). Then Eq. (3) can be approximately recast as

$$W_L = \mu_i H_i\{0\} + \mu_k H_k\{0\} + \mu_{ik}\nabla_i H_k\{0\}. \quad (4)$$

Now all the terms are written for some middle point (hereafter, the center of the system). Such a representation reflects the essence of multipole expansion.

Assume that the loops are made of a flexible wire. Then, the magnetic field orienting the magnetic moments can deform the loops. We take this perturbation into account by introducing the quadrupole polarizability $\alpha^{(m)}$ of the medium. The additional energy of local interaction of the system with a magnetic field is given by the expression

$$W_{NL} = \alpha^m_{ik} H_i H_k\{0\}, \quad (5)$$

which is nonlinear in field in this case. Eventually, we come to a qualitatively new property of the system of dipoles. It may be concluded that the external field H_i induces an additional magnetic dipole polarization of the medium $M_k = \alpha^{(m)}_{ik} H_i$, which can interact with the field H_k. Here, $\alpha^{(m)}_{ik}$ is the quadratic magnetic susceptibility of the medium.

Thus, we introduced a nonlinear interaction accounting for the nonlinear response of the medium to a magnetic field. Basically, the tensor α may have an arbitrary number of subscripts. In the expression for energy, it must be convolved with the corresponding symmetric tensor for the vector of the external magnetic field strength. A system consisting of electric dipoles may naturally experience deformations similar to those described above. If the electric susceptibility α^e is substituted for the magnetic one, the expression for the energy of interaction with an external electric field will be completely identical to Eq. (5). In real systems, atoms and molecules are in continuous motion due to external and internal factors. Therefore, interrelation between the sources via the fields may give rise to mixed polarizabilities. For example, in the case of electromagnetic polarizability, the interaction energy is given by

$$W_{NL} = \beta^{(em)}_{ijk} E_i H_j H_k. \quad (6)$$

Bearing in mind the organization of computer memory, we introduce the polarizability of the third rank, corresponding to the Hall effect symmetry, and the respective interaction energy

$$W_{NL} = \frac{1}{c}\alpha^{(em)}_{ikl} E_i H_k J_l, \quad (7)$$

where J_l is the external current passing through the system.

In particular, this current may be a displacement current when the system is inside a capacitor. Assume that the tensor $\alpha^{(em)}_{ikl}$ is asymmetric with respect to subscripts i and k. Then, for a given experimental configuration, the asymmetric part of the tensor $E_i H_k$ can be related to the Poynting vector of light passing through a crystal with appropriate nonlinear properties. Following the preceding considerations, we may introduce the induced toroidal polarization

$$T_l = \frac{1}{c}\alpha^{(em)}_{[ik]l} E_i H_k \equiv \alpha^{(em)}_{jl} S_j, \quad (8)$$

responding to the current J_l passing through the crystal.

Indeed, an electromagnetic wave may reorient the polarization of a ferromagnetic dielectric, turning it into a weak ferromagnet with a toroidal polarization. Such behavior is typical of perovskites, manganites [23], and other complex compounds.

In experiments, approaches to measuring the toroidal polarization are simpler. For example, Popov *et al*. [10] studied the response of Fe–Ga oxide single crystals to a magnetic field. It is known that such materials combine the ferromagnetic and piezoelectric properties (by analogy with elastic loops). The authors of [10] measured the electric polarization induced in the crystal by a magnetic field **H**:

$$P_i = \alpha_{ij}^{(em)} H_j, \tag{9}$$

where $\alpha_{ij}^{(em)}$ is the tensor of electromagnetic susceptibility.

The asymmetric part of the tensor $\alpha_{ij}^{(em)}$ defines the dipole toroidal moment in accordance with the duality relation

$$\alpha_{ij}^{(em)} - \alpha_{ji}^{(em)} = \epsilon_{ijk} T_k, \tag{10}$$

where ϵ_{ijk} is the Levi-Civita symmetric unit tensor.

The polarization per unit volume of the oxide gives the dipole toroidal moment $|\mathbf{T}| = 24\ \mu_B$ Å. Similar results were obtained for Cr_2O_3 magnetoelectric subjected to a strong magnetic field (exceeding the field of spin flop transition).

CONTRIBUTIONS OF THE POLARIZABILITIES TO THE VECTOR POLARIZATIONS AND THEIR MULTIPOLE EXPANSIONS

In this section, we develop a comprehensive formalism of contributions of the highest (mixed) polarizabilities to four basis densities of polarization distribution for an arbitrary medium: **P**, **M**, $\mathbf{T}^{(\mu)}$, and $\mathbf{T}^{(d)}$. Further multipole expansion of the energy of their interaction with external polarizations would generalize the above formulas for the case of nonlinear media. Based on the fundamental interactions of each of the polarizations, we introduce polarizabilities that take into account all possible mechanisms whereby external fields and currents induce a given polarization. The most general expression for electric polarization up to contributions of the fourth order is given by

$$\begin{aligned}
\mathbf{P} &= \alpha_2^{(e)} \cdot E + \alpha_3^{(e)} : EE + \alpha_4^{(3)} \vdots EEE \\
&+ \ldots + \alpha_3^{(em)} : EH + \alpha_4^{(em)} \vdots EEH + \ldots + \beta_2^{(m)} \cdot H \\
&+ \beta_3^{(m)} : HH + \beta_4^{(m)} \vdots HHH + \ldots + \beta_3^{(me)} : HE \\
&+ \beta_4^{(me)} \vdots HHE + \ldots + \gamma_2^{(\tau)} \cdot J^{(m)} + \gamma_3^{(\tau)} : J^m J^m \\
&+ \gamma_4^{(\tau)} \vdots J^m J^m J^m + \ldots + \gamma_3^{(e\tau)} : EJ^m + \gamma_3^{(m\tau)} : HJ^m \\
&+ \ldots + \gamma_4^{(em\tau)} \vdots EHJ^m + \ldots + \delta_2^{(\tau)} \cdot J^e + \delta_3^{(\tau)} : J^e J^e \\
&+ \delta_4^{(\tau)} \vdots J^e J^e J^e + \ldots + \delta_3^{(e\tau)} : EJ^e \\
&+ \delta_3^{(m\tau)} : HJ^e + \ldots + \delta_4^{(em\tau)} \vdots EHJ^e + \ldots .
\end{aligned} \tag{11}$$

Here, dots denote the number of contracted subscripts in the corresponding tensor, $\mathbf{J}^m := 1/c(\mathbf{j} + \dot{\mathbf{D}}) \equiv \mathrm{curl}\mathbf{H}$, $\mathbf{J}^e := \mathrm{curl}\mathbf{E}$, and subscripts at the polarizabilities denote their tensor rank. One can pass to irreducible tensors by introducing half-sums and half-differences for $\alpha_3^{(em)}$ and $\beta_3^{(me)}$, etc. For magnetic polarization, we obtain

$$\begin{aligned}
\mathbf{M} &= \hat{\alpha}_2^{(m)} \cdot H + \hat{\alpha}_3^{(m)} : HH + \hat{\alpha}_4^{(m)} \vdots HHH + \ldots \\
&+ \hat{\alpha}_3^{(em)} : HE + \hat{\alpha}_4^{(em)} \vdots HEE + \ldots + \hat{\beta}_2^{(e)} \cdot E \\
&+ \hat{\beta}_3^{(e)} : EE + \hat{\beta}_4^{(e)} \vdots EEE + \ldots + \hat{\beta}_3^{(me)} : EH \\
&+ \hat{\beta}_4^{(me)} \vdots EEH + \ldots + \hat{\gamma}_2^{(\tau)} \cdot J^{(m)} + \hat{\gamma}_3^{(\tau)} : J^m J^m \\
&+ \hat{\gamma}_4^{(\tau)} \vdots J^m J^m J^m + \ldots + \hat{\gamma}_3^{(e\tau)} : HJ^m + \hat{\gamma}_3^{(m\tau)} : EJ^m \\
&+ \ldots + \hat{\gamma}_4^{(me\tau)} \vdots HEJ^m + \ldots + \hat{\delta}_3^{(\tau)} J^e + \hat{\delta}_3^{(\tau)} : J^e J^e \\
&+ \hat{\delta}_4^{(\tau)} \vdots J^e J^e J^e + \ldots + \hat{\delta}_3^{(e\tau)} : EJ^e \\
&+ \hat{\delta}_3^{(m\tau)} : HJ^e + \ldots + \hat{\delta}_4^{(em\tau)} \vdots EHJ^e + \ldots;
\end{aligned} \tag{12}$$

for toroidal polarization of magnetic nature,

$$\begin{aligned}
\mathbf{T}_M &= \bar{\alpha}_2^{(\tau)} \cdot J^m + \bar{\alpha}_3^{(\tau)} : J^m J^m + \bar{\alpha}_4^{(\tau)} \vdots J^m J^m J^m + \ldots \\
&+ \bar{\alpha}_3^{(\tau e)} : J^m E + \bar{\alpha}_4^{(\tau e)} \vdots J^m EE + \ldots + \bar{\beta}_2^{(e)} \cdot E \\
&+ \bar{\beta}_3^{(e)} : EE + \bar{\beta}_4^{(e)} \vdots EEE + \ldots + \bar{\beta}_3^{(e\tau)} : EJ^m \\
&+ \bar{\beta}_4^{(e\tau)} \vdots EEJ^m + \ldots + \bar{\gamma}_2^{(m)} \cdot H + \bar{\gamma}_3^{(m)} : HH \\
&+ \bar{\gamma}_4^{(m)} \vdots HHH + \ldots + \bar{\gamma}_3^{(\tau m)} : J^m H + \bar{\gamma}_3^{(\tau e)} : J^m E \\
&+ \ldots + \bar{\gamma}_4^{(\tau me)} \vdots J^m HE + \ldots + \bar{\delta}_2^{(\tau)} \cdot J^e + \bar{\delta}_3^{(\tau)} : J^e J^e \\
&+ \bar{\delta}_4^{(\tau)} \vdots J^e J^e J^e + \ldots + \bar{\delta}_3^{(e\tau)} : EJ^e \\
&+ \bar{\delta}_3^{(m\tau)} : HJ^e + \ldots + \bar{\delta}_4^{(me\tau)} \vdots EHJ^e + \ldots;
\end{aligned} \tag{13}$$

and for toroidal polarization of electric nature,

$$\begin{aligned}
\mathbf{T}_p &= \tilde{\alpha}_2^{(\tau)} \cdot J^e + \tilde{\alpha}_3^{(\tau)} : J^e J^e + \tilde{\alpha}_4^{(\tau)} \vdots J^e J^e J^e + \ldots \\
&+ \tilde{\alpha}_3^{\tau e} : J^e E + \tilde{\alpha}_4^{(\tau e)} \vdots J^e EE + \ldots + \tilde{\beta}_2^{(e)} \cdot E \\
&+ \tilde{\beta}_3^{(e)} : EE + \tilde{\beta}_4^{(e)} \vdots EEE + \ldots + \tilde{\beta}_3^{(e\tau)} : EJ^e \\
&+ \tilde{\beta}_4^{(e\tau)} \vdots EEJ^e + \ldots + \tilde{\gamma}_2^{(m)} \cdot H + \tilde{\gamma}_3^{(m)} : HH \\
&+ \tilde{\gamma}_4^{(m)} \vdots HHH + \ldots + \tilde{\gamma}_3^{(\tau m)} : J^e H + \tilde{\gamma}_3^{(\tau e)} : J^e E \\
&+ \ldots + \tilde{\gamma}_4^{(\tau me)} \vdots J^e HE + \ldots + \tilde{\delta}_2^{(\tau)} \cdot J^m + \tilde{\delta}_3^{(\tau)} : J^m J^m \\
&+ \tilde{\delta}_4^{(\tau)} \vdots J^m J^m J^m + \ldots + \tilde{\delta}_3^{(e\tau)} : EJ^m \\
&+ \tilde{\delta}_3^{(m\tau)} : HJ^m + \ldots + \tilde{\delta}_4^{(me\tau)} \vdots EHJ^m + \ldots .
\end{aligned} \tag{14}$$

With allowance for dual symmetries, $\beta_2^{(m)} = \hat{\beta}_2^{(e)}$, $\bar{\delta}_2^{(\tau)} = \tilde{\delta}_2^{(\tau)}$, etc. All the contributions to the polarizations containing vectors J^m and J^e are called nonlocal. In several recent experiments, they were separated from local effects [24]. In these formulas, some of the field strength vectors can be replaced by their derivatives. The rank of the polarizabilities will increase in this case, and these contributions to the corresponding polarizations will become nonlocal. A set of nonlinear cross terms may include not only magnetoelectric ones, but also mechanomagnetic, thermoelectric, and others.

Let us proceed to multipole expansion of the vector polarizations with regard for the properties of the polarizabilities. For this purpose, we use formulas (1) and (2) for the interaction energy between these polarizations and external fields. We start from the analysis of electric polarization. The presence of the electric field vector in the associated expression modifies the definition of the multipole moments. The dipole moment $\mathbf{Q} = \int \mathbf{P}(\mathbf{r}) d^3r$ acquires additional terms due to the polarizabilities $\alpha_2^{(e)}$, $\alpha_3^{(e)}$, …, $\gamma_4^{(em\tau)}$, … . For example, the contribution related to $\alpha_3^{(e)}$ is

$$\mathbf{Q}_3^{(e)} = \alpha_3^{(e)} : \int EEd^3r. \tag{15}$$

Such a contribution can be considerable if the medium exhibits strong nonlinear properties and $\alpha_3^{(e)}$ is large. This addition also depends on the integral. In spherical coordinates, it can be rewritten as

$$\int E_i E_k d^3r = \int E^2 r^2 dr \int n_i n_k d\Omega, \tag{16}$$

where $\mathbf{n} = \mathbf{E}/|\mathbf{E}|$.

Such integrals are common in electrodynamics and statistical mechanics. They are found from the general formula

$$\int \underbrace{n_i n_j n_k \ldots n_l}_{L} = \frac{1}{L!}\{\delta_{ij}\delta_k \ldots \delta_{\ldots l} - \delta_{ik}\delta_j \ldots \delta_{\ldots l} + \ldots\}. \tag{17}$$

These integrals become difficult to take if div$\mathbf{E} = 0$ or div$\mathbf{H} = 0$. Fields applied to crystals in physical experiment may be of different natures (in particular, of different frequencies). Specifically, the Pockels effect, usually considered as linear, is actually a second-order effect, since it involves the simultaneous action of a low-frequency electric field and a light wave where the electric component usually prevails. The well-known Kerr effect is believed to be quadratic, because the refractive index of a crystal is proportional to the square of the low-frequency refractive field. However, judging from the expression for interaction energy, this effect is generally of the third order. One should also distinguish between static and dynamic effects. In the latter, one must take into account sum and difference frequencies, which may strongly influence the amount of the effect.

Using expression (11), one can easily derive a formula for **Q** where all the nonlinear local contributions are included. If in (11)–(14) we capitalize the letters denoting the nonlocal polarizabilities, the contribution, for example, from the nonlocal polarizability of the second rank to the dipole moment will be represented as

$$\mathbf{Q}_3^{(e)} = A_3^{(e)} : \int \nabla E d^3r. \tag{18}$$

The tensor in the integrand can be separated into irreducible parts. Then its trace, equal to div**E**, will be the external charge density distribution. The symmetric and antisymmetric parts are responsible for interaction with the quadrupole and axial components of an external field, respectively.

CONCLUSION

A comprehensive formalism of such multipole expansions is easy to develop. The parametrization of the expression for the interaction energy between a polarization and an external field will lead to a cubic lattice infinitely occupying the space, for example, in the positive directions of the *X*, *Y*, and *Z* axes. On the *X* (principal) axis, we will plot the values of the moments describing a complex electromagnetic system in order of ascending *l* polarity. On the *Y* axis, we will plot the contributions from the radii (to the power 2*n*) of the distributions for each of the moments in order of ascending power *n*. Recall that multipole analysis is mathematically based on the theory of generalized functions. Therefore, in spite of the erroneous views regarding the domination of the moments, the values of the radii may appear to be much more important for the description of a particular system. Specifically, an ellipsoidal double layer with a small eccentricity is more appropriately described by the rms charge radius than by the electric quadrupole moment. From the geometrical point of view, the radii describe the deviation of each of the *l*-pole moments from its value specified at the point with actually zero dimensions [25]. As follows from [25], both moments and radii could be corrected for nonlinearity in the interaction of a given system with an external field. The corrections can be plotted vertically in parallel to the *Z* axis (in order of ascending number of subscripts in the polarizability tensors) over each position of the matrix fixing the values of the moments or radii.

Note that the majority of experiments employ either static or harmonic fields. It should be noted that, to reveal first-order vortex structures (toroidal dipoles), the "true" response function (i.e., free of other multipole interactions) is obtained by applying fields linearly increasing with time. The contributions of the poloidal dipoles (second-order vortex structures or two-tori [26]) can be revealed by applying a magnetic field that

grows quadratically with time. In general, the power in the time dependence of the field must equal the order of an *n*-torus. It is natural that the response amplitude will decrease sharply with increasing *n* owing to the relativistic factor $(1/c)^n$ in the expression for the energy of interaction of an *n*-torus with an external field.

Note that the separation of the vector contributions is most convenient in analyzing, for example, small-size crystal objects. The advantages of the method become even more obvious in studying the electromagnetic responses from low-dimension structures used in nanotechnology.

ACKNOWLEDGMENTS

This work was supported by the Russian Interdisciplinary Scientific and Technical Program "Fundamental Metrology," project no. 2.51.

REFERENCES

1. Yu. N. Venevtsev, V. V. Gagulin, and V. N. Lyubimov, *Ferroelectric Magnets* (Nauka, Moscow, 1982).
2. V. M. Dubovik and A. A. Cheshkov, Fiz. Élem. Chastits At. Yadra **5**, 791 (1974) [Sov. J. Part. Nucl. **5**, 318 (1974)].
3. V. M. Dubovik and L. A. Tosunyan, Fiz. Élem. Chastits At. Yadra **14**, 1193 (1983) [Sov. J. Part. Nucl. **14**, 504 (1983)].
4. V. M. Dubovik, in *Proceedings of the 3rd Seminar "Group-Theoretical Methods in Physics* (Nauka, Moscow, 1986), Vol. 2, pp. 356–362.
5. J. M. Blatt and V. F. Weisskopf, *Theoretical Nuclear Physics* (Wiley, New York, 1952; Inostrannaya Literatura, Moscow, 1954).
6. I. S. Zheludev, Izv. Akad. Nauk SSSR, Ser. Fiz. **33**, 204 (1969).
7. H. Schmid, Int. J. Magn. **4**, 337 (1974).
8. E. Ascher, in *Magnetoelectric Interaction Phenomena in Crystals*, Ed. by A. Freeman and H. Schmid (Gordon and Breach, New York, 1975), p. 69.
9. V. M. Dubovik, S. S. Krotov, and V. V. Tugushev, Kristallografiya **32**, 540 (1987) [Sov. Phys. Crystallogr. **32**, 314 (1987)].
10. Yu. F. Popov, A. K. Zvezdin, A. M. Kadomtseva, *et al.*, Zh. Éksp. Teor. Fiz. **114**, 263 (1998) [JETP **87**, 146 (1998)].
11. Yu. F. Popov, A. M. Kadomtseva, A. K. Zvezdin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69** (4), 302 (1999) [JETP Lett. **69**, 330 (1999)].
12. V. M. Dubovik, L. A. Tosunyan, and V. V. Tugushev, Zh. Éksp. Teor. Fiz. **90**, 590 (1986) [Sov. Phys. JETP **63**, 344 (1986)].
13. M. A. Martsenyuk and N. M. Martsenyuk, Pis'ma Zh. Éksp. Teor. Fiz. **53** (5), 229 (1991) [JETP Lett. **53**, 243 (1991)].
14. N. A. Tolstoĭ and A. A. Spartakov, Pis'ma Zh. Éksp. Teor. Fiz. **52** (3), 796 (1990) [JETP Lett. **52**, 161 (1990)].
15. G. N. Afanas'ev, Fiz. Élem. Chastits At. Yadra **21**, 172 (1990) [Sov. J. Part. Nucl. **21**, 74 (1990)].
16. X. G. He and B. H. J. McKellar, Phys. Rev. A **47**, 3424 (1993); G. N. Afanas'ev, Fiz. Élem. Chastits At. Yadra **24**, 512 (1993) [Phys. Part. Nucl. **24**, 219 (1993)].
17. V. M. Dubovik, M. A. Martsenyuk, and N. M. Martsenyuk, Fiz. Élem. Chastits At. Yadra **24**, 1056 (1993) [Phys. Part. Nucl. **24**, 453 (1993)].
18. B. N. Filippov, L. G. Korzunin, *et al.*, Fiz. Met. Metalloved. **87** (6), 17 (1999).
19. N. A. Usov, J. Magn. Magn. Mater. **203**, 277 (1999).
20. V. M. Dubovik and V. E. Kuznetsov, J. Mod. Phys. A **143**, 5257 (1998).
21. E. N. Bukina, V. M. Dubovik, and V. E. Kuznetsov, Phys. Lett. B **435**, 134 (1998).
22. V. M. Koryukin, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 10, 119 (1996).
23. L. M. Sandratskii and J. Kübler, Phys. Rev. Lett. **76**, 4963 (1996).
24. V. P. Drachev, S. V. Perminov, S. G. Rautian, and V. P. Safonov, Pis'ma Zh. Éksp. Teor. Fiz. **68** (8), 618 (1998) [JETP Lett. **68**, 651 (1998)].
25. V. M. Dubovik and V. V. Tugushev, Phys. Rep. **187** (4), 145 (1990).
26. G. N. Afanasiev and V. M. Dubovik, Fiz. Élem. Chastits At. Yadra **29**, 891 (1998) [Phys. Part. Nucl. **29**, 366 (1998)].

Translated by A. Chikishev

Technical Physics, Vol. 46, No. 2, 2001, pp. 139–146. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 8–15.

On the Theory of Classic Mesodiffusion

V. V. Uchaikin and V. V. Saenko

Ul'yanovsk State University, Institute for Theoretical Physics, Ul'yanovsk, 432700 Russia

e-mail: uchaikin@sv.uven.ru

Received March 27, 2000

Abstract—A simple model of the classical random walk of particles with a constant speed and anisotropic angular distribution is used to study the characteristic features of mesodiffusion, that is, of an intermediate stage between the ballistic regime (short times) and ordinary diffusion (long times). In the extreme case of anisotropy, namely, walking along a straight line, the process can be described by the telegraph equation, whose solution contains δ-functions accounting for the ballistic component. As the anisotropy becomes less pronounced, the δ-singularity transforms into a frontal burst (the quasi-ballistic component), beyond which the distribution can be satisfactorily described by the telegraph approximation. In the other extreme case of isotropic walking, the frontal burst disappears and the telegraph approximation, contrary to general belief, proves to be cruder than the diffusion approximation. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Research into the preparation of high-purity semiconductor materials with perfect crystal structure gave birth to a new branch in condensed matter physics, the study of mesoscopic systems [1–3]. Mesoscopic nanostructures present a unique possibility to experimentally study the transfer process in a medium with a well-defined potential field not perturbed by random impurities and other defects. The existence of three temporal intervals is essential to this process in which the transfer mechanisms are different: an interval $(0, t_1)$ where the ballistic transfer is dominant; an interval of (t_2, ∞); an interval $(t_2 > t_1)$ of ordinary (gaussian) diffusion, and an intermediate interval (t_1, t_2) of a transfer mechanism called mesoscopic diffusion [4], which we will call mesodiffusion.

The quantum-mechanical analysis of the one dimensional problem carried out in [4] has shown that the distinctive feature of mesodiffusion is a deviation from Fick's law

$$j = -D\frac{\partial \rho}{\partial x},$$

which is replaced in this region by the Maxwell–Cattaneo relationship

$$j = -D\frac{\partial \rho}{\partial x} - \Theta\frac{\partial j}{\partial t}, \quad \Theta > 0.$$

In combination with the continuity equation

$$\frac{\partial \rho}{\partial t} = -\frac{\partial j}{\partial x},$$

it gives an equation for the distribution density $\rho(x, t)$ of the particles

$$\frac{\partial \rho}{\partial t} + \Theta\frac{\partial^2 \rho}{\partial t^2} = D\frac{\partial^2 \rho}{\partial x^2},$$

which is called the telegraph equation. Its solution for a short-time (for example, instantaneous) source represents a diffusion front, beyond which there are no diffusing particles, and in the vicinity of which they travel in the ballistic regime. At long times, the ballistic component decays and the remaining part of the solution for large samples transforms into a gaussian packet that satisfies an ordinary diffusion equation.

The results obtained in [4] apply, however, to a rather artificial model of a particle walking along a fixed x-axis with zero transverse velocity component and an alternating longitudinal component (the one-dimensional walk). This model does not distinguish between the nonscattered particles and those multiply scattered in one (say, positive) direction of the x-axis, thus eliminating the differences between the telegraph and kinetic equations, on the one hand, and creating the specific δ-singularity at the diffusion wave front that is lacking in the real physical process on the other.

The other extreme case is considered to be the one-velocity model of isotropic scattering, a time-independent version of which was thoroughly studied in connection with its use in neutron physics and nuclear reactor problems [5–7]. The time-dependent version of the model is considered in relatively few papers (see reference in [8–10]) in which the telegraph equation appears as an approximation. However, the quantitative analysis of its accuracy has not still been carried out.

The present study considers the class of the mesodiffusion models relying on a parameter $\nu \in [1/\sqrt{3}, 1]$. It is assumed that the distribution of the cosine of the

angle ω between the *x*-axis and the direction of particle movement after scattering is independent of the direction of movement before scattering and has the density

$$W_\nu(\omega) = \frac{\alpha+1}{2}|\omega|^\alpha, \quad |\omega| \le 1, \quad \alpha = \frac{3\nu^2-1}{1-\nu^2} \ge 0. \quad (1)$$

The initial direction of a particle which starts to move from the origin of the coordinates at the time $t = 0$ has the same density distribution. The parameter

$$\nu = \left[\int_{-1}^{1} \omega^2 W_\nu(\omega)\,d\omega\right]^{1/2} = \left(\frac{\alpha+1}{\alpha+3}\right)^{1/2} \ge \frac{1}{\sqrt{3}}$$

is the root mean square cosine, $\nu = 1/\sqrt{3}$ for isotropic walking and $\nu = 1$ for one-dimensional walking. The smooth dependence of the scattering indicatrix (1) on the parameter ν enables the evolution of distributions with the variation of ν to be followed and a conclusion about the relationship between the solutions of the diffusion, the telegraph, and the kinetic equations, to be made.

THE TELEGRAPH EQUATION AND ONE-DIMENSIONAL MESODIFFUSION

The telegraph equation, derived (as pointed out in [11]) by Lord Kelvin in connection with the laying of the first transatlantic cable, has the form (in dimensionless time units *t*)

$$\frac{\partial^2 f_\nu}{\partial t^2} + \frac{\partial f_\nu}{\partial t} = \nu^2 \frac{\partial^2 f_\nu}{\partial x^2}. \quad (2)$$

Its solution $f_\nu(x, t)$ has the meaning of a current at a point *x* of the conductor at *t*. The parameter ν is related to the self-induction and the resistance of a conductor of unit length. Under the initial conditions

$$f_\nu(x, 0) = \delta(x), \quad [\partial f_\nu(x, t)/\partial t]|_{t=0} = 0,$$

it consists of two terms

$$f_\nu(x, t) = f_\nu^{(0)}(x, t) + f_\nu^{(s)}(x, t). \quad (3)$$

The first term describes two instant pulses moving away from the origin of the coordinates with a velocity ν:

$$f_\nu^{(0)}(x, t) = \frac{1}{2}[\delta(x - \nu t) + \delta(x + \nu t)]\exp\{-t/2\}.$$

The second term gives a continuous part of the solution that fills the interval between these pulses ($-\nu t < x < \nu t$):

$$f_\nu^{(s)}(x, t) = \frac{1}{4\nu}[I_0(\sqrt{(t^2 - x^2/\nu^2)/4}) + tI_1(\sqrt{(t^2 - x^2/\nu^2)/4})/\sqrt{t^2 - x^2/\nu^2}]e^{-t/2}.$$

Note the following properties of this solution:

$$f_\nu(x, t) > 0, \quad |x| \le \nu t,$$
$$f_\nu(x, t) = 0, \quad |x| > \nu t, \quad (4)$$

$$\int_{-\nu t}^{\nu t} f_\nu(x, t)\,dx = 1, \quad (5)$$

$$\int_{-\nu t}^{\nu t} x^2 f_\nu(x, t)\,dx = 2\nu^2[t + \exp\{-t\} - 1], \quad (6)$$

$$f_\nu(x, t) \sim g_\nu(x, t) = \frac{1}{\sqrt{4\pi\nu^2 t}}\exp\left\{-\frac{x^2}{4\nu^2 t}\right\}, \quad (7)$$
$$t \longrightarrow \infty.$$

These turned out to be just the properties needed for the description of a quite different process—symmetrical random walk along a straight line [12–15], in which $f_\nu(x, t)$ has the meaning of the probability density at time *t* and the parameter ν is the velocity of free movement of a particle. The term $f_\nu^{(0)}(x, t)$ describes the distribution of the particles which did not change their direction of movement during time *t*. They are found at points $x = \nu t$ and $x = -\nu t$ and form the front of the diffusion packet which occupies the interval $[-\nu t, \nu t]$; the probability to find a particle outside this interval is equal to zero. At $t \longrightarrow \infty$, the function $f_\nu(x, t)$ tends to the normal distribution $g_\nu(x, t)$ with the variance $2\nu^2 t$ [13] and satisfies the ordinary diffusion equation

$$\frac{\partial g_\nu}{\partial t} = \nu^2 \frac{\partial^2 g_\nu}{\partial x^2}$$

with the initial condition $g_\nu(x, 0) = \delta(x)$. Because of its approximate (asymptotic) nature, the function does not contain information about the diffusion front, but describes only the large central part of the diffusion packet (Fig. 1).

ISOTROPIC MESODIFFUSION

From the physical point of view, however, the one-dimensional model considered in [4] proves to be rather artificial, since, as a result of collisions, the velocity of a real particle may take other directions as well, not just the ones colinear to the *x*-axis. We chose the unit of length in such a way that the absolute value of the particle velocity between collisions (assumed to be constant) would be equal to 1. In the case of one-dimensional walking of such a particle along the *x*-axis, its distribution will be described by the density $f_1(x, t)$. If the particle acquires a new direction independent of the previous one and possessing the same distribution as the initial one after a collision then the density ρ(*x*, *t*) of

the distribution of its x-coordinate at time t satisfies the integral kinetic equation [10, 16]

$$\rho(x, t) = \rho^{(0)}(x, t) + \int_0^t dt' \int_{-t'}^{t'} dx' \rho^{(0)}(x', t') \rho(x - x', t - t'), \quad (8)$$

where $\rho^{(0)}(x, t)$ is the distribution density of the x-coordinate of the nonscattered particle.

From elementary probabilistic considerations, it can be found that this random coordinate is uniformly distributed between $-t$ and t for an isotropic source (we recall that the velocity is equal to 1)

$$\rho^{(0)}(x, t) = \begin{cases} (2t)^{-1} \exp\{-t\}, & |x| < t \\ 0, & |x| > t. \end{cases} \quad (9)$$

Let $\psi(x, \omega, t)$ be the joint distribution density of the x-coordinate and the cosine of the angle between the particle velocity and the x-axis at time t:

$$\int_{-1}^{1} \psi(x, \omega, t) d\omega = \rho(x, t).$$

It is known [16] that the integral equation (8) with kernel (9) is equivalent to the integrodifferential Boltzmann equation

$$\left[\frac{\partial}{\partial t} + \omega \frac{\partial}{\partial x} + 1\right] \psi(x, \omega, t) = \frac{1}{2} \int_{-1}^{1} \psi(x, \omega, t) d\omega \quad (10)$$

with the initial condition

$$\psi(x, \omega, t) = \frac{1}{2} \delta(x).$$

Expansion in terms of Legendre polynomials

$$\psi(x, \omega, t) = \sum_{l=0}^{\infty} (l + 1/2) P_l(\omega) \psi_l(x, t),$$

$$\psi_l(x, t) = \int_{-1}^{1} P_l(\omega) \psi(x, \omega, t) d\omega, \quad \psi_0(x, t) \equiv \rho(x, t)$$

transforms Eq. (10) to a chain of equations

$$\frac{\partial \psi_l}{\partial t} + \frac{1}{2l + 1} \left[(l + 1) \frac{\partial \psi_{l+1}}{\partial x} + l \frac{\partial \psi_{l-1}}{\partial x}\right] + (1 - \delta_{l0}) \psi_l = 0.$$

Retaining the first $L + 1$ equations and discarding in the last of them the term $\partial \psi_{L+1}/\partial x$, we obtain a P_L approximation that is well known in neutron physics [7–9]. In particular, in the P_1 approximation,

$$\frac{\partial \psi_0}{\partial t} + \frac{\partial \psi_1}{\partial x} = 0, \quad \frac{\partial \psi_1}{\partial t} + \frac{1}{3} \frac{\partial \psi_0}{\partial x} + \psi_1 = 0.$$

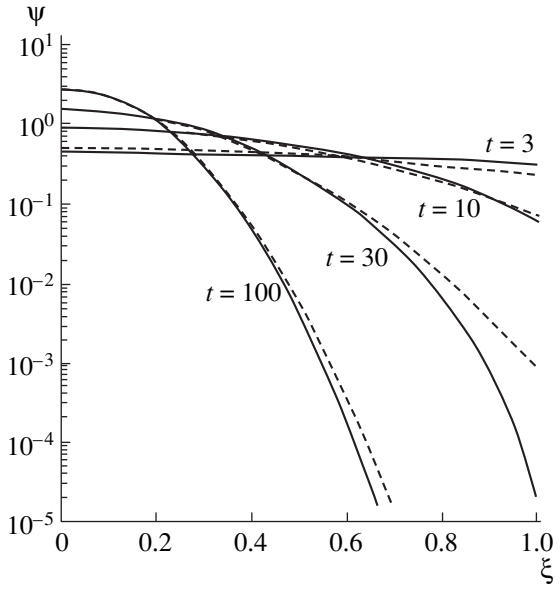

Fig. 1. The distributions $\psi(\xi, t) = tf_1(\xi t, t)$ (solid curve) and $\psi'(\xi, t) = tg_1(\xi t, t)$ (dashed curve) for one-dimensional random walking.

Eliminating ψ_1 from this equation, we obtain an equation for $\psi_0 \equiv \rho$ in the form

$$\frac{\partial^2 \rho}{\partial t^2} + \frac{\partial \rho}{\partial t} = \frac{1}{3} \frac{\partial^2 \rho}{\partial x^2},$$

so that

$$\rho(x, t) = f_{1/\sqrt{3}}(x, t). \quad (11)$$

The presence of $\nu = 1/\sqrt{3}$ is clear since the root mean square projection of the isotropically distributed unit vector of the velocity onto one of the axes is considered. However, the solution of Eq. (11) according to the properties (4)–(7) is nonzero only in the interval $[-t/\sqrt{3}, t/\sqrt{3}]$ and has δ-singularities at the boundary points, whereas the exact solution covers the interval $[-t, t]$ and at the boundary points has finite discontinuities equal to $(2t)^{-1}\exp\{-t\}$ due to nonscattered radiation (9). Figure 2 shows that in this problem, the telegraph approximation describes the transfer process less accurately than the diffusion approximation. This result, unexpected at first glance, calls for care regarding the conclusions of studies in which the time-dependent diffusion approximation is refined with the use of the telegraph equation [17, 18].

ANISOTROPIC MESODIFFUSION

As is evident from the foregoing, the relations between the diffusion and telegraph equations and the kinetic equation in the considered extreme cases are opposite in a certain sense; namely, in the one-dimensional case, the diffusion solution is a poorer approxi-

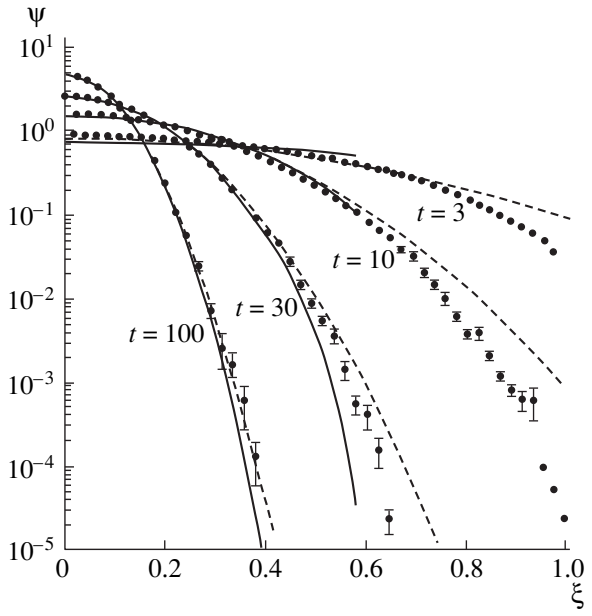

Fig. 2. The distributions $\psi(\xi, t) = t\rho(\xi t, t)$ for isotropic random walking. Dots are the results of the Monte Carlo simulation, solid curves are the telegraph approximation, and dashed curves are the diffusion approximation.

mation, while in the three-dimensional isotropic case, conversely, the telegraph approximation is less accurate than the diffusion approximation. A smooth transition from one case to the other can be conveniently achieved by taking the angle distribution of a particle after exit from the source and after each scattering in form (1).

The telegraph approximation in this scheme is easy to obtain using a Fourier–Laplace transform of the integral equation (8) with the distribution of nonscattered particles corresponding to the indicatrix of scattering (1):

$$\rho^{(0)}(x, t) = t^{-1}W_\nu(x/t)\exp\{-t\}, \quad |x| < t. \qquad (12)$$

As it follows from Eq. (8), the transform

$$\tilde\rho(k, \lambda) = \int_0^\infty dt \int_{-t}^{t} dx \exp\{ikx - \lambda t\}\rho(x, t)$$

satisfies the equation

$$\tilde\rho(k, \lambda) = \tilde\rho^{(0)}(k, \lambda)[1 + \tilde\rho(k, \lambda)], \qquad (13)$$

where

$$\tilde\rho^{(0)}(k, \lambda) = \int_0^\infty dt \int_{-t}^{t} dx \exp\{ikx - \lambda t\}\rho^{(0)}(x, t)$$

$$= (\alpha + 1)(\lambda + 1)\int_0^1 \frac{\omega^\alpha d\omega}{(\lambda + 1)^2 + (\omega k)^2}$$

$$= \frac{\alpha + 1}{\lambda + 1}\sum_{n=0}^{\infty} \frac{(-1)^n}{2n + \alpha + 1}\left(\frac{k}{\lambda + 1}\right)^{2n}.$$

At $k \longrightarrow 0$, this transform has the asymptotics

$$\tilde\rho^{(0)}(k, \lambda) \sim \frac{\lambda + 1}{(\lambda + 1)^2 + \nu^2 k^2}.$$

Substituting it into Eq. (13) and executing simple transformations, we obtain

$$[\lambda^2 + \lambda + \nu^2 k^2]\tilde\rho(k, \lambda) = \lambda + 1.$$

This relationship is nothing else than the Fourier–Laplace transform of the telegraph equation

$$\frac{\partial^2\rho}{\partial t^2} + \frac{\partial\rho}{\partial t} = \nu^2\frac{\partial^2\rho}{\partial x^2}, \quad t > 0 \qquad (14)$$

with the initial conditions $\rho(x, 0) = \delta(x)$ and $[\partial\rho(x, t)/dt]|_{t=0} = 0$.

A comparison of solution (3) above to Eq. (14) with the results of numerical solution of the integral equation (8) with kernel (12) by Monte Carlo simulation (see the Appendix) is shown in Fig. 3 in variables $\xi = x/t$ and

$$\psi(\xi, t) = t\rho(\xi t, t).$$

In the top part of the figure, it is seen that even at $\nu^2 = 1/2$, where the angle distribution linearly depends on ω, in the region $\xi < \nu$ the telegraph solution is closer to the exact solution than the diffusion one. As the anisotropy is further increased, a frontal burst appears in the exact solution, which describes the distribution of the particles that always move with a positive x-projection of the velocity. In the limit $\nu^2 \longrightarrow 1$, this burst transforms to the limit δ-function of the telegraph solution. At $\nu^2 \geq 0.7$ outside the frontal burst, the telegraph approximation is in good agreement with the exact solution.

FRONTAL BURST

An analytical description of the frontal burst can be conveniently obtained using the so-called method of moments [19, 20]. Let $z = t - x$ be a lag of the particle from the extreme point of the front t and $\rho^+(z, t)$ the distribution over z at time t of particles moving all the time in the positive x-axis direction, more specifically, those invariably having a positive projection of the velocity on this axis. For this density, an equation similar to Eq. (8) will also be valid (corrected for the nonnegativity of z):

$$\rho^+(z, t) = \rho^{(0)}(z, t) + \int_0^t dt' \int_0^{t'} dz' \rho^{(0)}(z', t')\rho^+(z - z', t - t'), \qquad (15)$$

where

$$\rho^{(0)}(z, t) \equiv t^{-1} W_\nu(1 - z/t)\exp\{-t\}, \quad 0 < z < t. \quad (16)$$

It is convenient now to perform the Laplace transform over both variables, z and t:

$$\tilde{\rho}(s, \lambda) = \int_0^\infty dt \int_0^\infty dz \exp\{-\lambda t - sz\}\rho^+(z, t),$$

giving the formula

$$\tilde{\rho}^+(s, \lambda) = \frac{\tilde{\rho}^{(0)}(s, \lambda)}{1 - \tilde{\rho}^{(0)}(s, \lambda)}. \quad (17)$$

Obviously,

$$[\partial^n \tilde{\rho}^+(s, \lambda)/\partial s^n]_{s=0} = (-1)^n \tilde{m}_n(\lambda),$$

where

$$\tilde{m}_n(\lambda) = \int_0^\infty \exp\{-\lambda t\} m_n(t) dt$$

is the Laplace transform of the nth moment of the distribution $\rho^+(z, t)$

$$m_n(t) = \int_0^\infty z^n \rho^+(z, t) dz.$$

For nonscattered radiation,

$$\tilde{\rho}^{(0)}(s, \lambda) = \frac{\alpha + 1}{2} \int_0^1 \frac{\omega^\alpha d\omega}{1 + \lambda + (1 - \omega)s},$$

from whence

$$\tilde{m}_0^{(0)}(\lambda) = \frac{1}{2(1 + \lambda)}, \quad \tilde{m}_1^{(0)}(\lambda) = \frac{1}{2(1 + \lambda)^2(\alpha + 2)},$$

$$\tilde{m}_2^{(0)}(\lambda) = \frac{2}{(1 + \lambda)^3(\alpha + 2)(\alpha + 3)},$$

and so forth. Differentiating Eq. (17) with respect to s, we obtain the corresponding expressions for the moments of the distribution of all particles forming the frontal burst:

$$\tilde{m}_0(\lambda) = \frac{1}{1 + 2\lambda},$$

$$\tilde{m}_1(\lambda) = \frac{2}{(\alpha + 2)(2\lambda + 1)^2},$$

$$\tilde{m}_2(\lambda) = 4\frac{2(2\lambda + 1)(\alpha + 2) + (\alpha + 3)}{(\alpha + 2)^2(\alpha + 3)(\lambda + 1)(2\lambda + 1)^3},$$

and so forth. The inverse Laplace transform by the res-

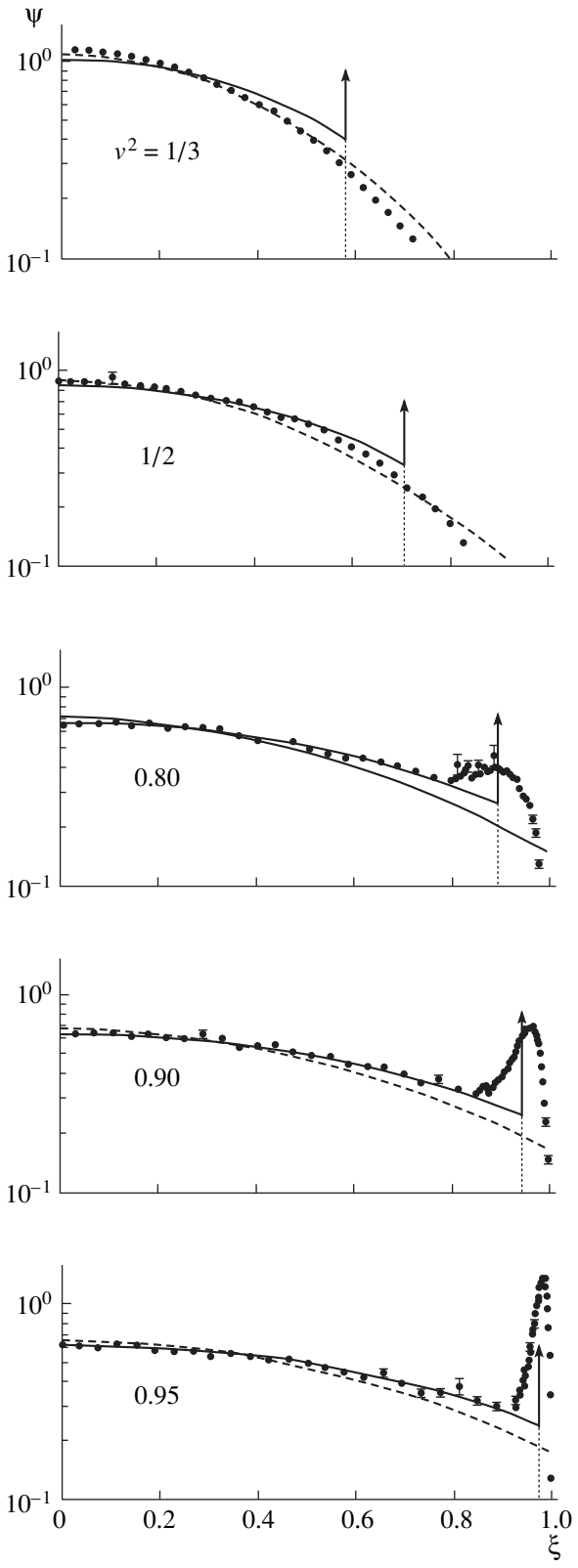

Fig. 3. The distributions $\psi(\xi, t)$ for $t = 5$ and the specified values of ν^2. The top plot corresponds to isotropic walk. The notations are the same as in Fig. 2. The δ-singularities in the telegraph approximation are denoted by vertical arrows.

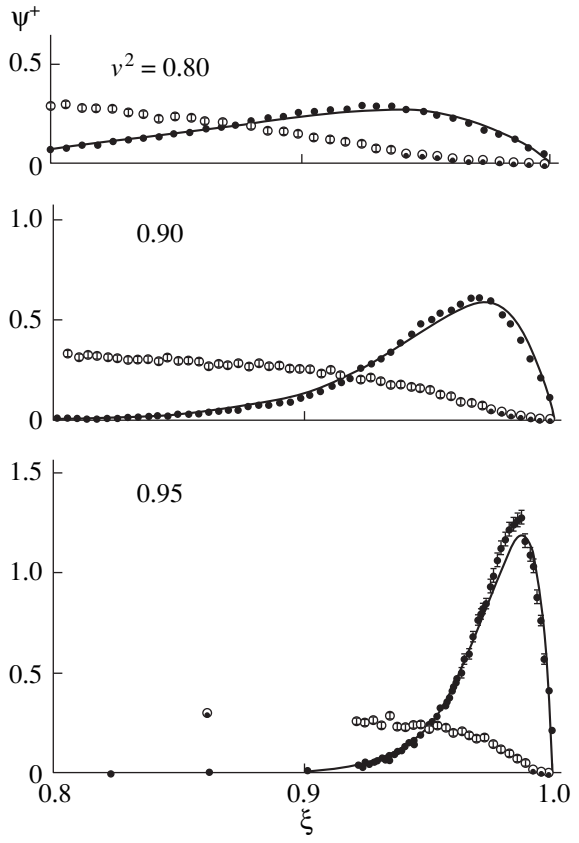

Fig. 4. Distributions $\psi^+(\xi, t) = t\rho^+(\xi t, t)$ of particles moving in the positive x-axis direction. Solid circles are the results of Monte Carlo simulation; solid curves are approximations by Eq. (21) and for particles which changed their x-component sign of the velocity at least once. Open circles are the results of the Monte Carlo simulation.

idue method gives

$$m_0(t) = (1/2)\exp\{-t/2\},$$
$$m_1(t) = \frac{t}{2(\alpha+2)}\exp\{-t/2\}, \tag{18}$$

$$m_2(t) = \left[\frac{4-2t+t^2/2}{(\alpha+2)^2} + \frac{4(t-2)}{(\alpha+2)(\alpha+3)}\right]\exp\{-t/2\} + \frac{4(\alpha+1)}{(\alpha+2)^2(\alpha+3)}\exp\{-t\}. \tag{19}$$

Using the obtained relationships for the moments, it is easy to find the center of gravity of the considered particle distribution:

$$\bar{x} = t - \bar{z} = t - m_1(t)/m_0(t) = \frac{\alpha+1}{\alpha+2}t = \frac{2t}{1+\nu^{-2}},$$

whose velocity

$$\dot{\bar{x}} = \frac{2}{1+\nu^{-2}}$$

at ν^2 close to the limiting value of 1 is only slightly less than the velocity of the front formed by the nonscattered particles (16) moving parallel to the x-axis. The dispersion of the considered packet

$$\sigma_x^2 = \sigma_z^2 = m_2(t)/m_0(t) - [m_1(t)/m_0(t)]^2$$

increases at $t \longrightarrow \infty$ as $4(\alpha+1)t/[(\alpha+2)^2(\alpha+3)]$, so that its relative width

$$\delta(t) = \sigma_x(t)/\bar{x}(t) \sim \frac{2}{\sqrt{(\alpha+1)(\alpha+3)t}}$$

tends to zero with time as $t^{-1/2}$. The propagation regime of such particles will be referred to as quasiballistic. It also includes particles in the ballistic regime; that is, the nonscattered particles whose distribution is given by density (16) and characterized by a finite surge

$$\Delta\rho^{(0)} = \rho^{(0)}(0, t) = \frac{\alpha+1}{2t}\exp\{-t\}. \tag{20}$$

The fraction of ballistic particles in a quasi-ballistic packet is

$$\eta(t) = m_0^{(0)}(t)/m_0(t) = \exp\{-t/2\}.$$

Except for step (20), the solution of Eq. (15) can be approximated by the formula

$$\rho^+(z, t) = \frac{\gamma\mu^2 z}{2\Gamma(2/\gamma)}\exp\{-t/2 - (\mu z)^\gamma\}.$$

It meets the normalizing condition

$$\int_0^\infty \rho^+(z, t)dz = (1/2)\exp\{-t/2\}$$

and has the moments

$$\int_0^\infty z^n\rho^+(z, t)dz = \frac{\Gamma((n+2)/\gamma)}{2\mu^n\Gamma(2/\gamma)}\exp\{-t/2\}.$$

Equating the first and second moments to those given by Eqs. (18) and (19), respectively, we obtain a system of equations to determine parameters γ and μ as functions of t and ν^2:

$$\frac{2\Gamma(4/\gamma)\Gamma(2/\gamma)}{[\Gamma(3/\gamma)]^2} = \frac{m_2(t)}{[m_1(t)]^2}\exp\{-t/2\},$$

$$\mu = \frac{\Gamma(3/\gamma)}{2m_1(t)\Gamma(2/\gamma)}\exp\{-t/2\}.$$

The results of the numerical calculations of $\gamma(t, \nu^2)$ and $\mu(t, \nu^2)$ are presented in the table. In Fig. 4, the

The parameters γ and μ in formula (21)

t	ν^2							
	0.80		0.90		0.95		0.99	
	γ	μ	γ	μ	γ	μ	γ	μ
2	0.854	12.69	0.765	35.73	0.729	84.69	0.703	484.8
4	1.103	3.777	0.978	9.908	0.928	22.70	0.892	126.5
6	1.401	1.808	1.229	4.505	1.161	10.07	1.111	55.01
8	1.756	1.093	1.521	2.622	1.428	5.752	1.362	30.98
10	2.184	0.758	1.861	1.767	1.737	3.822	1.650	20.36
12	2.709	0.573	2.262	1.308	2.096	2.799	1.981	14.78
14	3.378	0.459	2.747	1.032	2.521	2.189	2.368	11.48
16	4.280	0.384	0.348	0.851	3.035	1.793	2.828	9.359
18	5.609	0.332	4.129	0.726	3.678	1.521	3.391	7.900
20	7.914	0.294	5.213	0.635	4.524	1.324	4.107	6.849

approximation

$$\psi^+(\xi, t) = t\rho^+((1-\xi)t, t) = \frac{t^2\gamma\mu^2(1-\xi)}{2\Gamma(2/\gamma)}\exp\{-t/2-(\mu t(1-\xi))^\gamma\} \quad (21)$$

is compared with the results of the Monte Carlo simulation of the distribution of particles moving all the time in the positive x-axis direction. Good agreement between these results enables the solution of the kinetic equation to be presented in the form of a sum of the slowly varying component of the solution of the telegraph equation and two bursts $\psi^+(\xi, t)$ and $\psi^-(\xi, t)$, which are symmetric about the origin of coordinates.

CONCLUSIONS

The main conclusions of the study are as follows. The spatial distribution of particles emitted by an instantaneous point source in the case of mesodiffusion differs from the ordinary diffusion distribution by the occurrence of a front, beyond which the density is zero. The front moves away from the source with the velocity of a freely moving particle.

In the case of one-dimensional mesodiffusion, there are δ-singularities at both fronts accounting for the ballistic regime. Together with the continuous part of the solution of the telegraph equation, they give an exact solution of Boltzmann's kinetic equation.

Taking into account the angle distribution of the particles results in the disappearance of the singularities and the emergence instead of well pronounced frontal bursts in the case of strong anisotropy when $\nu^2 > 0.8$, which merge into the background of the smooth component at $\nu^2 < 0.7$. The shape of the surges corresponding to the ballistic regime is well approximated by formula (21).

At $\nu^2 > 0.7$, the continuous part of the telegraph equation is in satisfactory agreement with the solution of the kinetic equation (if ξ is not too close to 1) and, together with the above-mentioned approximation of the frontal burst, can be used as an approximation of the exact solution.

In the case of weak anisotropy, the telegraph approximation approximates the exact solution less accurately than the diffusion approximation.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research, project nos. 000100284 and 000217507.

APPENDIX

SOLUTIONS OF EQUATIONS (8) AND (15) USING MONTE CARLO SIMULATION

The analog simulation of the considered process presents no special problems; however, the obtained histograms are not very convenient for describing the density in the region of high gradient. Indeed, in the analog simulations, the estimation of the density ρ at point x^* at time t^* is obtained by building an elementary section (layer) of length Δx^* in the vicinity of the point x^* and calculating the quantity

$$\hat{\rho}(\Delta x^*, t^*) = \frac{N(\Delta x^*, t^*)}{N_0 \Delta x^*}, \quad (A1)$$

where N_0 is the total number of statistical sampling trajectories; $N(\Delta x^*, t^*)$ is the number of trajectories which are found in the layer Δx^* at the moment of time t^*.

The estimator (A1) is not unbiased and its mathematical expectation

$$M\hat{\rho}(\Delta x^*, t^*) = \frac{1}{\Delta x^*}\int\limits_{\Delta x^*} \rho(x, t^*)dx$$

is equal to the function $\rho(x, t^*)$ averaged over Δx^* rather than its sought-for value $\rho(x^*, t^*)$. According to the mean value theorem,

$$\frac{1}{\Delta x^*}\int\limits_{\Delta x^*} \rho(x, t^*) = \rho(x', t^*),$$

where $x' \in \Delta x^*$, but the exact position of x' is not known.

This uncertainty is the source of a horizontal error. Estimator (A1) itself contains a statistical "vertical" error.

Obviously, the horizontal error component can be reduced by decreasing Δx^*, which in turn results in an increase in the vertical error. It is possible to completely eliminate the horizontal error from the results by transition from the analog scheme to the following modification of the method. Let

$$J_h(t^*) = \int\limits_0^{t^*} dt \int dx h(x, t)\rho(x, t), \qquad \text{(A2)}$$

and X_1, T_1, X_2, T_2, …, $X_{N(t^*)}$, $T_{N(t^*)}$ be random coordinates and instants of time of the collisions of the particle in a given trajectory; $N(t^*)$ is a random number of collisions in the interval $(0, t^*)$. It is obvious that

$$\hat{J}_h(t^*) \equiv \sum_{i=1}^{N(t^*)} h(X_i, T_i) \qquad \text{(A3)}$$

is the unbiased estimate of the functional (A2)

$$M\hat{J}_h(t^*) = J_h(t^*).$$

Now represent Eq. (8) in the form

$$\rho(x^*, t^*) = \rho_0(x^*, t^*) + J_h(t^*),$$

where

$$h(x, t) = \rho_0(x^* - x, t^* - t).$$

According to Eqs. (A2) and (A3), the unbiased estimator of the functional $J_h(t^*)$ is

$$\hat{J}_h = \sum_{i=1}^{N(t^*)} \rho_0(x^* - X_i, t^* - T_i). \qquad \text{(A4)}$$

Since $\rho_0(x^* - X_i, t^* - T_i)$ in the adopted system of units describes the density of the flux and the density of collisions of nonscattered particles, then estimator (A4) can be considered as a sort of local estimation [21].

REFERENCES

1. I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems* (Nauka, Moscow, 1982; Wiley, New York, 1988).
2. V. F. Gantmakher and M. V. Feĭgel'son, Usp. Fiz. Nauk **168**, 113 (1998) [Phys. Usp. **41**, 105 (1998)].
3. *Direction in Condensed Matter Physics. Physics of Mesoscopic Systems*, Ed. by G. Grinstein and G. Mazenko (World Scientific, Singapore, 1986).
4. S. Godoy and L. S. García-Colin, Physica A (Amsterdam) **258**, 414 (1998).
5. B. Davison, *Neutron Transport Theory* (Clarendon, Oxford, 1957; Atomizdat, Moscow, 1960).
6. G. I. Marchuk, *Computing Methods of Nuclear Reactors* (Gosatomizdat, Moscow, 1961).
7. K. M. Case and P. F. Zweifel, *Linear Transport Theory* (Addison-Wesley, Reading, Mass., 1967; Mir, Moscow, 1972).
8. A. M. Weinberg and E. P. Wigner, *The Physical Theory of Neutron Chaim Reactors* (Univ. of Chicago Press, Chicago, 1959; Inostrannaya Literatura, Moscow, 1961).
9. K. H. Beckurts and K. Wirtz, *Neutron Physics* (Springer-Verlag, Berlin, 1964; Atomizdat, Moscow, 1968).
10. R. García-Pelayo, Physica A (Amsterdam) **258**, 365 (1998).
11. R. García-Pelayo, Physica A (Amsterdam) **216**, 299 (1995).
12. P. Frank and R. von Mises, *Die Differential- und Integralgleichungen der Mechanik und Physik* (Vieweg, Brunswick, 1930, 1935; ONTI, Moscow, 1937), Vols. 1, 2.
13. A. S. Monin, Izv. Akad. Nauk, Ser. Geofiz., No. 3, 234 (1995).
14. M. Kac, *Some Stochastic Problems in Physics and Mathematics* (Magnolia Petrolium Co., Dallas, 1956; Nauka, Moscow, 1967).
15. A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (MIT, Cambridge, 1971, 1975; Gidrometeoizdat, St. Petersburg, 1992).
16. V. V. Uchaikin, Physica A (Amsterdam) **255**, 65 (1998).
17. S. L. Sobolev, Usp. Fiz. Nauk **167**, 1095 (1997) [Phys. Usp. **40**, 1043 (1997)].
18. V. A. Lykov and V. G. Nikolaev, Zh. Vychisl. Mat. Mat. Fiz. **38**, 1907 (1998).
19. A. Ya. Énder and I. A. Énder, Zh. Tekh. Fiz. **69** (6), 22 (1999) [Tech. Phys. **44**, 628 (1999)].
20. A. Ya. Énder and I. A. Énder, Zh. Tekh. Fiz. **69** (9), 6 (1999) [Tech. Phys. **44**, 1005 (1999)].
21. S. M. Ermakov and G. A. Mikhaĭlov, *Course of Statistical Modeling* (Nauka, Moscow, 1976).

Translated by M. Lebedev

Technical Physics, Vol. 46, No. 2, 2001, pp. 147–153. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 16–22.

On the Nature of the Aharonov–Bohm Effect

A. G. Chirkov* and A. N. Ageev**

**St. Petersburg State Technical University,*
Politekhnicheskaya ul. 29, St. Petersburg, 195251 Russia
***Ioffe Physicotechnical Institute, Russian Academy of Sciences,*
Politekhnicheskaya ul. 26, St. Petersburg, 194021 Russia

Received March 31, 2000

Abstract—It is shown that zero-field potentials may be responsible for the Aharonov–Bohm effect. A magnetic field $\mathbf{B} = \mathrm{curl}\mathbf{A}^f$ has a physical (gauge-invariant) meaning for field potentials $\mathbf{A}^f$, whereas a circulation $\oint_C \mathbf{A}^0 d\mathbf{r}$ has a physical meaning for zero-field potentials $\mathbf{A}^0$. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The Aharonov–Bohm effect, an intriguing quantum-mechanical phenomenon, was predicted in 1939 and 1949 [1, 2] and then discovered and theoretically treated in 1959 [3]. Its essence is that a charged quantum particle moving in a region where a stationary magnetic (or electric) field is absent but the vector (or scalar) potential is nonzero is subjected to electromagnetic influence. A deeper insight into the nature of the Aharonov–Bohm effect (ABE) has been given in the course of long-standing discussion [4, 5]. However, inexact and sometimes incorrect assertions on this problem are still appearing in the physics literature. The necessary condition for the ABE is the presence of gauge-invariant potentials that do not produce electromagnetic fields. Since such potentials (zero-field potentials) must have the form

$$\mathbf{A}^0 = \mathrm{grad}\Psi, \quad \varphi^0 = -\frac{1}{c}\frac{\partial\Psi}{\partial t},$$

the function Ψ was, generally speaking, mistaken for the function of gradient transformation in the vast majority of works. This resulted in a paradox in both classical and quantum cases: "…, we may retain the present local theory and, instead, we may try to give a further new interpretation to the potentials. In other words, we are led to regard $\mathbf{A}(\mathbf{r})$ as a physical variable. This means that we must be able to define the physical difference between two quantum states which differ only by gauge transformation" [3]. In papers devoted to the ABE, the necessary condition for its existence has virtually not been discussed. In the stationary case, it was generally assumed that "a nontrivial topology of the region of charged particle propagation" is necessary for the ABE [6].

After the cogent experiments of Tonomura *et al*. [7], the potentialities of studying the stationary ABE had apparently been exhausted and physicists began to study the nonstationary ABE [8–12]. In [8, 9], the possibility of the quasi-ABE occurring in the presence of solely field potentials was considered. In [10–12], a variable magnetic flux was specified but corresponding fields and potentials were not discussed. Nevertheless, the origin and structure of the fields are of fundamental importance for the ABE. Indeed, as was shown (see, e.g., [13]), in the stationary case, the ABE is due to zero-field potentials, changing only the phase of the wave function. The necessary condition for the ABE is the presence of zero-field potentials that cannot be eliminated by gauge transformation.

The notion of zero-field (redundant) potentials was first introduced for solving boundary problems in electrodynamics of anisotropic media [14–16] and is very seldom encountered in the physics literature. The authors of [17] argue that "electric and magnetic field vectors cannot be expressed in terms of vector potentials" in anisotropic media. The conventional approach to such problems was to use electric and magnetic field strengths or, for zero scalar potentials, vector potentials proportional to them as unknown functions (Coulomb gauge). Such an approach turned out to exclude the possibility of satisfying boundary conditions in anisotropic media due to the intrinsic structure of Maxwell's equations (see Section 2).

The use of electromagnetic potentials with nonzero scalar potential in order to regularly satisfy boundary conditions was first proposed by academician Tikhonov in 1959 [18]. The evolution of these ideas has given rise to a general method for solving boundary problems in electrodynamics of anisotropic media, which is referred to as the method of redundant potentials [14–16, 19]. Its basic idea is the use of potentials that do not generate an electromagnetic field to regularly satisfy boundary conditions. The main results of this method are reported in Sections 1 and 2. However, even its authors state that "these potentials make no physical sense, since, being a direct consequence of gauge invariance, they generate zero electromagnetic fields" [19].

1063-7842/01/4602-0147$21.00

As is shown in Section 5, zero-field potentials, generally speaking, differ from gauge transformation despite the same form. The fact that two equivalence relations for vector potentials **A** and **A'** (curl**A** = curl**A'** and **A** – **A'** = gradχ) do not coincide in the general case was apparently first mentioned in [20] and rigorously proven in the cohomology theory [21].

1. ZERO-FIELD POTENTIALS IN A UNIAXIAL DIELECTRIC MEDIUM

Consider an anisotropic dielectric medium with a permittivity tensor $\hat{\varepsilon} = \mathrm{diag}(\varepsilon, \varepsilon, \varepsilon_1)$, i.e., a uniaxial medium, which is the simplest special case of a homogeneous anisotropic medium.

The free-field Maxwell equations for electromagnetic potentials have the form (in any gauge)

$$\mathrm{graddiv}\mathbf{A} - \Delta\mathbf{A} + \frac{\mu}{c^2}\frac{\partial^2}{\partial t^2}\hat{\varepsilon}\mathbf{A} + \frac{\mu}{c}\frac{\partial}{\partial t}\hat{\varepsilon}\mathrm{grad}\varphi = 0, \quad (1a)$$

$$\mathrm{div}\left[\hat{\varepsilon}\left(\frac{1}{c}\frac{\partial\mathbf{A}}{\partial t} + \mathrm{grad}\varphi\right)\right] = 0, \quad (1b)$$

where δ is the Laplacian operator. Projecting these equations on the principal axes of anisotropy, we obtain

$$-\frac{\partial^2 A_x}{\partial y^2} - \frac{\partial^2 A_x}{\partial z^2} + \frac{\mu\varepsilon}{c^2}\frac{\partial^2 A_x}{\partial t^2} + \frac{\partial^2 A_y}{\partial x\partial y} + \frac{\partial^2 A_z}{\partial x\partial z} + \frac{\mu\varepsilon}{c}\frac{\partial^2\varphi}{\partial t\partial x} = 0, \quad (2a)$$

$$\frac{\partial^2 A_x}{\partial y\partial x} - \frac{\partial^2 A_y}{\partial x^2} - \frac{\partial^2 A_y}{\partial z^2} + \frac{\mu\varepsilon}{c^2}\frac{\partial^2 A_y}{\partial t^2} + \frac{\partial^2 A_z}{\partial y\partial z} + \frac{\mu\varepsilon}{c}\frac{\partial^2\varphi}{\partial t\partial y} = 0, \quad (2b)$$

$$\frac{\partial^2 A_x}{\partial z\partial x} + \frac{\partial^2 A_y}{\partial z\partial y} - \frac{\partial^2 A_z}{\partial x^2} - \frac{\partial^2 A_z}{\partial y^2} + \frac{\mu\varepsilon_1}{c^2}\frac{\partial^2 A_z}{\partial t^2} + \frac{\mu\varepsilon_1}{c}\frac{\partial^2\varphi}{\partial t\partial z} = 0, \quad (2c)$$

$$\frac{\varepsilon}{c}\frac{\partial^2 A_x}{\partial x\partial t} + \frac{\varepsilon}{c}\frac{\partial^2 A_y}{\partial y\partial t} + \frac{\varepsilon_1}{c}\frac{\partial^2 A_z}{\partial z\partial t} + \varepsilon\frac{\partial^2\varphi}{\partial x^2} + \varepsilon\frac{\partial^2\varphi}{\partial y^2} + \varepsilon_1\frac{\partial^2\varphi}{\partial z^2} = 0. \quad (2d)$$

Let us use the Chetaev method [16, 19] to study system (2). Its determinant, involving the operator coefficients, is identically zero. This means that the equations in system (2) are linearly dependent. In this case, one of the equations, e.g., the last one, may be eliminated. Solving the obtained system by the Cramer rule, we find the equations of differential constraints imposed on the unknown functions:

$$\frac{\partial}{\partial x_0}LL_2\mathbf{A} = -\nabla LL_2\varphi, \quad (3)$$

where $x_0 = ct$ and the operators L and L_2 have the form

$$L = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \mu\varepsilon\frac{\partial^2}{\partial x_0^2}, \quad (4a)$$

$$L_1 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \mu\varepsilon_1\frac{\partial^2}{\partial x_0^2}, \quad (4b)$$

$$L_2 = L_1 - \left(1 - \frac{\varepsilon_1}{\varepsilon}\right)\frac{\partial^2}{\partial z^2} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\varepsilon_1}{\varepsilon}\frac{\partial^2}{\partial z^2} - \mu\varepsilon_1\frac{\partial^2}{\partial x_0^2}. \quad (4c)$$

Relation (3) implies that the general potentials **A** and φ are representable as the sums $\mathbf{A} = \mathbf{A}^f + \mathbf{A}^0$ and $\varphi = \varphi^f + \varphi^0$. The field-producing terms $\mathbf{A}^f$ and φ^f satisfy the basic (dispersion) relation (irrespective of the gauge)

$$LL_2\mathbf{A}^f = 0, \quad LL_2\varphi^f = 0. \quad (5)$$

The terms $\mathbf{A}^0$ and φ^0, which identically satisfy system (2) but do not satisfy (5), are related by the condition

$$\frac{\partial\mathbf{A}^0}{\partial x_0} + \mathrm{grad}\varphi^0 = 0;$$

i.e., they do not generate fields.

In quantum problems, just as in the stationary ABE, allowing for zero-field potentials causes only a change in the phase of the wave function. Therefore, the presence of zero-field potentials is the necessary condition for the ABE.

In the isotropic case ($\varepsilon_1 = \varepsilon$), each of the Cartesian coordinates of the vector potential is an independent solution of the wave equation. In order to satisfy the boundary conditions, three independent solutions are required [9]. Hence, the system of boundary conditions is closed. For an anisotropic medium, this is not the case: the introduction of zero-field potentials becomes a must to satisfy the boundary conditions. This assertion will be clarified in the next section.

2. GENERAL REPRESENTATIONS FOR A VECTOR POTENTIAL IN A UNIAXIAL MEDIUM

To complete the definition of system (2), we replace its last equation by a linear equation in the general form [19]

$$L_xA_x + L_yA_y + L_zA_z + L_0\varphi = 0, \quad (6)$$

which is the most general gauge condition (L_k are linear operators of differentiation with respect to coordinates x_k). Calculating the determinant of the resulting system, we find that the potentials have to satisfy the equations

$$LL_2S\mathbf{A} = 0, \quad LL_2S\varphi = 0 \tag{7}$$

with the operator

$$S = \frac{\partial}{\partial x}L_x + \frac{\partial}{\partial y}L_y + \frac{\partial}{\partial z}L_z - \frac{\partial}{\partial x_0}L_0.$$

Thus, the general representation of the potentials has the form $\mathbf{A} = \mathbf{A}^f_{(1)} + \mathbf{A}^f_{(2)} + \mathbf{A}^0$, where

$$L\mathbf{A}^f_{(1)} = 0, \quad \mathbf{A}^f_{(2)} = 0, \quad S\mathbf{A}^0 = 0 \tag{8}$$

and the potentials must satisfy the additional condition

$$\frac{\partial \mathbf{A}^0}{\partial x_0} + \mathrm{grad}\varphi^0 = 0;$$

i.e., $\mathbf{A}^0 = \mathrm{grad}\Psi$ and $\varphi^0 = -\partial\Psi/\partial x_0$, where Ψ is any differentiable function.

The last equation in (8) is referred to as redundant [19].

Let us put $L_\alpha = e_{\alpha\beta}\partial/\partial x_\beta$ (α and $\beta = x, y, z$) and $L_0 = \mu\partial/\partial x_0$. Then, Eq. (6) may be represented in the form of the generalized Lorentz condition

$$\mathrm{div}\hat{e}\mathbf{A}^f + \mu\frac{\partial\varphi^f}{\partial x_0} = 0, \tag{9}$$

and the operator of the redundant equation is given by

$$S = \mathrm{div}\hat{e}\mathrm{grad} - \mu\frac{\partial^2}{\partial x_0^2}. \tag{10}$$

Choosing the tensor $\hat{e}$ in the form

$$\hat{e} = \mathrm{diag}\left(\frac{1}{e}, \frac{1}{e}, \frac{1}{e_1}\right),$$

we obtain

$$S = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{e}{e_1}\frac{\partial^2}{\partial z^2} - \mu e\frac{\partial^2}{\partial x_0^2}. \tag{10a}$$

For $e = e_1 = \varepsilon$, the redundant operator S coincides with L; for $e = \varepsilon_1$ and $e_1 = \varepsilon$, $S = L_2$. The former case coincides with the Chetaev optimal gauge condition [19]; and the latter, with the Tikhonov condition [18]. In these and only in these cases, the redundant equation appears, showing up as one of the second-order equations into which the general equation decomposes [19]. For $e = e_1 = \varepsilon$, general system (2) takes the form

$$LA^f_x = 0, \quad LA^f_y = 0, \tag{11a, 11b}$$

$$L_2A^f_z = \left(1 - \frac{\varepsilon_1}{\varepsilon}\right)\frac{\partial}{\partial z}\left(\frac{A^f_x}{\partial x} + \frac{A^f_y}{\partial y}\right). \tag{11c}$$

From these equations,

$$A^{f(2)}_x = 0, \quad A^{f(2)}_y = 0, \quad A^f_z = A^{f(1)}_z + A^{f(2)}_z \tag{12}$$

with $LA^{f(1)}_z = 0$ and $L_2A^{f(2)}_z = 0$ in accordance with (8). From the last equations in (11) and (12), we come to the equation

$$L_2A^{f(1)}_z = \left(1 - \frac{\varepsilon_1}{\varepsilon}\right)\frac{\partial}{\partial z}\left(\frac{\partial A^{f(1)}_x}{\partial x} + \frac{\partial A^{f(1)}_y}{\partial y}\right). \tag{13}$$

In view of the Chetaev gauge condition, Eq. (13) yields

$$\frac{\partial A^{f(1)}_z}{\partial x_0} + \frac{\partial\varphi^{f(1)}}{\partial z} = 0; \tag{14}$$

i.e., $A^{f(1)}_z$ and $\varphi^{f(1)}$ coincide with the zero-field potentials. Therefore, $A^{f(1)}_z = 0$, which results in

$$\frac{\partial A^{f(1)}_x}{\partial x} + \frac{\partial A^{f(1)}_y}{\partial y} = 0. \tag{15}$$

Finally, for the Chetaev gauge condition, the general representation of the vector potential in a uniaxial medium (for $A_k, \varphi \sim e^{i\omega t}$) has the form [19]

$$A_x = A^{f(1)}_x + \frac{1}{ik_0}\frac{\partial\varphi^{(1)}_0}{\partial x}, \tag{16a}$$

$$A_y = A^{f(1)}_y + \frac{1}{ik_0}\frac{\partial\varphi^{(1)}_0}{\partial y}, \tag{16b}$$

$$A_z = Z^{f(1)}_z + \frac{1}{ik_0}\frac{\partial\varphi^{(1)}}{\partial z}, \tag{16c}$$

where $k_0 = \omega/c$ and $\omega^{(1)}_0$ satisfy the redundant equation $L\varphi^{(1)}_0 = 0$.

The general representation for the potentials (16) makes it possible to solve boundary problems in electrodynamics of uniaxial media. Unlike the case of an isotropic medium, here there appears additional condition (15), relating $A^{f(1)}_x$ to $A^{f(1)}_y$. However, the boundary conditions cannot be satisfied using solely the components of the field potentials. Indeed, let it be required to solve the Fock–Sommerfeld problem of a dipole field generalized for an anisotropic medium. In this case, it is necessary to determine seven functions (bounded at infinity): three components (A^0_x, A^0_y, and A^0_z) of the vector potential in a vacuum ($z < 0$), three components ($A^{f(1)}_x$, $A^{f(1)}_y$, and $A^{f(1)}_z$) of the vector

potential, and the function $\varphi_0^{(1)}$ in the anisotropic half-space ($z > 0$). Thus, to satisfy the boundary conditions, we have 14 constants. They can be determined from the seven conditions of boundedness at infinity. Three more equations result from continuity conditions for the vector potential components at $z = 0$. The continuity conditions for the normal derivatives of the tangential components of the vector potential yield two additional equations. One equation results from relation (15) between $A_x^{f(1)}$ and $A_y^{f(1)}$. Finally, we should take into account the Tikhonov field excitation condition (the jump of the normal derivative of the vector potential component that is parallel to the direction of current), which allows for the presence of a dipole.

Thus, one can uniquely solve this problem only by introducing zero-field potentials. This also means that, along with field potentials, zero-field potentials are also uniquely determined in a given gauge.

3. RELATION BETWEEN ZERO-FIELD POTENTIALS AND GAUGE TRANSFORMATION

Let initial potentials **A** and φ be related to each other by general gauge relation (6); that is,

$$\sum_{k=1}^{3} L_k A_k + L_0\varphi = 0, \quad \varphi = -L_0^{-1}\sum_{k=1}^{3} L_k A_k, \qquad (17)$$

and new potentials **A**' and φ' are related as

$$\sum_{k=1}^{3} L_k' A_k' + L_0'\varphi' = 0, \quad \varphi' = -(L_0')^{-1}\sum_{k=1}^{3} L_k' A_k'. \quad (18)$$

Here, L_k, L_k', L_0, and L_0' are any linear operators with constant coefficients.

It is well known that electromagnetic fields **E** and **B** are gauge-invariable:

$$\mathbf{A}' = \mathbf{A} + \mathrm{grad}\chi, \quad \varphi' = \varphi - \frac{1}{c}\frac{\partial\chi}{\partial t}.$$

Substituting these expressions into Eq. (18), we arrive at

$$\sum_{k=1}^{3} L_k'\frac{\partial\chi}{\partial x_k} - \frac{1}{c}L_0'\frac{\partial\chi}{\partial t} = -\sum_{k=1}^{3}(L_k' - L_0'L_0^{-1}L_k)A_k. \quad (19)$$

The last equation represents the condition that must be satisfied by the function χ of gauge transformation if the potentials are related to each other by gauge conditions (17) and (18). Generally, this equation differs from the redundant equation $S\Psi = 0$. The only special case where Eq. (19) coincides with the redundant one is the restricted gauge transformation $L_k = L_k'$ and $L_0 = L_0'$.

Finally, we should take into account that the function χ in gradχ must be single-valued (which is always true in a simply connected domain), since this is the only case when the circulations of **A** and **A**' coincide:

$$\oint \mathbf{A}d\mathbf{l} = \oint \mathbf{A}d\mathbf{l}. \qquad (20)$$

We have come to the conclusion that the equalities curl**A** = curl**A**' and **A**' = **A** + gradχ are equivalent only in a simply connected domain [21]. In contrast to differential equations describing local properties of the vector fields **A** and **A**', condition (20) describes their global properties.

Thus, even if the redundant equation for zero-field potentials coincides with that for the function of gauge transformation, their solutions [in view of (20)] may be different in multiply connected domains. It is this case that takes place in the ABE [3].

4. THE STRUCTURE OF THE VECTOR POTENTIAL IN THE CASE OF MAGNETOSTATIC ABE

In the initial version of the magnetostatic ABE [3], the influence of the region outside an infinite solenoid with constant current on electron motion was studied. Let us consider the structure of the vector potential in this problem on the basis of the above-developed theory. It is worth mentioning that, if the solenoid is considered as a smooth thin-walled continuous cylinder, we can take the entire space to be a homogeneous anisotropic medium with a conductivity that is nonzero only in the $\mathbf{e}_a$ direction at $\rho = R$ ($\rho = \sqrt{x^2 + y^2}$ and the z axis is aligned with the axis of solenoid symmetry). The thinner the wire and the less the winding pitch, the more accurate such a treatment. Therefore, the potential must have the form $\mathbf{A} = \mathbf{A}^f + \mathbf{A}^0$, where $\mathbf{A}^f$ is the field-generating potential and $\mathbf{A}^0 = \mathrm{grad}\Psi$ is the zero-field potential.

In the magnetostatic case ($\varphi = 0$ and $\mu = \varepsilon = 1$), $L = L_2 = \Delta$. The equations of differential constraints (3) will take the form

$$\frac{\partial}{\partial z}\Delta A_x = \frac{\partial}{\partial x}\Delta A_z, \qquad (21a)$$

$$\frac{\partial}{\partial z}\Delta A_y = \frac{\partial}{\partial y}\Delta A_z. \qquad (21b)$$

Hence, either $\Delta A_x = \Delta A_y = \Delta A_z = 0$ or

$$\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} = B_y = 0,$$

$$\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = B_x = 0.$$

Completing the definition of this system by operator (6) with $L_x = \partial/\partial x$, $L_y = \partial/\partial y$, $L_z = \partial/\partial z$, and $L_0 = 0$, i.e., by the condition div**A** = 0, we obtain Eq. (7) in the form $\Delta^2\mathbf{A} = 0$. This means that the operator of the redundant equation is $S = \Delta$ and coincides with L. By virtue of the

symmetry of the problem, one can put $A_\rho = A_z = 0$ and $A_\alpha = A_\alpha(\rho)$. In view of the boundedness of the solution at the origin, its form in the range $0 \le \rho < R$ may be $A_{1\alpha} = c_1\rho$ ($A_1^f = A_{1\alpha}$ and $A_1^0 = 0$). In the outer region ($\rho > R$), the general solution has the form $A_{2\alpha} = c_3\rho + c_2/\rho$. However, one cannot require the potential to be bounded at infinity, since the system is infinite. Physically, it is clear that the magnetic field outside the solenoid is equal to zero ($A_2^f = 0$). Therefore, in the outer region, the redundant equation $\Delta\mathbf{A}^0 = 0$ subject to $\mathrm{curl}\mathbf{A}^0 = 0$ should be solved. This condition selects the proper solution $A_{2\alpha} = c_2/\rho$. This fact was not found out earlier due to the coincidence of the redundant and general operators in this problem. In this case, the Tikhonov condition for field excitation is given by

$$\frac{\partial A_{1\alpha}}{\partial\rho} - \frac{\partial A_{2\alpha}}{\partial\rho}\bigg|_{\rho = R} = \frac{2I}{cR},$$

where I is the current per unit solenoid length. This condition is the same as that of the jump of the magnetic field tangential component in the usual statement of the problem.

Using the Tikhonov condition and the continuity condition for the potential at $\rho = R$, we obtain the final solution: $A_{1\alpha} = I\rho/cR$ ($0 \le \rho < R$) and $A_{2\alpha} = IR/c\rho$ ($R < \rho < \infty$). The constants in these formulas can be expressed in terms of the magnetic field flux inside the solenoid: $A_{1\alpha} = (\Phi/2\pi R^2)\rho$ and $A_{2\alpha} = \Phi/2\pi\rho$.

The potential in the outer region is a zero-field potential, so that $\mathbf{A}_2 = \mathrm{grad}\Psi$. However, since this region ($R < \rho < \infty$) is doubly connected, the function Ψ is multiple-valued and $\oint\mathbf{A}_2\,d\mathbf{r} \ne 0$ despite the fact that $\mathrm{curl}\mathbf{A}_2 \equiv 0$. Here, the nonequivalence mentioned in Section 3 takes place. This means that the Stokes theorem (in the form $\oint\mathbf{A}\,d\mathbf{r} = \oint\mathbf{B}\,d\mathbf{s}$, where the contour C encloses the solenoid in the outer region and S is the area inside the contour C) is inapplicable [21]. Nevertheless, this formula is incorrectly used in all the papers devoted to the ABE.

Outside the solenoid, the Stokes theorem applies to the region between two closed contours C_1 and C_2 that enclose the solenoid. A gap between the contours makes the region simply connected. In this case, $\oint_{C_1}\mathbf{A}_1\,dr = \oint_{C_2}\mathbf{A}_2 = \omega_1$, where ω_1 is the cyclic constant [21] that equals the flux $\Phi = 2\pi IR/c = \pi R^2 B$ in our case, where B is the field inside the solenoid. For a restricted gauge transformation, the function χ of gauge transformation (the same both inside and outside the solenoid) has to satisfy the Laplace equation $\Delta\chi = 0$ ($0 \le \rho < \infty$). It is well known [22, 23] that the solution of the Laplace equation in the entire space subject to $\chi|_{\rho=\infty} = 0$ is identically zero. Thus, the potentials $A_{1\alpha}$ and $A_{2\alpha}$ are defined uniquely (in the gauge $\mathrm{div}\mathbf{A} = 0$).

5. THE STRUCTURE OF THE VECTOR POTENTIAL FOR AN AC SOLENOID

Consider nonstationary excitation of an infinite cylindrical solenoid with infinitely thin walls to illustrate the existence of nonstationary zero-field potentials as a generalization of the stationary magnetostatic ABE. Choosing a cylindrical coordinate frame (ρ, α, z) with the z axis directed along the solenoid axis, we can represent the volume distribution of current as

$$j_\alpha(\rho, \alpha, z) = I_0\delta(\rho - R)\exp i(-n\alpha + \omega t), \qquad j_\rho = j_z = 0, \tag{22}$$

where R is the solenoid radius and ω is the circular frequency of the current.

The nonzero components of the vector potential, A_ρ and A_α, have the form [17] (further, we will omit the factor showing the harmonic dependence on time)

$$A\boldsymbol{\rho} = \int_V j_\alpha(\boldsymbol{\rho}')\sin(\alpha - \alpha')G(\boldsymbol{\rho}, \boldsymbol{\rho}')dV', \tag{23a}$$

$$A_\alpha = \int_V j_\alpha(\boldsymbol{\rho}')\cos(\alpha - \alpha')G(\boldsymbol{\rho}, \boldsymbol{\rho}')dV', \tag{23b}$$

where $G(\boldsymbol{\rho}, \boldsymbol{\rho}') = -\dfrac{i\pi}{c}H_0^{(2)}(k|\boldsymbol{\rho} - \boldsymbol{\rho}')$ is the Green function for the Helmholtz equation [17], $H_0^{(2)}$ is the Hankel function, and $k = \omega/c$.

The integrals appearing in (23) are easily evaluated using the addition formula for the Hankel functions [17]:

$$H_0^{(2)}(k\sqrt{\rho^2 + R^2 - 2\rho R\cos(\alpha - \alpha')}) = \sum_{m=-\infty}^{+\infty} e^{-im(\alpha - \alpha')}\begin{cases} H_m^{(2)}(kR)J_m(k\rho), & \rho < R \\ J_m(kR)H_m^{(2)}(k\rho), & \rho > R. \end{cases} \tag{24}$$

Eventually, we obtain

$$A_\alpha = -\frac{i\pi^2 I_0 R}{c}e^{-in\alpha} \times \begin{cases} H_{n+1}^{(2)}(kR)J_{n+1}(k\rho) + H_{n-1}^{(2)}(kR)J_{n-1}(k\rho), & \rho < R \\ J_{n+1}(kR)H_{n+1}^{(2)}(k\rho) + J_{n-1}(kR)H_{n-1}^{(2)}(k\rho), & \rho > R, \end{cases} \tag{25a}$$

$$A_\rho = -\frac{\pi^2 I_0 R}{2c} e^{-in\alpha} \times \begin{cases} H^{(2)}_{n+1}(kR) J_{n+1}(k\rho) - H^{(2)}_{n-1}(kR) J_{n-1}(k\rho), & \rho < R \\ J_{n+1}(kR) H^{(2)}_{n+1}(k\rho) - J_{n-1}(kR) H^{(2)}_{n-1}(k\rho), & \rho > R. \end{cases} \quad (25b)$$

In the case $n = 0$, i.e., in the absence of angle modulation, formula (11) yields

$$A_\alpha = -\frac{2i\pi^2 I_0 R}{c} \begin{cases} H^{(2)}_1(kR) J_1(k\rho), & \rho < R \\ J_1(kR) H^{(2)}_1(k\rho), & \rho > R, \end{cases} \quad (26)$$
$$A_\rho = 0.$$

In the stationary case ($\omega \longrightarrow 0$), these relations result in the well-known formulas $A_\alpha = J\rho/cR$ and $A_\alpha = JR/c\rho$ ($\rho > R$), where $J = 2\pi R I_0$ is the current per unit length in the solenoid wall.

The only nonzero magnetic-field component corresponding to potentials (26) has the form

$$B_z = -\frac{2i\pi^2 I_0 R k}{c} \begin{cases} H^{(2)}_1(kR) J_0(k\rho), & \rho < R \\ J_1(kR) H^{(2)}_0(k\rho), & \rho > R, \end{cases} \quad (27)$$

for alternating current or

$$B_z = \frac{4\pi I_0}{c} \ (\rho < R), \quad B_z = 0 \ \ (\rho > R) \quad (28)$$

in the stationary case.

In order to divide the vector potential outside the solenoid into two terms according to (1), we separate out the term whose curl is equal to zero in (12). This can be done uniquely as follows:

$$A_\alpha = -\frac{2i\pi^2 I_0 R}{c} J_1(kR) \tilde{H}^{(2)}_1(k\rho) + \frac{4\pi I_0 R J_1(kR)}{\rho} \equiv A^f_\alpha + A^0_\alpha, \quad (29)$$

where A^f_α is the component of the field potential, A^0_α is the component of the zero-field potential, and $\tilde{H}^{(2)}_1(k\rho) = H^{(2)}_1(k\rho) - 2i/\pi k\rho$.

Let us write the real parts of the potential components in (29):

$$\mathrm{Re} A^f_\alpha = W\left\{\pi J_1(k\rho)\sin\omega t - \left[\frac{2}{k\rho} + \pi Y_1(k\rho)\right]\cos\omega t\right\}, \quad (30a)$$

$$\mathrm{Re} A^0_\alpha = W\frac{2}{k\rho}\cos\omega t, \quad (30b)$$

where

$$W = \frac{2\pi I_0 R J_1(kR)}{c}$$

and Y_1 is the Neumann function.

In the stationary case, $A^f_\alpha = 0$ and $A^0_\alpha = JR/c\rho$. From the condition of absence of an electric field, one can obtain the expression for the scalar zero-field potential:

$$\varphi^0 = -\frac{4\pi i I_0 R}{c} J_1(kR)\alpha, \quad (31)$$

where α is the azimuthal angle. In the stationary case, $\varphi^0 = 0$.

Note that there exists a relation between the solenoid radius and the electromagnetic field wavelength. This relation is defined by the roots of the equation $J_1(kR) = 0$ and yields the zero field outside the solenoid, as follows from (15). However, in this nonstationary case, zero-field potentials also vanish; here, we are dealing with a closed waveguide. Zero-field potentials are equal to zero for all $n \neq 0$ as well. One can easily evaluate the flux inside the solenoid:

$$\Phi = \int_0^R \int_0^{2\pi} B_{1z}(\rho)\rho d\rho d\alpha = -\frac{i4\pi^3 R^2 I_0}{c} J_1(k_0 R) H^{(2)}_1(k_0 R). \quad (32)$$

The circulation of the vector $\mathbf{A}^0$ on any contour enclosing the solenoid (cyclic constant) is given by

$$\omega_1 = \oint_C \mathbf{A}_0 d\mathbf{r} = \frac{8\pi^2 I_0 R}{c k_0} J_1(k_0 R). \quad (33)$$

Obviously, $\omega_1 \neq \Phi$ in the general (nonstationary) case. Therefore, their coincidence in the stationary case ($k_0 \longrightarrow 0$) should be considered as accidental.

Generally speaking, this fact completely changes our understanding of the ABE. The reason for the ABE is zero-field potentials. A magnetic field $\mathbf{B} = \mathrm{curl}\,\mathbf{A}^f$ has a physical (gauge-invariant) meaning for field potentials $\mathbf{A}^f$, whereas a circulation $\oint_C \mathbf{A}^0 dr$ has a physical meaning for zero-field potentials $\mathbf{A}^0$.

REFERENCES

1. W. Franz, Verh. Dtsch. Phys. Ges. **20**, 65 (1939).
2. W. Ehrenberg and R. E. Siday, Proc. Phys. Soc. London, Sect. B **62**, 8 (1949).
3. V. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
4. S. Olariu and I. I. Popesku, Rev. Mod. Phys. **57**, 339 (1985).
5. M. Peskin and A. Tonomura, Lect. Notes Phys. **340**, 115 (1989).

6. V. G. Bagrov, D. N. Gitman, and V. D. Skarzhinskiĭ, Tr. Fiz. Inst. Akad. Nauk SSSR **176**, 151 (1986).
7. A. Tonomura, N. Osakabe, T. Matsuda, *et al.*, Phys. Rev. Lett. **56**, 792 (1986).
8. B. Lee, E. Yin, T. K. Gustafson, and R. Chiao, Phys. Rev. A **45**, 4319 (1992).
9. A. N. Ageev, S. Yu. Davydov, and A. G. Chirkov, Pis'ma Zh. Tekh. Fiz. **26** (9), 70 (2000) [Tech. Phys. Lett. **26**, 392 (2000)].
10. A. Vourdas, Europhys. Lett. **32**, 289 (1995).
11. A. Vourdas, Phys. Rev. A **56**, 2408 (1997).
12. G. Afanasiev and M. Nelhiebel, Phys. Scr. **54**, 417 (1996).
13. D. H. Kobe, Ann. Phys. **123**, 381 (1979).
14. D. N. Chetaev, Zh. Tekh. Fiz. **32**, 1342 (1962) [Sov. Phys. Tech. Phys. **7**, 991 (1963)].
15. D. N. Chetaev, Izv. Akad. Nauk SSSR, Fiz. Zemli, No. 10, 45 (1966).
16. D. N. Chetaev, Dokl. Akad. Nauk SSSR **174**, 775 (1967) [Sov. Phys. Dokl. **12**, 555 (1967)].
17. G. T. Markov and A. F. Chaplin, *Excitation of Electromagnetic Waves* (Moscow, 1967).
18. A. N. Tikhonov, Dokl. Akad. Nauk SSSR **126**, 967 (1959) [Sov. Phys. Dokl. **4**, 566 (1959)].
19. M. G. Savin, *Problems of Lorentz Gauge in an Anisotropic Medium* (Moscow, 1979).
20. F. Valz-Gris and C. Magni, J. Math. Phys. **36**, 177 (1995).
21. B. A. Dubrovin, S. P. Novikov, and A. T. Fomenko, *Modern Geometry: Methods of Homology Theory* (Nauka, Moscow, 1984).
22. V. I. Smirnov, *A Course of Higher Mathematics* (Nauka, Moscow, 1968; Addison-Wesley, Reading, 1964), Vol. 2.
23. N. E. Kochin, *Vector Calculus* (Moscow, 1967).

Translated by M. Fofanov

Technical Physics, Vol. 46, No. 2, 2001, pp. 154–157. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 23–26.

Transformation Properties of Wave Functions of a Two-Dimensional Harmonic Oscillator

I. E. Mazets

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

Received May 16, 2000

Abstract—A transformation formula for wave functions of type $\psi_{n_1}(x_1)\psi_{n_2}(x_2)$, where ψ_n is the eigenfunction of a harmonic oscillator for the nth energy level, was found for an arbitrary rotation of the (x_1, x_2) Cartesian system. By way of example, matrix elements for contact interaction of two particles in an external harmonic potential field were calculated. © *2001 MAIK "Nauka/Interperiodica".*

STATEMENT OF THE PROBLEM

Let us have an isotropic harmonic oscillator of mass m and fundamental frequency ω. Its complete Hamiltonian can be represented as the sum of one-dimensional Hamiltonians describing oscillation along two orthogonal axes x_1 and x_2:

$$\hat{H} = \hat{H}_1 + \hat{H}_2,$$
$$\hat{H}_j = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x_j^2} + \frac{1}{2}m\omega^2 x_j^2, \quad j = 1, 2. \tag{1}$$

The properties of the eigenfunctions of operators $\hat{H}_j$ are well known [1]. We will take advantage of the expression for the eigenfunction for the nth energy level ($n = 0, 1, 2, \ldots$) normalized to unity:

$$\psi_n(x_j) = \frac{1}{\sqrt{n!}}(\hat{a}_j^\dagger)^n \psi_0(x_j), \tag{2}$$

where

$$\psi_0(x_j) = \left(\frac{m\omega}{\pi\hbar}\right)^{1/4} \exp\left(-\frac{m\omega x_j^2}{2\hbar}\right) \tag{3}$$

is the wave function of the ground state. The quantum production operator is explicitly written as

$$\hat{a}_j^\dagger = \sqrt{\frac{m\omega}{2\hbar}}x_j - \sqrt{\frac{\hbar}{2m\omega}}\frac{\partial}{\partial x_j}, \tag{4}$$

and its conjugate (annihilation operator), as

$$\hat{a}_j = \sqrt{\frac{m\omega}{2\hbar}}x_j + \sqrt{\frac{\hbar}{2m\omega}}\frac{\partial}{\partial x_j}. \tag{5}$$

Our goal is to see how the wave functions of type $\psi_{n_1}(x_1)\psi_{n_2}(x_2)$ transform when the coordinate system rotates through an angle χ in the plane (x_1, x_2); in other words, how the wave functions transform in going to the new (primed) coordinates

$$x_1' = \cos\chi x_1 + \sin\chi x_2, \quad x_2' = -\sin\chi x_1 + \cos\xi x_2. \tag{6}$$

As far as we know, such a transformation has not been considered in the literature. This is associated largely with the fact that it is natural to represent the wave function of an axially symmetric harmonic oscillator as the product of the radial part, which depends on the polar radius alone and hence is invariant under rotations, and a function like $\exp(iM\varphi)$, where φ is an angular variable and M is an integer. Moreover, the addition theorem for Hermitian polynomials, or, which is the same, for eigenfunctions of a harmonic oscillator, has also not been covered in the mathematical literature concerning special functions (see, e.g., the comprehensive monograph [2]).

Before calculating the coefficients in the transform

$$\psi_{n_1}(x_1)\psi_{n_2}(x_2) = \sum_{n_1', n_2'} K_{n_1 n_2}^{n_1' n_2'}(\chi)\psi_{n_1'}(x_1')\psi_{n_2'}(x_2'), \tag{7}$$

we will point to their obvious properties. First, when the coordinate axes are rotated, only degenerate states belonging to the same eigenvalue of the complete Hamiltonian $\hat{H}$ mix. Hence, the necessary condition for $K_{n_1 n_2}^{n_1' n_2'}(\chi)$ to be other than zero is the fulfillment of the equality

$$n_1 + n_2 = n_1' + n_2'. \tag{8}$$

Second, since the basis functions are real, so are the coefficients of the transform; for them, the unitary property changes to the property of orthogonality. Thus, the inverse transform (in going from the new

coordinates to the old) has the form

$$\psi_{n_1'}(x_1')\psi_{n_2'}(x_2') = \sum_{n_1, n_2} K_{n_1 n_2}^{n_1' n_2'}(\chi)\psi_{n_1}(x_1)\psi_{n_2}(x_2). \quad (9)$$

CALCULATION OF THE TRANSFORM COEFFICIENTS

Direct calculation of the coefficients $K_{n_1 n_2}^{n_1' n_2'}(\chi)$ implies the definition of the quantum production operators $\hat{a}_1'^{\dagger}$ and $\hat{a}_2'^{\dagger}$ in the new coordinates. This is done by replacing x_j in (4) by x_j'. With (6), we find

$$\begin{aligned} \hat{a}_1^{\dagger} &= \cos\chi\hat{a}_1'^{\dagger} - \sin\chi\hat{a}_2'^{\dagger}, \\ \hat{a}_2^{\dagger} &= \sin\chi\hat{a}_1'^{\dagger} + \cos\chi\hat{a}_2'^{\dagger}. \end{aligned} \quad (10)$$

These expressions are substituted into the definitions of $\psi_{n_1}(x_1)$ and $\psi_{n_2}(x_2)$. Removing the parentheses in the $\psi_{n_1}(x_1)\psi_{n_2}(x_2)$ product, collecting similar terms, and bearing in mind that $\psi_0(x_1)\psi_0(x_2) = \psi_0(x_1')\psi_0(x_2')$, we find the coefficient preceding $\psi_{n_1'}(x_1')\psi_{n_2'}(x_2')$ in the form of a finite sum. After cumbersome computation using various identities known from the angular momentum theory [3], we eventually come to

$$K_{n_1 n_2}^{n_1' n_2'}(\chi) = d^{\frac{1}{2}(n_1 + n_2)}_{\frac{1}{2}(n_2' - n_1')\frac{1}{2}(n_2 - n_1)}(2\chi). \quad (11)$$

In view of the known property [3]

$$d^J_{MM'}(\beta) = d^J_{-M'-M}(\beta),$$

formula (11) can be recast as

$$K_{n_1 n_2}^{n_1' n_2'}(\chi) = d^{\frac{1}{2}(n_1 + n_2)}_{\frac{1}{2}(n_1 - n_2)\frac{1}{2}(n_1' - n_2')}(2\chi). \quad (12)$$

Here, $d^J_{MM'}$ is the function whereby the Wigner D function that corresponds to a rotation through Eulerian angles (α, β, γ) is expressed

$$D^J_{MM'}(\alpha, \beta, \gamma) = \exp(-iM\alpha)d^J_{MM'}(\beta)\exp(-iM'\gamma).$$

Under rotation, the wave functions with certain total angular momentum J and its projection M onto the z axis transform as

$$\begin{aligned} &\Psi_{JM'}(\vartheta', \varphi', \sigma') \\ &= \sum_{M=-J}^{J} \Psi_{JM}(\vartheta, \varphi, \sigma)D^J_{MM'}(\alpha, \beta, \gamma). \end{aligned} \quad (13)$$

The argument of the wave function involves two angular variables ϑ and φ and the spin variable σ; the primed quantities are their values in the new (rotated) coordinate system. In essence, relation (13) defines the Wigner functions.

We omit a routine procedure of directly computing the transform coefficients $K_{n_1 n_2}^{n_1' n_2'}(\chi)$, the more so as expressions (11) and (12) can be obtained by another, less trivial but faster, way. It uses isospin formalism, widely employed in the theory of two-level atom ensembles [4]. Let us introduce the operators of isospin projections onto three orthogonal axes in an auxiliary (isospin) space:

$$\begin{aligned} \hat{I}_1 &= \frac{1}{2}(\hat{a}_1\hat{a}_2^{\dagger} + \hat{a}_1^{\dagger}\hat{a}_2), \quad \hat{I}_2 = \frac{i}{2}(\hat{a}_1\hat{a}_2^{\dagger} - \hat{a}_1^{\dagger}\hat{a}_2), \\ \hat{I}_3 &= \frac{1}{2}(\hat{a}_1^{\dagger}\hat{a}_1 - \hat{a}_2^{\dagger}\hat{a}_2). \end{aligned} \quad (14)$$

It is easy to check that these operators satisfy the usual commutation relations for angular momentum

$$[\hat{I}_1, \hat{I}_2] = i\hat{I}_3; \quad (15)$$

two other similar relations are obtained by circularly permutating subscripts 1, 2, and 3. Also, each of the operators $\hat{I}_k$ (k = 1, 2, or 3) is commutatively related to the spin square operator:

$$\sum_{k=1}^{3} \hat{I}_k^2 = \hat{I}(\hat{I} + 1), \quad (16)$$

where

$$\hat{I} = \frac{1}{2}(\hat{a}_1^{\dagger}\hat{a}_1 + \hat{a}_2^{\dagger}\hat{a}_2). \quad (17)$$

Thus, the wave function $\psi_{n_1}(x_1)\psi_{n_2}(x_2)$ is the eigenfunction of the operators $\hat{I}$ and $\hat{I}_3$, and the eigenvalues of the operators are $1/2(n_1 + n_2)$ and $1/2(n_1 - n_2)$, respectively.

Now consider a rotation through an angle χ in the (x_1, x_2) plane. It is evident that

$$\psi_{n_1'}(x_1')\psi_{n_2'}(x_2') = \exp\left(-\chi\frac{\partial}{\partial\phi}\right)\psi_{n_1}(x_1)\psi_{n_2}(x_2), \quad (18)$$

where the variable ϕ is a polar angle reckoned from the x_1 axis toward the x_2 axis. In addition [1, 3],

$$\frac{\partial}{\partial\phi} = x_1\frac{\partial}{\partial x_2} - x_2\frac{\partial}{\partial x_1}.$$

Using definitions (4) and (14), one can show that

$$\psi_{n_1'}(x_1')\psi_{n_2'}(x_2') = \exp(-2i\chi\hat{I}_2)\psi_{n_1}(x_1)\psi_{n_2}(x_2); \quad (19)$$

in other words, a rotation through an angle χ in the (x_1, x_2) plane is reduced to a rotation through an angle 2χ about the x_2 axis in the isospin plane. Thus, formula (12)

and its equivalent (11) immediately follow from (19) in view of (13).

MATRIX ELEMENTS OF CONTACT INTERACTION

By way of example, the relations obtained will be used in calculating matrix elements of contact interaction. As is known [5], in ensembles of ultracold atoms, where *s*-wave isotropic scattering of length *a* dominates because of the smallness of the collision energy, atomic interaction can be approximated by the contact pseudopotential

$$U(\mathbf{r}-\mathbf{r}') = g\delta^{(3)}(\mathbf{r}-\mathbf{r}'), \tag{20}$$

where $\mathbf{r}$ and $\mathbf{r}'$ are the positions of colliding atoms (μ is their reduced mass), $g = 2\pi\hbar^2\mu^{-1}a$ is the effective interaction constant, and $\delta^{(3)}$ is the three-dimensional Dirac delta function.

To refine the existing theory of degenerate atomic gas in a magnetic trap [5], it seems of interest to calculate matrix elements of contact interaction using the basis of eigenfunctions of a harmonic oscillation. Consider two atoms of equal mass m ($\mu = m/2$) that had quantum numbers n_j^k before collision (the subscript $j = 1, 2$ numbers the atoms, while the superscript $k = x, y, z$ stands for the Cartesian axes directed along the magnetic axes of the trap). The quantum numbers after collision will be primed. In the general case, all three fundamental frequencies of oscillation along the axes differ; we, however, will omit the super- and subscripts at ω to make the mathematical notation simpler. The desired matrix element is defined as

$$\langle n_j^{k'}|U|n_j^k\rangle = g\prod_{k=x}^{z} T_{n_2^{k'}n_1^{k'}n_1^k n_2^k}, \tag{21}$$

where

$$T_{n_2'n_1'n_1n_2} = \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2 \psi_{n_2'}(x_2)\psi_{n_1'}(x) \times \delta(x_1-x_2)\psi_{n_1}(x_1)\psi_{n_2}(x_2) \tag{22}$$

is the matrix element from the one-dimensional delta function. Two properties of $T_{n_2'n_1'n_1n_2}$ immediately follow from definition (22). First, this element remains invariable for any permutation of its four subscripts. Second, it is other than zero only if the sum $n_1 + n_2 + n_1' + n_2'$ is even.

To compute matrix element (22), we will transform the coordinates in such a way as to separate the center-of-mass motion of two particles and their relative motion:

$$x_1' = \frac{1}{\sqrt{2}}(x_1+x_2), \quad x_2' = \frac{1}{\sqrt{2}}(x_2-x_1). \tag{23}$$

Formally, this transformation is a rotation through an angle of $\pi/4$ in the (x_1, x_2) "plane." With (7), we find

$$T_{n_2'n_1'n_1n_2} = \frac{1}{\sqrt{2}}\sum_{\nu_1,\nu_2,\nu_1',\nu_2'} K_{n_1n_2}^{\nu_1\nu_2}(\pi/2)K_{n_1'n_2'}^{\nu_1'\nu_2'}(\pi/2) \times \delta_{\nu_1\nu_1'}\psi_{\nu_2'}(0)\psi_{\nu_2'}(0) \tag{24}$$

($\delta_{\nu_1\nu_1'}$ is Kronecker's symbol). Substituting the eigenfunctions of a harmonic oscillator in the explicit form (i.e., expressing them through Hermitian polynomials) [1] and using the basic result of this work [expression (11)], one can represent (24) in the form

$$T_{n_2'n_1'n_1n_2} = (-1)^{J'-J}\sqrt{\frac{m\omega}{2\pi\hbar}} \times \sum_s d^J_{M2s-J}(\pi/2)d^J_{M'+2s+J'-2J}(\pi/2) \times \frac{\sqrt{(2s)!(2s+2J'-2J)!}}{s!(s+J'-J)!2^{2s+J'-J}}, \tag{25}$$

where s under the summation sign runs over all integers for which the arguments of the factorials in (25) are nonnegative. Hereafter, we introduce the designations

$$J = \frac{1}{2}(n_1+n_2), \quad J' = \frac{1}{2}(n_1'+n_2'),$$

$$M = \frac{1}{2}(n_2-n_1), \quad M' = \frac{1}{2}(n_2'-n_1').$$

After tedious routine rearrangements that follow from the angular momentum theory [3], (25) is reduced to a form that is free of the Wigner functions but contains Clebsch–Gordan coefficients $C^{jm}_{j_1m_1j_2m_2}$ (such a form is very convenient when the computation is performed with the Mathematica 3.0 program package, where the Clebsch–Gordan coefficients are used as standard functions):

$$T_{n_2'n_1'n_2} = \zeta_{2(J+J')}\sqrt{\frac{m\omega}{2\pi\hbar}}\sum_{l=|J-J'|}^{J+J'}\sum_{J_3=|J-J'|}^{J+J'}\sum_{J_4=\max(|J_3-l|,|M-M'|)}^{J_3+l} \times (-1)^{(J_4+M'-M+J+J'-l)/2}\zeta_{l-J'+J} \times \zeta_{J_4-M+M'}C^{J_3M-M'}_{JMJ'-M'}C^{J_4M-M'}_{J_3M-M'l0}C^{J_3J-J'}_{Jl-J'J'J-l}C^{J_40}_{J_3J-J'lJ'-J} \tag{26}$$

$$\times \frac{\sqrt{(J_4 - M + M')!(J_4 + M - M')!}}{2^{J_4}[(J_4 - M + M')/2]![J_4 + M - M')/2]!}.$$

Here, $\zeta_s = \frac{1}{2}[(-1)^s + 1]$ (i.e., $\zeta_s = 1$ or 0 if s is even or odd, respectively). Matrix elements for powers of the difference between the coordinates of two atoms (and, hence, for the energy of their interaction in the form of a power series other than pseudopotential (20)) can be calculated in a similar manner.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (grant no. 99-02-17 076) and the Program "Universities of Russia" (grant no. 015.01.01.04).

The author thanks D.A. Varshalovich for valuable discussions.

REFERENCES

1. A. S. Davydov, *Quantum Mechanics* (Nauka, Moscow, 1973; Pergamon, Oxford, 1976); L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989, 4th ed.; Pergamon, New York, 1977, 3rd ed.).
2. V. T. Erofeenko, *Theory of Addition* (Nauka i Tekhnika, Minsk, 1989).
3. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Scientific, Singapore, 1988).
4. A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics* (Nauka, Moscow, 1988); M. M. Al'perin, Ya. D. Klubis, and A. I. Khizhnyak, *Introduction to the Physics of Two-Level Atoms* (Naukova Dumka, Kiev, 1987).
5. L. P. Pitaevskiĭ, Usp. Fiz. Nauk **168** (6), 641 (1998) [Phys. Usp. **41**, 569 (1998)].

Translated by V. Isaakyan

Technical Physics, Vol. 46, No. 2, 2001, pp. 158–166. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 27–35.

GASES AND LIQUIDS

Nonlinear Capillary Vibrations and Stability of a Highly Charged Drop under Single-Mode Initial Deformation of Large Amplitude

S. O. Shiryaeva

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

e-mail: shir@uniar.ac.ru

Received April 12, 2000

Abstract—Nonlinear vibrations of a charged drop that are caused by a virtual initial perturbation of the equilibrium spherical shape of the drop were considered. The perturbation can be proportional to any of the free vibration modes. The spectrum and stability of the vibrations were studied correct to second order of smallness.

(1) Charged drop stability is of considerable interest in geophysical applications, scientific instrumentation, technical physics, and engineering [1]. However, most of the theoretical investigations into the problem were based on a linear approximation. The nonlinearity of the phenomenon has been taken into account in analytical and numerical studies only recently [2–10], although some results of the nonlinear approach were also obtained within simpler linear approximations [11, 12].

We will consider a drop whose electric charge is smaller than critical in terms of Rayleigh instability. Let its surface undergo vibrations of large amplitude. Our aim is to calculate the shape of the drop and the parameters of surface vibrations. We formulate the problem in a more general nonlinear form than in [2–10] and solve it by the multiple scale method up to terms of second order of smallness in amplitude of the initial disturbance of the drop surface. We use a different method of satisfying initial conditions than in [2–10]. It allows us to obtain the solution up to second-order terms when the initial perturbation is associated with the excitation of an arbitrary vibrational mode.

(2) Consider a drop of an ideal incompressible perfectly conducting liquid with a density ρ and a surface tension coefficient σ and study the time evolution of its shape. We assume that the drop is in a vacuum, its charge equals Q, and its volume equals the volume of a sphere of radius R. At the initial time instant $t = 0$, the equilibrium spherical shape of the drop undergoes a virtual axisymmetric perturbation of fixed amplitude that is much smaller than the drop radius and is proportional to one of the Legendre polynomials. We aim at calculating the spectrum of drop capillary vibrations occurring under such conditions, i.e., at finding the shape of the drop at each subsequent time instant $t > 0$. In what follows, we will use the dimensionless variables for which $R = \rho = \sigma = 1$.

Since the initial perturbation of the drop is axisymmetric and small, we introduce the following simplifications. The shape of the drop is assumed to be axisymmetric both at the initial and at each subsequent time instant, and the equation for the shape of the drop written in polar coordinates with the origin at the center of the drop has the form

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

The motion of the liquid inside the drop is assumed to be potential. This means that the field of the liquid velocities $\mathbf{V}(\mathbf{r}, t)$ is wholly defined by the potential function $\psi(\mathbf{r}, t)$; i.e., $\mathbf{V}(\mathbf{r}, t) = \nabla\psi(\mathbf{r}, t)$.

Under the above conditions, the evolution of the drop is described by the Laplace equations for velocity potential $\psi(\mathbf{r}, t)$ and electrostatic potential $\Phi(\mathbf{r}, t)$:

$$\Delta\psi(\mathbf{r}, t) = 0, \quad \Delta\Phi(\mathbf{r}, t) = 0, \tag{2), (3}$$

where Δ is the Laplacian, with the boundary conditions

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

$$r \longrightarrow \infty\colon \Phi(\mathbf{r}, t) \longrightarrow 0, \tag{5}$$

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

$$\Delta p - \frac{\partial\psi}{\partial t} - \frac{1}{2}(-\nabla\psi)^2 + \frac{1}{8\pi}(\nabla\Phi)^2 = \nabla\mathbf{n}, \tag{7}$$

$$\Phi(\mathbf{r}, t) = \Phi_g(t). \tag{8}$$

To make system (2)–(8) closed, we formulate additional conditions:

conservation of drop charge

$$-\frac{1}{4\pi}\oint_S (\mathbf{n}\nabla\Phi)dS = Q, \quad \text{where} \quad S = \begin{cases} r = 1 + \xi(\Theta, t) \\ 0 \le \Theta \le \pi \\ 0 \le \phi \le 2\pi, \end{cases} \tag{9}$$

and conservation of drop volume

$$\int_V r^2 dr \sin\Theta d\Theta d\phi = \frac{4}{3}\pi,$$

where

$$V = \begin{cases} 0 \le r \le 1 + \xi(\Theta, t) \\ 0 \le \Theta \le \pi \\ 0 \le \phi \le 2\pi. \end{cases} \tag{10}$$

The initial conditions for the problem will be set by taking the initial axisymmetric perturbation of the equilibrium spherical surface of the drop in the form

$$t = 0: \xi(\Theta) = \xi_0 + \varepsilon P_k(\mu) \quad (k \ge 2)$$
$$\mu \equiv \cos\Theta \tag{11}$$

and by equating the initial speed of the surface to zero:

$$t = 0: \frac{\partial \xi(\Theta)}{\partial t} = 0. \tag{12}$$

In (6)–(12), Δp is the difference between constant pressures inside and outside the drop in equilibrium, **n** is the unit normal vector to surface (1), $\Phi_g(t)$ is the position-independent potential on the drop surface, ε is the amplitude of the initial perturbation of the drop surface, $P_k(\mu)$ is the kth-order Legendre polynomial, and ξ_0 is a constant determined from (10) up to terms of second order of smallness:

$$\xi_0 = -\varepsilon^2 \frac{1}{(2k+1)} + O(\varepsilon^2). \tag{13}$$

(3) We will take advantage of the well-known multiple scale method to solve the problem up to quadratic terms in the small parameter ε [13]. To do this, we represent the required functions $\xi(\Theta, t)$, $\psi(\mathbf{r}, t)$, and $\Phi(\mathbf{r}, t)$ in the form of a series in ε and assume that the functions depend not merely on time but on various time scales $T_m \equiv \varepsilon^m t$; that is,

$$\xi(\Theta, t) = \sum_{m=0} \varepsilon^m, \xi^{(m)}(\Theta, T_0, T_1, T_2, \ldots),$$
$$\psi(\mathbf{r}, t) = \sum_{m=0} \varepsilon^m, \psi^{(m)}(\Theta, T_0, T_1, T_2, \ldots), \tag{14}$$
$$\Phi(\mathbf{r}, t) = \sum_{m=0} \varepsilon^m, \Phi^{(m)}(\Theta, T_0, T_1, T_2, \ldots).$$

The time differentiation is performed with respect to each time scale by the rule

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial T_0} + \varepsilon\frac{\partial}{\partial T_1} + \varepsilon^2\frac{\partial}{\partial T_2} + O(\varepsilon^2).$$

Substituting (14) into boundary problem (2)–(10) and equating terms of the same order of smallness to each other in each of the equations, we easily obtain a set of boundary problems from which one can successively determine the unknown functions $\xi^{(m)}$, $\psi^{(m)}$, and $\Phi^{(m)}$, where $m = 1, 2, 3, \ldots$. Since Eqs. (2)–(5) are linear, each of the functions $\psi^{(m)}$ and $\Phi^{(m)}$ from (14) must obey them. Therefore, these functions, as solutions to (2)–(5), may be given by

$$\psi^{(m)}(r, \Theta, T_0, T_1, T_2, \ldots) = \sum_{n=1}^{\infty} D_n^{(m)}(T_0, T_1, T_2, \ldots) r^n P_n(\mu),$$
$$\Phi^{(m)}(r, \Theta, T_0, T_1, T_2, \ldots) = \sum_{n=0}^{\infty} F_n^{(m)}(T_0, T_1, T_2, \ldots) r^{-(n+1)} P_n(\mu). \tag{15}$$

Successive corrections $\xi^{(m)}$ to the expression for drop shape will be sought as a series in Legendre polynomials:

$$\xi^{(m)}(r, \Theta, T_0, T_1, T_2, \ldots) = \sum_{n=1}^{\infty} M_n^{(m)}(T_0, T_1, T_2, \ldots) P_n(\mu). \tag{16}$$

(4) Up to the zeroth order in ε, the solution for the equilibrium system is

$$\xi^{(0)}(r, \Theta, T_0, T_1, T_2, \ldots) = 0;$$
$$\psi^{(0)}(r, \Theta, T_0, T_1, T_2, \ldots) = 0; \tag{17}$$
$$\Phi^{(0)}(r, \Theta, T_0, T_1, T_2, \ldots) = \frac{Q}{r},$$

as follows from (2)–(10).

(5) Up to the first order in ε ($m = 1$), boundary conditions (6)–(10) for solutions (15) and (16) to determine $D_n^{(m)}$, $F_n^{(m)}$ and $M_n^{(m)}$ are transformed to

$$r = 1: \frac{\partial \xi^{(1)}}{\partial T_0} = \frac{\partial \psi^{(1)}}{\partial r},$$
$$-\frac{\partial \psi^{(1)}}{\partial T_0} + \frac{1}{4\pi}\frac{d\Phi^{(0)}}{dr}\left(\frac{\partial \Phi^{(1)}}{\partial r}\right.$$

$$+\frac{d\Phi^{(0)}}{\partial r^2}\xi^{(1)}\Big) = -(2+\Delta_\Omega)\xi^{(1)},$$
$$\Phi^{(1)} + \frac{d\Phi^{(0)}}{dr} - \xi^{(1)} = \Phi_g^{(1)}, \quad (18)$$
$$\int_0^\pi \left[\frac{\partial\Phi^{(1)}}{\partial r} + \left(\frac{d^2\Phi^{(0)}}{dr^2} + 2\frac{d\Phi^{(0)}}{dr}\right)\xi^{(1)}\right]d\mu = 0,$$
$$\int_0^\pi \xi^{(1)} d\mu = 0, \quad \Delta_\Omega = \frac{\partial}{\partial\mu}\left((1-\mu^2)\frac{\partial}{\partial\mu}\right).$$

From (18), we easily obtain first-order corrections to the coefficients of series (15) and (16):

$$M_0^{(1)}(T_0, T_1, T_2, \ldots) = 0,$$
$$M_0^{(1)}(T_0, T_1, T_2, \ldots) = A_n(T_1, T_2, \ldots)\exp(i\omega_n T_0) + \text{c.c.} \quad (19)$$
$$(n \geq 1),$$
$$D_n^{(1)}(T_0, T_1, T_2, \ldots) = \frac{1}{n}\frac{\partial M_n^{(1)}(T_0, T_1, T_2)}{\partial T_0} \quad (n \geq 1),$$
$$F_n^{(1)}(T_0, T_1, T_2, \ldots) = QM_n^{(1)}(T_0, T_1, T_2, \ldots) \quad (20)$$
$$(n \geq 0),$$
$$A_n(T_1, T_2) \equiv a_n(T_1, T_2, \ldots)\exp[ib_n(T_1, T_2, \ldots)],$$
$$\Phi_g^{(1)} = 0, \quad \omega_n^2 \equiv n(n-1)[(n+2) - W], \quad W \equiv \frac{Q^2}{4\pi}, \quad (21)$$

where c.c. means complex conjugate terms. The functions a_n and b_n are determined up to higher orders of smallness. To complete the first-order problem, a_n and b_n are set to be constants and derived from initial conditions (11) and (12). It is of interest to find the error of the first-order approximation of the solution and also the time intervals where such an approximation is valid. It is clear that

$$a_n(T_1, T_2, \ldots) \approx \hat{a}_n + O(\varepsilon t);$$
$$b_n(T_1, T_2, \ldots) \approx \hat{b}_n + O(\varepsilon t), \quad (22)$$

where $\hat{a}_n$ and $\hat{b}_n$ are constants.

From (14), we have (in the linear approximation in ε)

$$\xi(\Theta, t) \approx \varepsilon\xi^{(1)}[\Theta, T_0, a_n(T_1, \ldots), b_n(T_1, \ldots)] + O(\varepsilon^2)$$

or, in view of (22),

$$\xi(\Theta, t) \approx \varepsilon\xi^{(1)}(\Theta, t, \hat{a}_n, \hat{b}_n) + \varepsilon O(\varepsilon t). \quad (23)$$

The error in this expansion is of the order of the first term if $t \approx O(\varepsilon^{-1})$. The expansion is not valid for $t \geq O(\varepsilon^{-1})$. Thus, (23) is applicable in the time interval $t \leq O(1)$; the error in this case is of the order of $\sim\varepsilon^2$. However, (23) can also be used in the time interval $t \leq O(\varepsilon^{-1})$ to analyze surface motion if the first-order solution is comparable to the initial perturbation. A more detailed estimate of the applicability of (23) can be obtained in the next (second) order of smallness in ε.

Substituting (23) into initial conditions (11)–(13) with regard for (16) and (19)–(22) and equating the terms of the same order of smallness to each other, we easily find

$$\hat{a}_n = \frac{1}{2}\delta_{nk}, \quad b_n = 0 \ (n \geq 1),$$

where δ_{nk} is the Kronecker symbol ($\delta_{nk} = 1$, if $n = k$, and 0 if $n \neq k$).

Eventually, the function describing the surface evolution in the linear approximation in ε has the form

$$\xi(\Theta, t) = \varepsilon\cos(\omega_k t)P_k(\mu) + O(\varepsilon^2). \quad (24)$$

It follows from this solution that, in the first-order approximation in the perturbation amplitude ε, the drop surface executes harmonic vibrations around the equilibrium sphere that are associated with the kth (initially excited) mode.

In the same approximation, the velocity and electrostatic potentials have the form

$$\psi(\mathbf{r}, t) = -\varepsilon\frac{\omega_k}{k}\sin(\omega_k t)P_k(\mu) + O(\varepsilon^2),$$
$$\Phi(\mathbf{r}, t) = \frac{Q}{r} + \varepsilon\cos(\omega_k t)\frac{Q}{r^3}P_k(\mu) + O(\varepsilon^2). \quad (25)$$

(6) To find the functions $\xi^{(2)}(\Theta, t)$, $\psi^{(2)}(\mathbf{r}, t)$, and $\Phi^{(2)}(\mathbf{r}, t)$ (that is, second-order corrections to the solutions), we first derive a set of equations that is obtained from (6)–(10) by equating terms proportional to $\sim\varepsilon^2$:

$$r = 1: \frac{\partial\xi^{(2)}}{\partial T_0} + \frac{\partial\xi^{(1)}}{\partial T_1}$$
$$= \frac{\partial\psi^{(2)}}{\partial r} + \frac{\partial^2\Psi^{(1)}}{\partial r^2}\xi^{(1)} - \frac{\partial\xi^{(1)}}{\partial\Theta}\frac{\partial\psi^{(1)}}{\partial\Theta},$$
$$-\frac{\partial\psi^{(2)}}{\partial T_0} - \frac{\partial\psi^{(1)}}{\partial T_1} - \frac{\partial^2\psi^{(1)}}{\partial r\partial T_0}\xi^{(1)}$$
$$-\frac{1}{2}\left[\left(\frac{\partial\psi^{(1)}}{\partial r}\right)^2 + \left(\frac{\partial\psi^{(1)}}{\partial\Theta}\right)^2\right] + \frac{1}{8\pi}\left[2\frac{\partial\Phi^{(0)}}{\partial r}\frac{\partial\Phi^{(2)}}{\partial r}\right.$$
$$\left. + \left(\frac{\partial\Phi^{(1)}}{\partial r}\right)^2 + \left(\frac{\partial\Phi^{(1)}}{\partial\Theta}\right)^2 + 2\frac{d\Phi^{(0)}}{dr}\frac{d^2\Phi^{(0)}}{dr^2}\xi^{(2)}\right.$$
$$\left. + 2\left(\frac{d\Phi^{(0)}}{dr}\frac{\partial^2\Phi^{(1)}}{\partial r^2} + \frac{d^2\Phi^{(0)}}{dr^2}\frac{\partial\Phi^{(1)}}{\partial r}\right)\xi^{(1)}\right.$$

$$+\left(\frac{d\Phi^{(0)}}{dr}\frac{d^3\Phi^{(0)}}{dr^3}+\left(\frac{d^2\Phi^{(0)}}{dr^2}\right)\right)^2(\xi^{(1)})^2\Bigg] \qquad (26)$$

$$+(2+\Delta_\Omega)\xi^{(2)}-2\xi^{(1)}(1+\Delta_\Omega)\xi^{(1)}=0,$$

$$\Phi^{(2)}+\left[\frac{d\Phi^{(0)}}{dr}\xi^{(2)}+\frac{\partial\Phi^{(1)}}{\partial r}\xi^{(1)}+\frac{1}{2}\frac{d^2\Phi^{(0)}}{dr^2}(\xi^{(1)})^2\right]=\Phi_g^{(2)},$$

$$\int_0^\pi\Bigg[\frac{d\Phi^{(2)}}{dr}\left(\frac{\partial\Phi^{(1)}}{\partial r^2}+2\frac{\partial\Phi^{(1)}}{\partial r}\right)\xi^{(1)}+\left(\frac{\partial\Phi^{(0)}}{\partial r^2}+2\frac{\partial\Phi^{(0)}}{\partial r}\right)$$

$$\times\xi^{(2)}+\left(\frac{1}{2}\frac{d^3\Phi^{(0)}}{dr^3}+2\frac{d\Phi^{(0)}}{dr^2}+\frac{d\Phi^{(0)}}{dr}\right)$$

$$\times(\xi^{(1)})^2-\frac{\partial\Phi^{(1)}}{\partial\Theta}\frac{\partial\xi^{(1)}}{\partial\Theta}\Bigg]d\mu=0,$$

$$\int_0^\pi[\xi^{(2)}(\Theta,t)+(\xi^{(1)}(\Theta,t))^2]d\mu=0.$$

Now, substituting (15) and (16) ($m = 2$), along with solutions (17) and (19)–(21), into boundary conditions (26), one obtains differential equations for the coefficients $M_n^{(2)}(T_0, T_1, T_2, \ldots)$. The solutions to this system do not have secular terms if

$$\frac{\partial a_n}{\partial T_1}=0,\quad \frac{\partial b_n}{\partial T_1}=0.$$

Hence, a_n and b_n do not depend on time T_1. Their dependence on the "slower" time scales T_2, T_3, etc. can be found in higher orders of approximation. Finally, the solution to the equations derived from (26) has the form

$$M_0^{(2)}(T_0,T_2,\ldots)=-\sum_{n=1}^{\infty}\frac{1}{(2n+1)}\{A_n(T_2,\ldots)$$

$$\times\bar{A}_n(T_2,\ldots)+[A_n(T_2,\ldots)]^2\exp(i2\omega_nT_0)+\text{c.c.}\},$$

$$M_0^{(2)}(T_0,T_2,\ldots)=\left\{\left[\frac{c_n(T_2,\ldots)}{a_n(T_2,\ldots)}+id_n(T_2,\ldots)\right]\right.$$

$$\left.\times A_n(T_2,\ldots)\exp(i\omega_nT_0)+\text{c.c.}\right\}+N_n(T_0,T_2,\ldots),$$

$$N_n^{(2)}(T_0,T_2,\ldots)\equiv\sum_{n=1}^{\infty}\sum_{l=1}^{\infty}\{\lambda_{mln}^{(+)}A_m(T_2,\ldots)A_l(T_2,\ldots)$$

$$\times\exp[i(\omega_m+\omega_l)T_0]+\lambda_{mln}^{(-)}A_m(T_2,\ldots)\bar{A}_l(T_2,\ldots)$$

$$\times\exp[i(\omega_m-\omega_l)T_0]+\text{c.c.}\},$$

$$\lambda_{mln}^{(+)}=\frac{\gamma_{mln}\pm\omega_m\omega_l\eta_{mln}}{\omega_n^2-(\omega_m\pm\omega_l)^2}, \qquad (27)$$

$$\gamma_{mln}\equiv K_{mln}\Bigg[\omega_m^2(n-m+1)+2n[l(l+1)-1]+[l(m+1)$$

$$-m(2m-2n+7)+3]n\frac{W}{2}\Bigg]+\alpha_{mln}\left[\frac{1}{m}\omega_m^2+n\frac{W}{2}\right],$$

$$\eta_{mln}\equiv K_{mln}\left(\frac{n}{2}-m+1\right)+\alpha_{mln}\frac{1}{m}\left(1+\frac{n}{2l}\right),$$

$$K_{mln}\equiv[C_{mln}^{000}]^2,\quad \alpha_{mln}\equiv-\sqrt{m(m+1)l(l+1)}C_{mln}^{000}C_{mln}^{-110},$$

$$C_{mln}^{000}\equiv\begin{cases}0 \text{ if } m+l+n=2g+1,\\ \text{where } g \text{ is an integer,}\\ \dfrac{(-1)^{g-n}\sqrt{2n+1}g!}{(g-m)!(g-l)!(g-n)!}\\ \times\left[\dfrac{(2g-2m)!(2g-2l)!(2g-2n)!}{(2g+1)!}\right]^{1/2}\\ \text{if } m+l+n=2g(g \text{ is an integer}),\end{cases}$$

$$C_{mln}^{-110}\equiv\sqrt{2n+1}n!$$

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

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

In the last series, the summation is performed over all integer z for which the expressions under the factorial sign are nonnegative. The bar over A_n in (27) means the complex conjugate. The Clebsch–Gordan coefficients C_{mln}^{000} and C_{mln}^{110} are other than zero only if their subscripts satisfy the relations [14]

$$|m-l|\le n\le(m+l),$$
$$m+l+n=2g\ (g \text{ is an integer}). \qquad (28)$$

The coefficients in series (15) for $\psi^{(m)}(\mathbf{r}, t)$ and $\Phi^{(m)}(\mathbf{r}, t)$ are related to solutions (19) and (27) as follows:

$$D_n^{(2)}(T_0,T_2,\ldots)=\frac{1}{n}\left\{\frac{\partial M_n^{(2)}(T_0,T_2,\ldots)}{\partial T_0}\right.$$

$$\left.\times\sum_{m=1}^{\infty}\sum_{l=1}^{\infty}[m(m-1)K_{mln}-\alpha_{mln}]\frac{1}{m}\right.$$

$$\times \frac{\partial M_m^{(1)}(T_0, T_2, \ldots)}{\partial T_0} M_l^{(1)}(T_0, T_2, \ldots) \Big\},$$

$$F_0^{(2)}(T_0, T_2, \ldots) = 0,$$

$$F_n^{(2)}(T_0, T_2, \ldots) = Q\{M_n^{(2)}(T_0, T_2, \ldots) + \sum_{m=1}^{\infty}\sum_{l=1}^{\infty} mK_{mln}(T_0, T_2, \ldots) \times M_n^{(1)}(T_0, T_2, \ldots) M_l^{(1)}(T_0, T_2, \ldots)\}.$$

The coefficients $c_n(T_2, \ldots)$ and $d_n(T_2, \ldots)$ in (27) are unknown functions of time. Like the functions a_n and b_n, they do not depend on the time scales T_0 and T_1. By analogy with the linear approximation, we complete the study of the problem in the second-order approximation in ε by assuming that the unknown functions c_n, d_n, a_n, and b_n are constants determined by initial conditions (11)–(13). Then, series (14) for drop surface perturbation takes the form

$$\xi(\Theta, t) \approx \varepsilon\xi^{(1)}(\Theta, T_0, \hat{a}_n, \hat{b}_n) + \varepsilon^2\xi^{(2)}(\Theta, T_0, \hat{a}_n, \hat{b}_n, \hat{c}_n, \hat{d}_n) + \varepsilon O(\varepsilon^2 t), \quad (29)$$

where $\hat{a}_n$, $\hat{b}_n$, $\hat{c}_n$, and $\hat{d}_n$ are constants.

Approximation (29) is valid in the time interval $t \le O(1)$ with an error $\sim\varepsilon^3$. In the time interval $O(1) < t \le O(\varepsilon^{-1})$, the error is comparable to the second term in (29), which is of second order of smallness. Therefore, only the first term, associated with the linear approximation, remains valid. Thus, approximate linear solutions (24) and (25) are strictly applicable (uniformly suitable) in the time interval $t \le O(\varepsilon^{-1})$. The entire expansion (29) can be used in the time interval $t \le O(\varepsilon^{-1})$ provided that the term quadratic in ε is a small correction to the linear term.

The characteristic time scale for the dimensionless variables adopted is $t_* = (R^3\rho/\sigma)^{1/2}$. It is about 0.004 s for a water drop of radius 1 mm and about 0.12 s when $R = 1$ cm. Hence, (29) adequately describes time evolution of the water drop shape at times $t \le 0.01$ s for $R =$ 1 mm and $t \le 0.3$ s for $R = 1$ cm. If the perturbation amplitude ε is about 0.1, solution (24) remains valid up to times one order of magnitude greater than those given above.

Substituting (29) in view of (16) into initial conditions (11)–(13) and equating the terms of the same order of smallness, we obtain the following relationships:

$$t = 0: M_n^{(1)} = \delta_{nk}, \quad M_n^{(2)} = -\frac{\delta_{n0}}{2k+1},$$

$$\frac{\partial M_n^{(1)}}{\partial t} = 0, \quad \frac{\partial M_n^{(2)}}{\partial t} = 0 \ (n = 0, 1, 2, \ldots).$$

From these expressions and using (19), (20), and (27), we determine the constants $\hat{a}_n$, $\hat{b}_n$, $\hat{c}_n$, and $\hat{d}_n$:

$$\hat{a}_n = \frac{1}{2}\delta_{nk}, \quad \hat{b}_n = 0,$$

$$\hat{c}_n = -\frac{1}{2}N_n(t=0), \quad \hat{d}_n = 0 \ (n = 0, 1, 2, \ldots).$$

Then, the coefficients $M_n^{(m)}(t)$ from (19), (20), and (27) take the final form

$$M_n^{(1)}(t) = \delta_{nk}\cos(\omega_n t),$$
$$M_0^{(2)}(t) = -\frac{1}{2(2k+1)}[1 + \cos(2\omega_k t)],$$
$$M_n^{(2)}(t) = -N_n(0)\cos(\omega_n t) + N_n(t) \ (n \ge 1), \quad (30)$$
$$N_n(t) = \frac{1}{2}(\lambda_{kkn}^{(-)} + \lambda_{kkn}^{(+)}\cos(2\omega_k t)).$$

It can be seen that the coefficients $M_n^{(2)}(t)$ are proportional to the quantities $\lambda_{kkn}^{(\pm)}$. Due to (27), the latter are proportional to the Clebsch–Gordan coefficients C_{kkn}^{000} and C_{kkn}^{-110} and, according to (28), are nonzero only if $n = 2j$, where $j = 0, \ldots, k$.

Substituting (30) into (29), we find that the evolution of the drop surface is described up to terms of second order of smallness (in the time interval $t \le O(\varepsilon^{-1})$ by the function

$$\xi(\Theta, t) \approx \varepsilon\cos(\omega_k t)P_k(\mu) - \varepsilon^2\frac{1}{2}\Big\{\frac{1}{(2k+1)} \times [1 + \cos(2\omega_k t)] + \sum_{j=1}^{k}[(\lambda_{k,k,2j}^{(-)} + \lambda_{k,k,2j}^{(+)})\cos(\omega_{2j} t) - (\lambda_{k,k,2j}^{(-)} + \lambda_{k,k,2j}^{(+)}\cos(2\omega_k t))]P_{2j}(\mu)\Big\} + O(\varepsilon^3 t). \quad (31)$$

(7) It follows from (31) that the initial perturbation of any even or odd kth single mode of capillary vibrations results in the excitation (in the second order of smallness) of only even modes with numbers from the interval $[0, 2k]$. This is illustrated in Figs. 1a–1d. These figures show the time dependences of the amplitudes $M_n^{(2)}$ of various second-order modes of charged drop

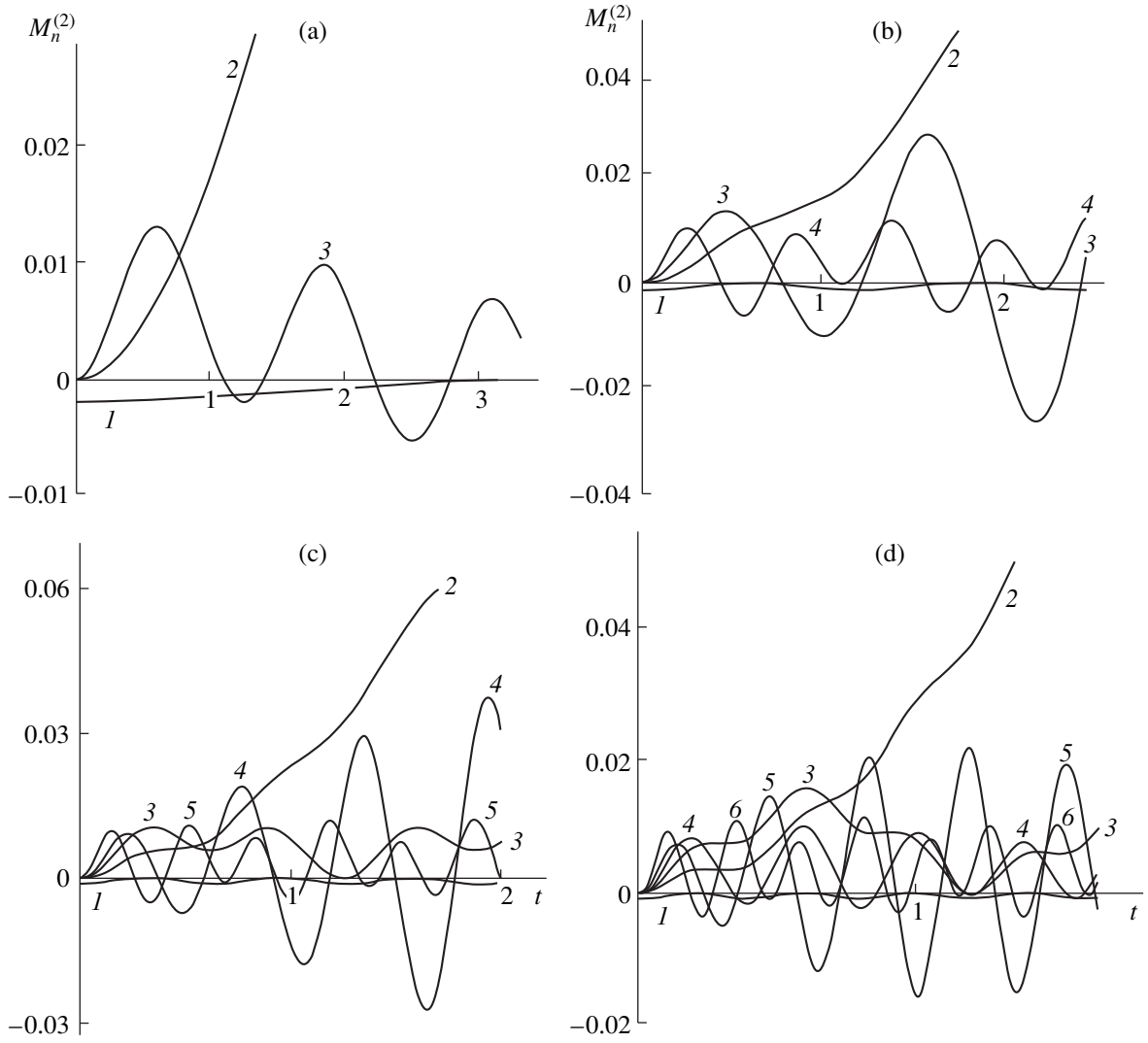

Fig. 1. Dimensionless time dependence of the dimensionless amplitudes $M_n^{(2)}$ of the (*1*) zeroth, (*2*) second (principal), (*3*) fourth, (*4*) sixth, (*5*) eighth, and (*6*) tenth modes of charged drop capillary vibrations at $W = 3.9$. The modes are excited as a result of second-order coupling. The initial deformation of the equilibrium spherical shape is due to the perturbation of the (a) second, (b) third, (c) fourth, and (d) fifth modes. The dimensionless amplitude of the perturbation is ε.

capillary vibrations excited as a result of coupling [see (31)]. The initial perturbation of the equilibrium spherical shape is specified by the excitation of modes *2–5* (Figs. 1a–1d, respectively) at $W = 3.9$. (A spherical drop becomes unstable when the parameter W attains the critical value $W_* = 4$.) It is seen that a drop with a charge somewhat smaller than critical may exhibit its charge instability through a fast increase in the principal mode amplitude ($n = 2$). Such behavior is independent of the initial perturbation of the equilibrium spherical shape. This result is in conflict with the predictions from the linear theory but is in qualitative agreement with [8], where nonlinear vibrations of a charged drop were numerically evaluated. When the initial perturbation is associated with the fifth mode, the time dependences of the excitation amplitudes are also in quantitative agreement with data from [8] in the second order of smallness.

It follows from Fig. 1 that the rate of increase in the principal mode amplitude grows with the number of the mode that specifies the initial deformation. As the number of this mode increases, so does the total number of modes of capillary vibrations due to coupling. Odd modes are excited only in the third-order approximation as a result of coupling between the excited odd kth mode and second-order even modes produced by the kth one.

Figure 2 illustrates drop shapes calculated using (1) and (31) at various dimensionless time instants with W values close to critical. In this case, the initial deformation of the equilibrium spherical shape is specified by the perturbation of the third (Fig. 2a) and fourth

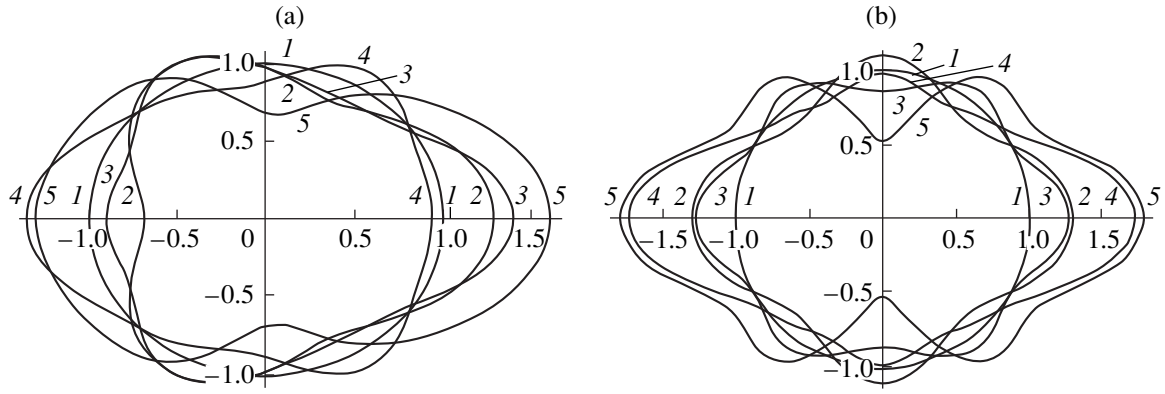

Fig. 2. Shapes taken by the drop at various time instants $t\varepsilon$ when the initial deformation $\varepsilon = 0.3$ of the equilibrium spherical shape is due to the virtual perturbation of various modes at $W = 3.8$. In both figures, curves *1* depict the equilibrium spherical shape and curves *2*, the shape at the initial time instant [the spherical shape deformed by a perturbation of type $\varepsilon P_n(\mu)$]. (a) $n = 3$ and $t\varepsilon =$ (*3*) 0.065, (*4*) 0.26, and (*5*) 0.585; (b) $n = 4$ and $t\varepsilon =$ (*3*) 0.24, (*4*) 0.36, and (*5*) 0.57.

(Fig. 2b) modes. The axis of symmetry is horizontal. It should be noted that, according to perturbation theory, formula (31) is uniformly suitable when $t < \varepsilon^{-1}$. It follows from the figures that the restriction on time t is actually even more stringent. Curves *4* in all the figures were drawn at the boundary of the domain of uniform expansion applicability. This follows from a comparison of the resulting deviation of the drop shape (curves *4*) from that at the initial time instant (curves *2*). Moreover, the volume of the drop associated with curves *4* is apparently different from the initial one. Yet, it is seen that, when the initial deformation of the equilibrium shape is specified by even Legendre polynomials, its time evolution is also described by even Legendre polynomials and the shape remains symmetric about the origin. For large enough t (at the boundary of the domain where the solution is uniformly suitable), the drop tends to break down into two equal parts. If the initial deformation is due to odd Legendre polynomials, the drop shape at each subsequent time instant is asymmetric about the origin even though only even modes are excited by second-order mode coupling. At large t, such drops tend to break down asymmetrically.

It is clear from physical reasoning that viscosity, which has not been taken into account in our study, would lead to damping of all the modes. However, the damping decrement for higher order modes is larger than for lower order ones. If a time interval is large enough, the amplitude of an initially excited higher order odd mode may go to zero faster than the amplitude of lower even modes excited by it. Then, further vibrations of the drop and its possible breakdown into two parts will be symmetric.

It follows from the above discussion and from the figures that the amplitude of the principal mode of capillary vibrations increases fastest, irrespective of the form of the initial deformation. The calculations used in our study are valid as long as the amplitudes of second-order modes are smaller than the amplitude of the initial perturbation. Therefore, an increase in the principal mode amplitude to a value of the order of ε means an extension of the drop into a spheroid with the eccentricity squared, $e^2 \approx 3\varepsilon - 5.25\varepsilon^2$ [15]. Even small values of $\varepsilon \sim 0.1$ will lead to a considerable extension of the drop and, according to [16], to a decrease in the W value critical for the onset of charge instability of the drop. In the approximation linear in e^2, this critical value for a spheroidal drop has the form

$$W_*(e^2) = 4(1 - 2e^2/7) \approx 4[1 - 2(3\varepsilon - 4.25\varepsilon^2)/7]. \quad (32)$$

Thus, if the Rayleigh parameter W is close to the critical value, the drop may become unstable. Let the Rayleigh parameter of the drop $W = W_+$ be slightly smaller than its critical value $W_* = 4$. Then, from (32), one can find the current dimensionless amplitude a_2 of the principal mode when the drop becomes unstable:

$$a_2 \approx \frac{1}{3.5}\{1 - [1 - 8.17(1 - 0.25W_+)]^{1/2}\}.$$

If, for example, $W_+ = 3.6$, the drop becomes unstable when the dimensionless amplitude of the principal mode reaches a value of $a_2 \approx 0.16$. Being unstable, the drop loses some amount of charge by emitting many fine, heavily charged daughter droplets [1, 17].

(8) It is seen from Fig. 1 that the amplitudes of all second-order modes oscillate around some midlines above the abscissa. This means that drop vibrations occur not around a sphere but around a lower symmetry body. Time-independent terms in (31) specify this body:

$$r(\Theta) \approx 1 - \varepsilon^2\frac{1}{2}\left\{\frac{1}{(2k+1)} - \sum_{j=1}^{k}\lambda^{(-)}_{k,k,2j}P_{2j}(\mu)\right\} + O(\varepsilon^3 t), \quad (33)$$

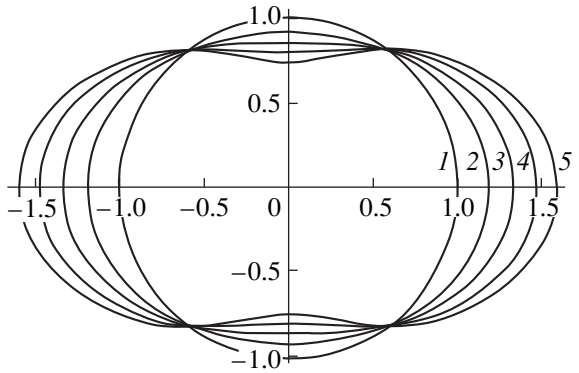

Fig. 3. Bodies around which the drop nonlinearly vibrates at $W = 3$ when the initial deformation of amplitude $\varepsilon = 0.3$ is due to the virtual excitation of various modes. Curve *1* is the equilibrium spherical shape, and curves *2–5* correspond to initially excited modes with numbers 2–5.

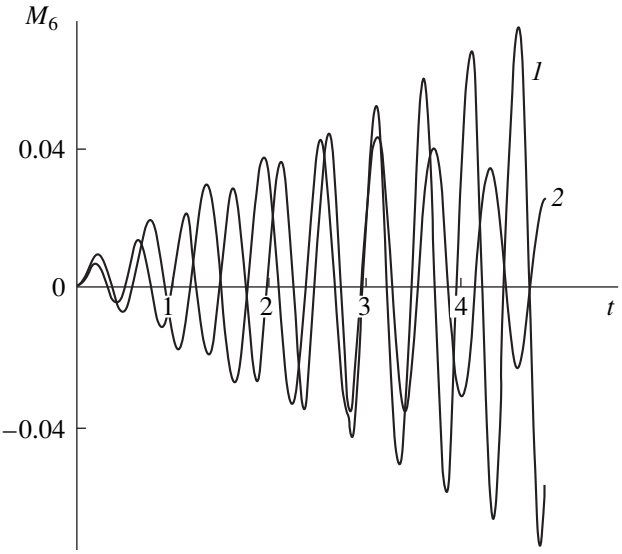

Fig. 4. Time dependence of the dimensionless amplitude of the 6th mode when the 4th mode is initially excited with $\varepsilon = 0.1$ (*1*) under resonance and (*2*) far from it.

where k is the number of an initially excited mode.

It is easy to check that the body around which second-order vibrations occur is composed only of even modes independently of the type of first-order modes excited. Hence, this body is a prolate spheroid, not a sphere as predicted in the Rayleigh linear analysis [1]. It is interesting that the parameters of such a spheroid depend on the drop charge, i.e., on the parameter W, through the coefficients $\lambda^{\pm}_{k,\,k,\,2j}$. The shapes of the bodies discussed above and calculated with (33) for various k are shown in Fig. 3. It is seen that, as the number of an initially excited mode increases, the stretching of the body around which vibrations take place grows.

(9) The solutions obtained also describe the resonance coupling between individual modes of drop vibrations that was analyzed in detail in [3]. Such coupling arises when second-order modes oscillate with frequencies satisfying the condition $\omega_m^2 = j^2\omega_n^2$ for some value of charge Q, where j is an integer and $m \neq n$. As a result, the amplitude of one of the modes increases in time periodically and infinitely (within our approximation). In [3], the resonance between the initially excited fourth and sixth modes at $W = 2.67$, i.e., when $\omega_0^2 = 4\omega_4^2$, was studied. The same resonance takes place in our study when only the fourth mode is excited. The time dependence of the amplitude of the sixth mode excited by the initial generation of the fourth mode for $W = 2.67$ and small time intervals is similar to curve *4* in Fig. 1c. However, Fig. 1c demonstrates the results of calculation at $W = 3.9$; here, the sixth mode amplitude increases because of energy transfer between modes coupled in the second order. In a time interval larger than that used in Fig. 1c, the amplitude of the sixth mode decreases with time. The time dependence of the sixth mode amplitude is shown in Fig. 4 for comparison. The graphs are constructed for resonance conditions, $W \approx 2.667$ (curve *1*), and far from resonance, $W = 3.9$ (curve *2*), when the fourth mode was initially excited. The time interval in Fig. 4 is larger than in Fig. 1c. The run of the curves is obviously different. Resonance coupling takes place when terms with the factors $\sim(\omega_m^2 - j^2\omega_n^2)^{-1}$ appear in (31).

CONCLUSION

The instability of a drop bearing a charge that is slightly smaller than critical was studied in the quadratic approximation with respect to the amplitude of the initial deformation. Due to coupling between various modes of capillary vibrations, this instability can be initiated by virtual excitation of not only the principal mode but of any mode, no matter whether it is odd or even. The drop exhibits nonlinear vibrations around a prolate spheroid, not a sphere as follows from a linear analysis. In contrast to aperiodic instability of such a drop, which is a consequence of energy transfer between excited adjacent modes, vibrational instability may result from resonance coupling of nonadjacent modes.

REFERENCES

1. A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 3 (1994).
2. J. A. Tsamopoulos and R. A. Brown, J. Fluid Mech. **127**, 519 (1983).
3. J. A. Tsamopoulos and R. A. Brown, J. Fluid Mech. **147**, 373 (1984).
4. J. A. Tsamopoulos, T. R. Akulas, and R. A. Brown, Proc. R. Soc. London, Ser. A **401**, 67 (1985).
5. R. Natarajan and R. A. Brown, Proc. R. Soc. London, Ser. A **410**, 209 (1987).

6. N. A. Pelekasis, J. A. Tsamopoulos, and G. D. Manolis, Phys. Fluids A **2** (8), 1328 (1990).
7. T. G. Wang, A. V. Anilkumar, and C. P. Lee, J. Fluid Mech. **308**, 1 (1996).
8. Z. Feng, J. Fluid Mech. **333**, 1 (1997).
9. D. F. Belonozhko and A. I. Grigor'ev, Pis'ma Zh. Tekh. Fiz. **25** (15), 41 (1999) [Tech. Phys. Lett. **25**, 610 (1999)].
10. O. A. Basaran and L. E. Scriven, Phys. Fluids A **1** (5), 795 (1989).
11. A. I. Grigor'ev and S. O. Shiryaeva, Zh. Tekh. Fiz. **69** (7), 10 (1999) [Tech. Phys. **44**, 745 (1999)].
12. S. I. Shchukin and A. I. Grigor'ev, Zh. Tekh. Fiz. **68** (11), 48 (1998) [Tech. Phys. **43**, 1314 (1998)].
13. A.-H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973; Mir, Moscow, 1976).
14. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Scientific, Singapore, 1988).
15. S. O. Shiryaeva, A. I. Grigor'ev, and I. D. Grigor'eva, Zh. Tekh. Fiz. **65** (9), 39 (1995) [Tech. Phys. **40**, 885 (1995)].
16. S. O. Shiryaeva and A. I. Grigor'ev, Zh. Tekh. Fiz. **66** (9), 12 (1996) [Tech. Phys. **41**, 865 (1996)].
17. A. I. Grigor'ev and S. O. Shiryaeva, Zh. Tekh. Fiz. **61** (3), 19 (1991) [Sov. Phys. Tech. Phys. **36**, 258 (1991)].

Translated by V. Gursky

Technical Physics, Vol. 46, No. 2, 2001, pp. 167–173. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 36–42.

Influence of the Ion Charge Sign on the Stimulation of Plasmochemical Etching of Silicon

S. N. Pavlov

Scientific Center Institute for Nuclear Research, National Academy of Sciences of Ukraine, Kiev, 03028 Ukraine
e-mail: interdep@kinr.kiev.ua

Received February 1, 2000

Abstract—The influence of ions with different charge signs on the stimulation of silicon etching under plasma conditions is studied. Fluorine radicals are produced in a glow discharge with a nonuniform pressure. A beam of positive or negative ions is created using a Penning ion source. The flow of fluorine radicals and the ion beam are superposed on a silicon surface placed in a high vacuum. Positive ions may be converted into fast neutrals via resonance charge exchange in the parent gas. It is shown that fast neutrals have the highest catalytic effect. The catalytic effect of positive ions is about two times less. Negative ions occupy the intermediate position. For the first time, it is found that some kinds of ions (e.g., molecular oxygen) do not accelerate, but rather decelerate the etching process; i.e., they behave as inhibitors. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The phenomenon of ion-stimulated etching of silicon was discovered more than 20 years ago [1, 2] and, at present, is widely used in microelectronics [3, 4]. The phenomenon, in essence, is as follows. When a silicon surface is bombarded by ions, e.g., argon ions with energies of 500–1000 eV, the solid surface is sputtered. Under the above conditions, the sputtering yield is approximately unity. In microelectronics, this process is referred to as ion-beam etching and is characterized by the rate $V_{ph} = dh/dt$, where h is the plate thickness. Another limiting case is etching by halogen radicals (F*, Cl*, Br*, or I*, fluorine being most often used). This is a heterogeneous reaction of the gas–solid type. Silicon is removed as a result of a chemical reaction proceeding according to the generalized formula $Si + 4F \longrightarrow SiF_4$. Silicon tetrafluoride is a gas that spontaneously leaves the surface of the treated material. In this case, the plasma is only used to obtain fluorine radicals from the molecules of more stable initial gases (CF_4, SF_6, etc.). Etching proceeds at a rate of V_{ch}, which depends on the halogen concentration and the solid temperature. If the silicon surface is exposed to the simultaneous action of an ion flow and fluorine radicals, etching proceeds at a rate of V_i, which may be several times higher than the sum $V_{ph} + V_{ch}$. The process is synergistic in character [5, 6] and, hence, is of general scientific interest. It can be observed not only in silicon–halogen systems, but also in a number of gasification reactions, such as carbon–oxygen or tungsten–fluorine reactions. Keeping in mind the general character of the results, we nevertheless will focus on the Si–F system, because it is the most widely used in practice and, hence, the best studied.

The process of plasmochemical etching can be divided into the following main stages: (i) delivery of the working gas molecules to the plasma discharge region, (ii) transformation of the working gas molecules into chemically active particles via dissociation and ionization, (iii) delivery of the chemically active particles to the etched surface, (iv) physical and chemical adsorption of the chemically active particles on the surface, (v) the chemical reaction itself, (vi) desorption of the reaction products from the surface, and (vii) removal of the reaction products from the plasma region and vacuum chamber. This comprehensive classification is given according to [4]. Stages (i), (iii), and (vii) are important for designing plasma chemical reactors but do not directly affect ion stimulation itself. Stage (ii) may have only indirect influence via the chemical composition of the reacting system. It worth noting that mass spectroscopy and optical studies show [7, 9] that all of the neutral fragments of the molecules of the initial gas (as a rule, CF_4 or SF_6), as well as their positive and negative ions, have been recorded in an RF discharge plasma. However, more than 75% of dissociating CF_4 molecules decompose into CF_3^* and F* radicals [10].

When analyzing the mechanism for ion stimulation, most authors restrict themselves to considering stages (iv)–(vi). Let us consider the possible consequences of ion bombardment of a silicon surface. In stage (iv), ions may (a) enhance nondissociative chemisorption due to the breaking of Si–Si bonds in the crystal lattice, which results in the formation of free valence silicon bonds, and (b) cause the dissociation of complex radicals (like CF_3^* or SF_5^*) absorbed on the surface with subsequent chemisorption of their fragments. In stage (v), the reaction rate may increase due to surface loosening, which

leads to both the breaking of Si–Si bonds and the formation of channels in the treated material, along which chemical reagents can penetrate into deeper layers. In stage (vi), ion bombardment may cause forced desorption of (a) the intermediate reaction products (usually volatile SiF_2), (b) the final product SiF_4, and (c) the nonvolatile products.

There are arguments in favor of each of the effects listed above. The problem is that the results of each experiment are usually treated as supporting one of the mechanisms, but as discarding the other ones. For example, the authors of [5, 11] conclude that dissociative chemisorption plays a major role, assuming that surface loosening is of minor importance. The results of [12–14] contradict this opinion. In 1978, Mauer *et al*. [15] showed that ion bombardment may cause induced desorption of the intermediate reaction component SiF_2. This concept has been well developed in recent years. The results of ion-stimulated etching [16, 17] show that surface bombardment by heavy particles may result in the complete rearrangement of chemical bonds in the SF_6–Si system, which makes the situation even more intricate.

Thus, a unified opinion on the mechanism for ion stimulation of plasmochemical etching of silicon is still lacking; therefore, the problem calls for further analysis. Taking into account the previous long-term and comprehensive investigations, one can expect that only new approaches may lead to considerable progress. An attempt at such an approach, as well as the first experimental results, is described below.

FORMULATION OF THE PROBLEM

The following fact is worthy of attention. Etching of silicon in halogen-containing plasma can be accelerated not only by ion flows, but also by electron flows. The difference is that ion stimulation occurs in almost any medium, whereas electron stimulation can occur only in some specific media. Thus, electron beams with densities of 2–4 mA/cm^2 increase the etching rate by a factor of four to six when CF_4 or CF_3Cl is used; however, the effect is not observed in XeF_2 and SF_6 plasmas [11, 18]. This means that it is expedient to investigate the influence of the ion charge sign on the stimulation of silicon etching. This problem is of especial interest, because negative ions are proposed to be used to improve the treatment of semiconductors (see, e.g., [19]).

Finally, when comparing positive and negative ions, it would also be reasonable to investigate the role of fast neutrals, which can be regarded as ions with zero charge.

The prototype of our experimental device was that used in [5]. Such a system allows one to imitate plasma conditions, which offers wide experimental possibilities, because in a real plasma, it is impossible to change only one of the parameters and hold the other parameters constant. The design and features of the device are described in the next section.

Sulfur hexafluoride and oxygen were used as working gases because they form a very compatible pair. To study the problem in question, it is necessary to produce beams of positive ions, negative ions, and fast neutrals of the same gas with approximately the same intensities. Fluorine is less appropriate for such purposes. Our investigations showed that in a Penning discharge, which is often used as an ion source, the intensity of the beam of negative fluorine ions is almost one order of magnitude higher than that of F^+ ions. Due to inevitable loss during charge exchange, we could not obtain fast neutrals in the necessary amounts. As will be shown below, oxygen is free of these drawbacks. On the other hand, the silicon etching rate is constant over a wide range of the O_2 relative concentration in a mixture of sulfur hexafluoride with oxygen, which considerably facilitates the interpretation of the results. Finally, the yield of Si atoms per incident ion is much higher for etching in SF_6 in comparison with that in CF_4, which facilitates performing the experiments.

As reference beams, we used beams of argon ions and fast argon atoms, because argon has been used in a great number of experiments and its properties are well studied.

EXPERIMENTAL SETUP

As was mentioned above, the prototype of our experimental device was that used in [5]. The idea of the experiment is to use two independent sources, one of which produces an ion flow and the other produces a flow of radicals. Both flows are superposed on a silicon surface placed in a high vacuum. In this case, the system parameters can be varied independently over a wide range. The experiments were carried out in a cap-type device at a background pressure of $\sim 2 \times 10^{-3}$ Pa. A diagram of the device is shown in Fig. 1.

We used a Penning discharge with a cold cathode as an ion source. The source design was similar to that described in [20]. The source chamber was sealed and connected to a high-vacuum chamber through a 1.5-mm-diameter aperture. This allowed us to have a working pressure in the source of 2–10 Pa at a pressure in the high-vacuum chamber no higher than 10^{-2} Pa. A cell was placed in the gap of a magnetic core assembled from permanent magnets. The magnetic induction in the center of the source was ~0.07 T. The characteristic discharge current was 100–120 mA at a discharge voltage of 600–750 V (the cathode diameter was 2 cm, and the anode length was 2.2 cm). The working gases were argon and oxygen. The ions were extracted across the magnetic field. The extractor voltage was 2.5 kV. The beam was focused by a single lens and then decelerated to an energy of 1 keV.

As is known [21], when an ion beam is extracted from the source into a vacuum, intense ion charge

exchange occurs in the parent gas near the extraction aperture. Under our conditions, the conversion coefficient due to this effect can attain 30%. To prevent fast neutrals from reaching the silicon surface, the beam was turned through an angle of ~12° by a cylindrical electrostatic capacitor and then passed through a Winn filter, which was assembled from permanent magnets in an armored case. The transit base was ~6.5 cm. The magnetic field was almost uniform in the ~4-cm-long central region, where the magnetic induction attained 0.16 T. Outside of this region, the magnetic induction dropped steeply. The distance between the plates of the electrostatic capacitor in the filter was 8 mm. The plates were 14 mm high. The inlet and outlet slits of the filter were rectangular in shape (5 × 12 and 3 × 12 mm in size, respectively). About 2 cm behind the filter, the strip beam reached a grounded receiver (Fig. 2). The upper part of the beam was used to treat the silicon surface. A 2.5-mm-diameter aperture *3* (Fig. 2) in the main plate of the receiver was used to monitor the beam parameters. The system was supplied with an additional electrode *ES*, which either suppressed secondary electron emission or extracted secondary electrons from collector *2*, depending on the voltage applied. When operating with positive ions, electrode *ES* was used to suppress secondary emission.

In the Winn filter, the ion beam was separated either by mass (O_2^+ and O^+; O_2^- and O^-) or by charge (Ar^+ and Ar^{2+}). Figure 3 shows that the source produced negative and positive oxygen ions in comparable amounts in both atomic and molecular form. The separation degree was 70–90%. The beam of argon ions almost completely consisted of Ar^+.

Fast neutrals (Ar^0, O_2^0, and O^0) with an energy of 1 keV were produced from positive ions via resonant charge exchange in the parent gas (Ar^+ in Ar, O_2^+ in O_2, and O^+ in O_2). For this purpose, the sides of the electrostatic capacitor of the Winn filter were sealed to obtain a 64-mm-long tube of a rectangular cross section (8 × 14 mm). In the middle part of the tube, there were apertures for measuring the pressure and the gas flow rate. Such a design, together with a high-vacuum space under the cap of the device, formed a charge-exchange chamber with differential pumping. The inlet and outlet slits of the Winn filter increased the pressure gradient. At a gas pressure of ~0.4 Pa in the center of the charge-exchange chamber, the pressure in the space under the cap was no higher than 10^{-2} Pa.

The degree of beam charge exchange was measured according to the technique described in [22]. The technique is based on the assumption that ions and fast neutrals of the same elements eject electrons from the metal surface with the same efficiency; i.e., their coefficients of secondary emission are the same ($k_{i-e} = k_{n-e}$). Strictly speaking, this is true only in the energy range in which kinetic ejection prevails over potential ejection.

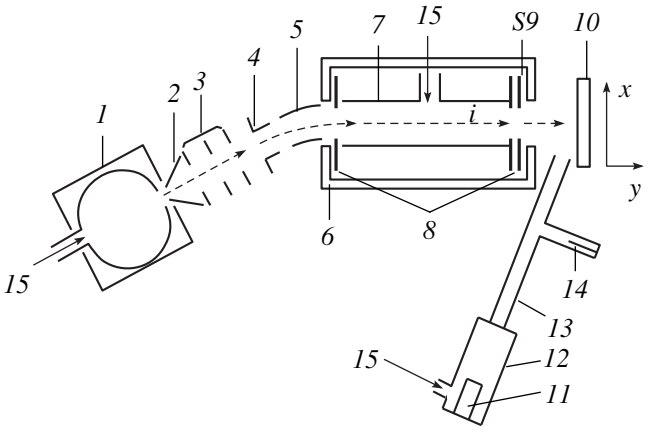

Fig. 1. Diagram of the experimental device: (*1*) ion source, (*2*) extractor, (*3*) single lens, (*4*) decelerating electrode, (*5*) electrostatic turning capacitor, (*6*, *7*) magnet yoke and electrostatic capacitor of the Winn filter, (*8*) inlet and outlet filter slits, (*S9*) auxiliary electrode, (*10*) receiving unit, (*11*) cathode, (*12*) quartz retort, (*13*) glass tube, (*14*) auxiliary anode, (*15*) gas inlet, and (*i*) trajectory of the ion beam.

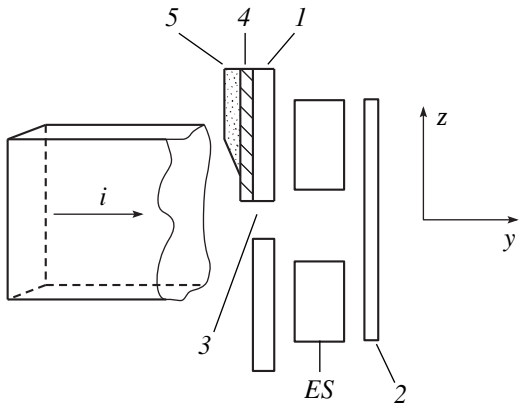

Fig. 2. Diagram of the receiving unit: (*1*) grounded case, (*2*) collector, (*3*) aperture for monitoring the beam parameters, (*4*) treated silicon plate, (*5*) shielding case, (*i*) ion beam, (*ES*) electrode for suppression of secondary emission or extraction of secondary electrons from the collector. The *x*- and *y*-axes in Fig. 1 and the *z*- and *y*-axes in Fig. 2 make the same left-hand Cartesian triple.

The measurement procedure was as follows. In the beginning, the gas inlet into the recharging chamber was shut, an auxiliary electrode *S9* (Fig. 1) was grounded, and the beam losses were almost absent. The Winn filter was tuned to a certain type of ions, e.g., Ar^+. By varying the voltage polarity on electrode *ES*, we could either suppress secondary electron emission or extract all secondary electrons from the collector. In the first case, we obtained the ion current I_i, whereas in the second case, we obtained the collector current $I_k = I_i + I_e$ (where I_e is the secondary electron current). Then, we have $k_{i-e} = I_e/I_i = (I_k - I_i)/I_i$. It turned out that this

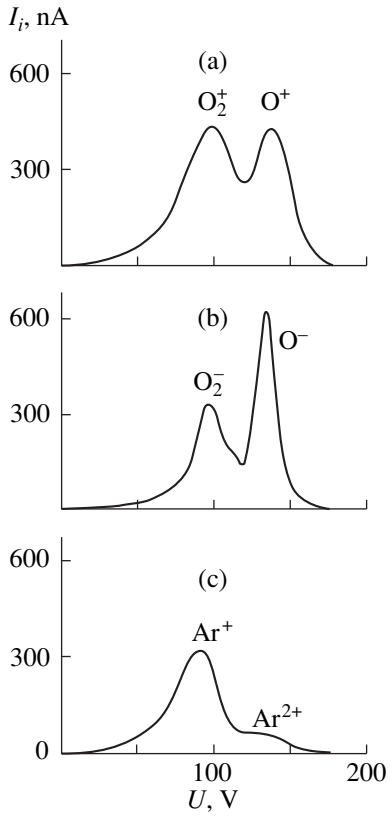

Fig. 3. Mass spectra of the ion beam: I_i is the beam current recorded by the collector of the receiving unit, U is the voltage on the electrostatic capacitor of the Winn filter, the pressure in the source is (a, b) ~6 and (c) ~4 Pa, and the discharge current is (a, b) 120 and (c) 80 mA.

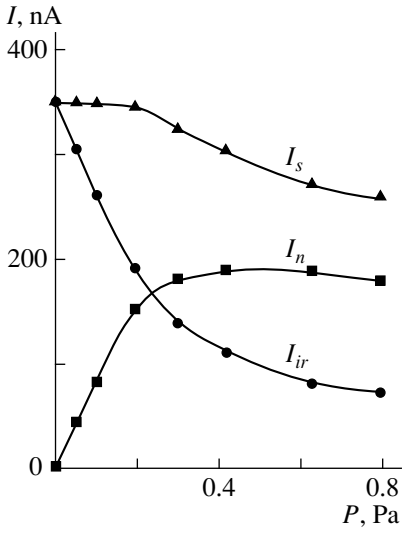

Fig. 4. Unrecharged ion current I_{ir}, electric equivalent of the flow of fast neutrals I_n, and their sum $I_s = I_{ir} + I_n$ as functions of the pressure in the charge-exchange chamber; the working gas is argon.

parameter varied in a rather wide range; hence, it was measured more than once at certain time intervals during every experiment.

Then, the gas was supplied to the charge-exchange chamber. A part of the ions was recharged, while the other part passed through the chamber unchanged. When a sufficiently high positive voltage was applied to electrode *S9*, all the recharged ions were reflected from it and were not able to leave the Winn filter. This can be easily verified by measuring the collector current in the regime of suppression of secondary emission. The measurements showed that the total reflection was observed at U_{S9} higher than 1100 V. Let us now change the polarity of electrode *ES* to positive and measure the current of secondary electrons. Since the ions do not reach the collector, the current I_e may be caused only by the flow of fast neutrals. According to [22], we have $I_n = I_e/k_{i-e}$. Here, the quantity I_n is the electric equivalent of the flow of neutrals, i.e., the current to the collector if every fast atom had a charge equal to the electron charge.

When the gas inlet to the charge-exchange chamber is open, electrode *S9* is grounded, and secondary emission from the collector is suppressed, one more parameter I_{ir} (the current of unrecharged ions) can be measured. This current was measured only in auxiliary experiments.

The dependence of I_{ir}, I_n, and their sum I_s on the gas pressure in the central part of the charge-exchange chamber is shown in Fig. 4. The working gas is argon. It is seen that I_n reaches its peak value at $P \sim 0.5$ Pa and then decreases slowly. The latter is associated with the scattering of fast neutrals by gas atoms. A similar dependence is observed for molecular and atomic oxygen, but at a pressure two to three times higher.

We draw attention to two important facts. First, I_s is equal to the initial current of the ion beam (before the gas was supplied to the charge-exchange chamber) with a high accuracy in the range $P \le 0.2$ Pa, which supports the validity of this measurement technique. Second, if we assume that the measurement error is high, the intensity of the actual flow of neutrals cannot be higher than a preset value, because the intensity of the total flow of neutrals and unrecharged ions cannot exceed that of the initial positive ion beam.

A glow discharge in a gas flow with a highly nonuniform pressure served as a source of fluorine radicals (Fig. 1). A duralumin cathode *11* with an area of 2 cm^2 was placed into quartz bulb *12* with a volume of ~10 cm^3, into which SF_6 gas was input. The bulb was connected to a vacuum chamber through a 22-cm-long glass tube *13* with an inner diameter of 8 mm. There was a side branch at a distance of 6 cm from the tube end, where the auxiliary anode *14* was placed. At a standard SF_6 flow rate of 900 cm^3 Pa/s, the gas pressure near the cathode was 25–30 Pa and, in the space under the cap, it was $\sim 7 \times 10^{-3}$ Pa. The auxiliary anode was used to help initiate the discharge with a highly nonuniform pressure. In the normal regime, the voltage across the discharge was 2.5–3 kV and the current was 4 mA.

The current was about equally shared between the auxiliary anode and the grounded elements of the cap.

Fluorine atoms were obtained due to dissociation of sulfur hexafluoride molecules in the positive column of the discharge. The measurements showed that, outside of the glass tube, the etching rate of Si, which is determined by the concentration P^*, is described well by the dependence $V_{et} \sim 1/R^2$, where R is the distance from the tube end. The axis of the radical source made an angle of ~20° with the plane of the receiver plate and with the silicon surface. The distance between the tube end and the center of the ion beam (Fig. 1) was 2 cm. A permanent magnet was placed behind the receiver plate to create a magnetic field (~7×10^{-3} T) directed along the y-axis (Fig. 1). This field prevented electrons of the glow discharge from reaching the region where the beam ions and fluorine radicals interacted with the treated surface.

MEASUREMENT TECHNIQUE AND EXPERIMENTAL RESULTS

The etching rate was determined by direct measurements of the etch depth. For this purpose, Si sample *4* was placed on receiving plate *1* (Fig. 2). Its upper part was protected from the influence of fluorine and the ion beam by casing *5*. The interferometer microscope measured the height of the step between treated and untreated parts of the sample; usually, it was about several microns. As was mentioned in the previous section, the rate of radical etching followed the law $V_{et} \sim 1/R^2$. For this reason, when the ion beam was switched off, the profile of the etch depth along the x-axis had the shape shown in Fig. 5a. The origin of the x-axis was at the end of the glass tube. When the silicon surface was simultaneously exposed to the radical and ion flows, the etching was accelerated in the region of their combined action and a step appeared in the profile of the etch depth (Fig. 5b). It is easy to extrapolate curves d and e and to obtain the value of Δh. The distance R was taken equal to 2 cm to obtain $\Delta h \cong h$, where h is the depth of silicon etching by fluorine radicals without ion stimulation.

The yield of silicon atoms per incident ion Y (the main measured quantity) was determined as follows. Let the increment of the etch depth due to ion bombardment during the time Δt be Δh. The number of atoms removed from the surface is determined by the expression $N_{Si} = \Delta h \rho N_a/\mu$, where ρ and μ are the density and molar weight of silicon, respectively, and N_a is Avogadro's number. The number of ions incident per unit area is $N_i = j\Delta t/e = I\Delta t/eS$, where j is the ion current density, I is the collector current, S is the collector area, and e is the electron charge. Since the beam current density was about several microamperes per square centimeter, the value Δt was typically several hours. It was impossible to keep the parameters of the source unchanged over such a long period. Hence, instead of the value $I\Delta t$, its integral value $\Phi = \int I\,dt$ was measured. To determine it, an integrating scheme was installed in the collector circuit. To find the yield of silicon atoms per incident fast neutral atom, we measured the value of Φ, in which the current I was replaced with the equivalent current I_n described above. As a result, we obtained $Y = \Delta h \rho N_a eS/\mu\Phi$. For each type of stimulating particle, a series of not less than five measurements was performed. The averaged data are given below.

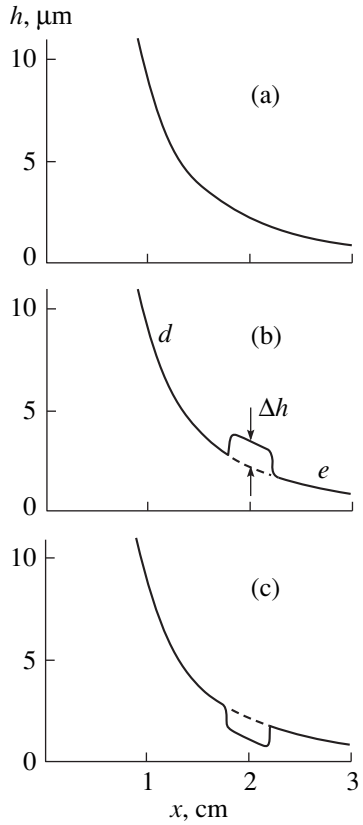

Fig. 5. Qualitative dependence of the silicon etch depth on the x coordinate (the point $x = 0$ is at the end of the radical source tube; see Fig. 1): (a) the ion source is switched off, and only the radical source operates; (b) etching is stimulated by the ion flow; and (c) the case when the process is decelerated by some external factors.

It is known [13] that the yield of silicon atoms strongly depends not only on the ion mass and ion energy, but also on the experimental conditions. The yield is typically from 4 [5] to 20 [23] for Ar^+ ions with an energy of 1 keV. In our experiments, we obtained $Y(Ar^+) = 5.7 \pm 1.3$ atom/ion. Since this value was closer to the results of [5], we used it as a reference value.

For fast argon atoms, we obtain $Y(Ar^0) = 10.6 \pm 1.5$ atom/neutral, i.e., $Y(Ar^0)/Y(Ar^+) \cong 1.85$. This result is unexpected for two reasons. First, as mentioned in the Introduction, the stimulating influence of ion bombardment is mainly associated with the breaking of Si–Si bonds, loosening of the silicon surface, and induced desorption of the intermediate products of the

SiF_x reaction (x = 1, 2, or 3). The existence of these effects and their substantial contribution to the etching process is undoubted. However, this influence is only mechanical; it cannot depend on the states of the upper electron shells (recall that the ion energy is 1 keV, while the electron transitions at the upper shells correspond to energies of one to two tens of electronvolts). This makes us assume the existence of an additional etching channel, which has not been previously accounted for. Second, it is known [24, 25] that when an ion approaches the surface, it is usually neutralized before colliding due to Auger or tunnel processes; i.e., almost always, it is fast neutrals that reach the surface. However, the place, time, and mechanism of neutralization are different. For this reason, we doubt the validity of the measurement technique for I_n. However, as was shown in the previous section, even if we assume the presence of a substantial error, the actual flow of fast neutrals cannot exceed a preset value, because this would contradict the current balance. On the other hand, $Y \sim 1/j$ and, consequently, the obtained value of $Y(Ar^0)$ is actually the lower limit. Therefore, it is proved that stimulating influence of ion bombardment on plasmochemical etching of silicon depends not only on the ion mass and ion energy, but also on the population of its upper electron levels.

The experiments showed that atomic oxygen also accelerates the etching process. In this case, $Y(O^+) = 1.9 \pm 0.2$ atom/ion, $Y(O^0) = 3.4 \pm 0.5$ atom/neutral, and $Y(O^-) = 2.1 \pm 0.5$ atom/ion. Fast neutrals have the strongest stimulating effect: $Y(O^0)/Y(O^+) \cong 1.8$. Negative ions take an intermediate position.

Unlike the particles mentioned above, the ions of molecular oxygen do not accelerate, but decelerate the etching process. As a result, a hole (rather than a step) is formed in the etch depth profile in the region exposed to the beam action (Fig. 5c). For convenience, this process can be regarded as negative catalysis, whereas its intensity can be characterized by negative values of Y, because Δh is negative. It was found that all kinds of molecular oxygen (O_2^+, O_2^-, and O_2^0) exhibit properties of inhibitors with $Y(O_2^+) = -2.9 \pm 0.3$ atom/ion, $Y(O_2^0) = -2.3 \pm 0.2$ atom/neutral, and $Y(O_2^-) = -2.6 \pm 0.3$ atom/ion. Negative ions again occupy an intermediate position between positive ions and fast neutrals. The observed decelerating action cannot be directly associated with the formation of SiO_2, which, as is known, is etched by fluorine at a much slower rate than pure silicon. Such effects are observed in the mixture of SF_6 and oxygen only when the relative concentration of O_2 is higher than 50%. In our case, the flow of SF_6 molecules onto the target is more than two orders of magnitude higher than that of molecular ions.

The qualitative difference between the influence of atomic and molecular oxygen is noteworthy. The former accelerates, while the latter decelerates the process of etching, although the chemical element is the same. We cannot exclude the possibility that the reason is the difference between the ion masses and, hence, the ion momentum at the same energy. However, it is unlikely because, as shown in [23], the yield of silicon atoms per incident ion is a very weak function of the ion energy, at least in the range from 500 to 3000 eV. Solving this problem requires further investigations.

CONCLUSIONS

The first results of investigations of the influence of the ion charge sign on the stimulation of plasmochemical etching of silicon show the following.

(1) Among the particles that accelerate the etching process, fast neutrals have the maximum catalytic effect. The catalytic influence of positive ions is about two times less. Negative ions occupy an intermediate position.

(2) It is found that some kinds of ions do not accelerate, but decelerate the etching process; i.e., they behave as inhibitors. Among such particles (at energies of 1 keV) are all kinds of molecular oxygen (O_2^+, O_2^-, and O_2^0).

(3) The results obtained allow us to conclude that the effects of loosening the silicon surface and forced desorption of the intermediate reaction products incompletely describe the influence of ion bombardment. There should be an additional channel of silicon etching, whose efficiency depends not only on the energy and mass of the ion, but also on the population of its upper electron levels.

REFERENCES

1. N. Hosokawa, R. Matsuzaki, and T. Asamaki, Jpn. J. Appl. Phys., Suppl. **2**, 435 (1974).
2. L. Holland and S. M. Ojha, Vacuum **26** (1), 53 (1976).
3. *Plasma Processing for VLSI*, Ed. by N. Einspruch and D. Brown (Academic, New York, 1984; Mir, Moscow, 1987).
4. B. S. Danilin and V. S. Kireev, *Applications of Low Temperature Plasmas for Etching and Purification of Materials* (Énergoatomizdat, Moscow, 1987).
5. J. W. Coburn and H. W. Winters, J. Appl. Phys. **50**, 3189 (1979).
6. D. L. Flamm and V. M. Donnelly, J. Vac. Sci. Technol. B **1**, 23 (1983).
7. M. J. Pabst, H. S. Tan, J. L. Franklin, *et al.*, Int. J. Mass Spectrom. Ion Phys. **20** (1), 191 (1976).
8. R. F. Williams, in *Plasma Processing of Semiconductors*, Ed. by R. F. Williams (Kluwer, Dordrecht, 1996) [NATO ASI Ser., Ser. E **336**, 321 (1996)].
9. J. Perrin, in *Plasma Processing of Semiconductors*, Ed. by R. F. Williams (Kluwer, Dordrecht, 1996) [NATO ASI Ser., Ser. E **336**, 397 (1996)].
10. D. I. Slovetskiĭ, in *Plasma Chemistry* (Énergoizdat, Moscow, 1983), Vol. 10, p. 108.

11. H. F. Winters and J. W. Coburn, J. Vac. Sci. Technol. B **3**, 1376 (1985).
12. V. M. Donnelly, N. Layadi, J. T. C. Lee, *et al.*, in *Plasma Processing of Semiconductors*, Ed. by R. F. Williams (Kluwer, Dordrecht, 1996) [NATO ASI Ser., Ser. E **336**, 243 (1996)].
13. H. F. Winters and J. W. Coburn, Surf. Sci. Rep. **14** (4–6), 161 (1992).
14. D. L. Flamm, in *Plasma Processing of Semiconductors*, Ed. by R. F. Williams (Kluwer, Dordrecht, 1996) [NATO ASI Ser., Ser. E **336**, 23 (1996)].
15. J. L. Mauer, J. S. Logan, L. B. Zielinski, and G. C. Schwartz, J. Vac. Sci. Technol. **15**, 1734 (1978).
16. V. Yu. Kireev, D. A. Nazarov, and V. I. Kuznetsov, Élektron. Obrab. Mater., No. 6, 37 (1986).
17. D. J. Oostra, A. Haring, A. E. de Vries, *et al.*, Nucl. Instrum. Methods Phys. Res. B **13**, 556 (1986).
18. V. Yu. Kireev and M. A. Kremerov, Élektron. Tekh., Ser. 3: Mikroélektron., No. 1, 3 (1985).
19. J. H. Keller, Plasma Phys. Controlled Fusion **39**, A437 (1997).
20. T. I. Danilina, E. V. Ivanova, Yu. E. Kreĭndel', and L. A. Levshchuk, Prib. Tekh. Éksp., No. 3, 158 (1968).
21. J. B. Hasted, *Physics of Atomic Collisions* (Butterworths, London, 1964; Mir, Moscow, 1965).
22. M. S. Ioffe, R. I. Sobolev, V. G. Tel'kovskiĭ, and E. E. Yushmanov, Zh. Éksp. Teor. Fiz. **39**, 1602 (1960) [Sov. Phys. JETP **12**, 1117 (1960)].
23. U. Gerlach-Meyer and J. W. Coburn, Surf. Sci. **103**, 177 (1981).
24. H. D. Hagstrum, Phys. Rev. **104**, 672 (1956).
25. H. D. Hagstrum and Y. Tekeishi, Phys. Rev. **139**, 526 (1965).

Translated by M. Astrov

Technical Physics, Vol. 46, No. 2, 2001, pp. 174–178. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 43–47.

Optical Studies of the Plasma–Liquid Transition Layer in Pulsed Corona Discharges in Strong Water Electrolytes

L. Z. Boguslavsky, S. A. Khaĭnatsky, and A. N. Shcherbak

Institute of Pulse Research and Engineering, National Academy of Sciences of Ukraine, Oktyabrskiĭ pr. 43a, Nikolaev, 54018 Ukraine

e-mail: ksa@aip.mk.ua

Received March 17, 2000

Abstract—The properties of corona discharge in strong water electrolytes are studied experimentally. An analysis of the measured time-integrated spectrum of the corona plasma emission shows the presence of a thin vapor layer between the plasma and liquid. The time dependences of the layer thickness are derived from the discharge shadowgraphs. The role of the transition layer in the energy conversion process in a corona discharge in strong electrolytes is determined. © *2001 MAIK "Nauka/Interperiodica".*

Studies of the electric and hydrodynamic characteristics of pulsed corona discharges in water electrolytes [1, 2] showed that, under certain conditions, the branched spray corona transforms into a continuous plasma formation, in which the level of hydrodynamic perturbations is comparable to that in spark underwater discharges [3]. At initial electric fields in the range $E_0 = 10^6$–10^8 V/m, the key condition for such a transformation is the high electrical conductivity of electrolytes (on the order of 10 S/m). In this case, if the surface of the point electrode provides an electric-current density and electric-field strength high enough to ignite a discharge over the entire surface of the point, then a plasma piston with any given configuration, depending on the geometry of the point, can be formed in a liquid [2]. These factors, along with the high stability of the electric and hydrodynamic characteristics, stimulate interest in corona discharges in strong electrolytes. Such discharges possess a high technical potential and have advantages over other types of electric explosions in liquids. They can be realized in very small volumes and are characterized by stable pressure levels, elevated resources of the electrode systems, and a wider range of electric conductivity of the media in which the electric-explosion energy conversion occurs.

A further investigation of corona discharges in strong electrolytes requires a detailed explanation of the mechanism for this phenomenon. In the existing electrodynamic models of corona discharge in a strong electrolyte [4, 5], the equivalent circuit for the discharge gap is represented by two series-connected nonlinear resistors formed by the expanding plasma piston and liquid. The total resistance varies due to variations in the plasma-piston radius and nonlinear electric-conductivity distribution, which is related to the temperature gradient caused by Joule heating. This model is based on the fact that, using conventional experimental techniques, it is impossible to divide the total energy dissipated in the discharge gap into equivalent components. In this case, the transition from plasma to liquid is regarded as a jump. However, studies of the underwater-spark channel show that, between the plasma and liquid, there is a thin transition layer in which the thermodynamic parameters T and ρ vary gradually [6, 7]. The existence of such a layer is explained by the facts that the characteristic pressure in the plasma channel attains 10^8–10^9 Pa and the water parameters are higher than the critical ones, so that the channel cannot have a sharp plasma–liquid boundary.

Studies of a pulsed diaphragm discharge [8, 9] (which is similar to a pulsed corona discharge because it occurs in the same media and also leads to the formation of a plasma bunch) demonstrate the presence of a transitional gas layer which plays a decisive role in the discharge evolution because of the high current density near the plasma bunch.

The facts listed above show that it is necessary to answer the question of whether a plasma–liquid transition layer arises in a corona discharge in a strong electrolyte when a continuous plasma region is formed, and to study the role of the transition layer in the discharge evolution.

EXPERIMENTAL SETUP

The experimental setup consisted of electrical and optical parts. The electrophysical part was the discharge circuit of a capacitor bank with the parameters $U_0 = 10$–30 kV, $C = 3$–6 μF, and $L = 2.4$ μH. The discharge current and the voltage across the discharge gap were recorded by an S8-17 oscilloscope with the help of a coaxial shunt and capacitive voltage divider. The optical part of the experimental setup (Fig. 1, top view) consisted of the spectral and shadowgraph sections.

1063-7842/01/4602-0174$21.00

A discharge was ignited in a discharge chamber, which was a parallelepiped 240 × 480 × 390 mm in size. The chamber had three windows made of highly transparent polished Plexiglas. The chamber material was Kh18N9T stainless steel. A brass rod electrode (with a curvature radius of the point of r_0 = 1.5 mm) insulated with polyethylene was mounted at the flange of the lateral wall of the chamber. The discharge chamber wall served as the negative electrode. As an electrolyte, we used an optically transparent distilled water solution of NaCl, which was preliminarily settled and filtered. The salt concentration in the electrolyte varied depending on the required electrolyte conductivity. In this experiment, it was σ_0 = 10 S/m. One of the windows was mounted perpendicular to the rod electrode to prevent defocusing of the optical scheme during the spectral measurements of the plasma evolution. With the help of a deflecting mirror and an intermediate achromatic quartz lens (f = 75 mm), the optical plasma radiation was focused at the entrance slit of a DFS-452 spectrograph. The spectrograph recorded the emission spectra over the range 190–1100 nm with a high resolution. In our experiments, we used a built-in 600-line/mm grating with maximum reflectivity at a wavelength of 500 nm. In this case, a film covers the wavelength range up to 360 nm, which is convenient when investigating the overview spectra of little-studied objects. To identify the spectrum under investigation, we used an IVS-28 standard spectrum source in whose discharge chamber an aluminum emission with a well-known spectrum was generated. With the help of a deflecting mirror and an intermediate lens, the aluminum emission was focused at the entrance slit of the spectrograph and was recorded on a free part of the film parallel to the spectrum under investigation. The mutual position of the spectrum under investigation and the reference spectrum was adjusted by a special film-carrier mechanism.

The shadowgraph section, which was developed for the visual identification of a thin plasma–liquid transition layer, operated on the principle of the defocused diaphragm [10]. A GOR-100M pulsed ruby laser was used as a light source. Using a plane-parallel light beam, we could partially prevent the lens effect of the shock front of the corona discharge. To enlarge the size of the light spot in which optical nonhomogeneities were studied, a collimator broadening the light beam to 50 mm was placed between the laser and the discharge chamber. A receiving Yupiter-36B objective (f = 250 mm) was placed on the optical axis behind the discharge chamber. In the recording part of the optical system, a Kaleinar-3B objective (f = 150 mm) was used to produce a sharp image of an object (which was positioned in the object plane) on the visualizing screen or on the photorecorder film. The use of the above objectives in the optical system was necessary because of fairly long (≥74 mm) working segments, which is important for the defocused diaphragm method with a high resolution (up to 45 line/mm) in the center of the light spot. A visualizing screen made of opal optical glass was placed behind the recording objective. Using this screen, we could visually observe the image when adjusting the system or preparing to film and could check the quality of the device adjustment. For high-speed photography, we used a VFU-1 photorecorder with a magnifying lens (5- and 10-fold magnification), which operated as a streak camera. The optical system as a whole permitted a resolution of 10^{-5}–10^{-4} m. As a visualizing diaphragm, a set of grids with a step of 0.05 to 1.0 mm was used. The diaphragm was positioned directly behind the object, because, in this case, light rays are deflected at large angles in the transition layer.

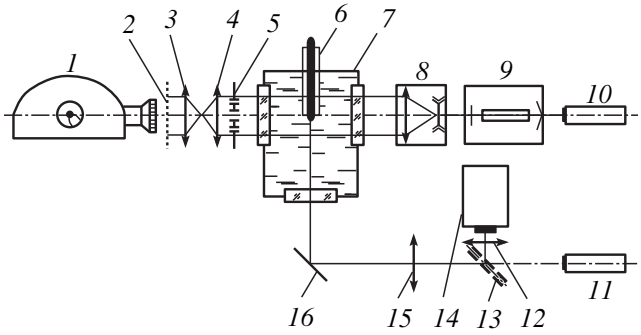

Fig. 1. Optical part of the experimental setup: (*1*) photorecorder, (*2*) visualizing screen, (*3*) recording objective, (*4*) receiving objective, (*5*) visualizing diaphragm, (*6*) electrode, (*7*) discharge chamber, (*8*) collimator, (*9*) pulsed laser, (*10*) adjusting laser, (*11*) diffraction-grating spectrograph, (*12*) intermediate lens, (*13*) removable deflecting mirror, (*14*) standard spectrum source, (*15*) intermediate lens, and (*16*) deflecting mirror.

All components of the optical system were mounted on a 3750-mm-long optical bench to prevent an adverse vibration effect. The frame of the optical system lay on 100-mm-thick rubber pads. The discharge chamber was installed on an individual platform fastened rigidly to the laboratory floor and was mechanically isolated from the optical system to prevent misalignment of the latter.

In the first stage of adjusting the optical system, we also carried out an experiment to determine the shape of the plasma bunch at given parameters of the electric circuit, point radius, and electrolyte. In this case, the time evolution of the plasma corona was recorded by a VFU-1 streak camera in the time loupe regime with a time resolution of 5×10^{-7} s. An IFK-2000 pulsed lamp was used as the illuminator.

EXPERIMENTAL RESULTS

Frame-by frame photography of the discharge demonstrated that, in the chosen range of electric-circuit parameters and σ = 10 S/m, a plasma bunch produced with a hemispherical point is shaped like a hemispherical layer. Investigations of the time-integrated emission spectrum of the continuous plasma of a corona dis-

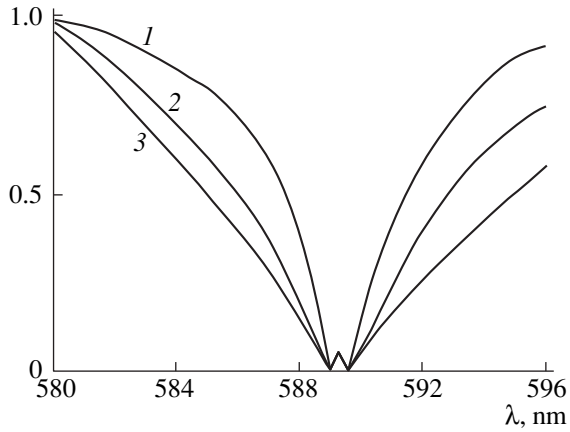

Fig. 2. Time evolution of the Na doublet profile in the plasma emission spectrum of the corona discharge at different values of U_0: (*1*) 10, (*2*) 15, and (*3*) 20 kV for $C = 3$ μF.

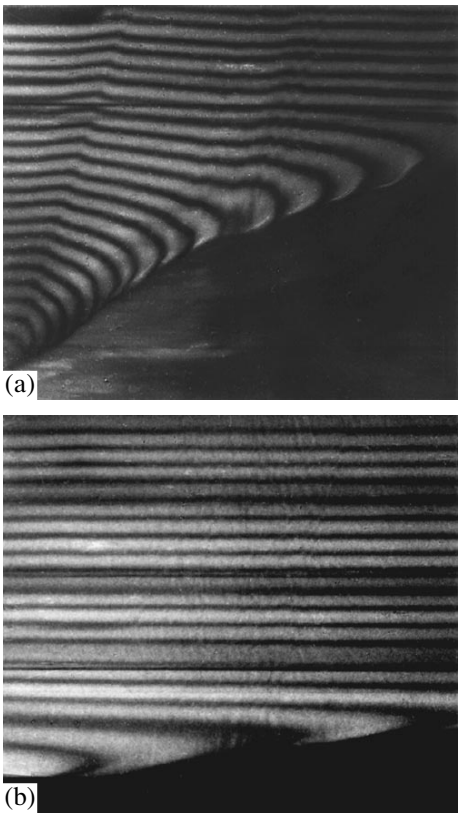

Fig. 3. Shadowgraphs of the corona discharge at different values of U_0: (a) 29 and (b) 15 kV for $C = 6$ μF.

charge have revealed that, against the background of an intense continuum, a wide absorption band corresponding to the sodium resonance doublet is observed in all of the discharge regimes under study. The band broadens with increasing initial voltage. Figure 2 shows the time evolution of the Na doublet profile in the emission spectrum of the corona-discharge plasma, measured with reference to the known emission spectrum of Al. The total absorption corresponding to the Na resonance doublet ($\lambda = 589$–589.6 nm) is most pronounced when the initial voltage U_0 is increased to 15 kV. This indicates that, between the plasma and electrolyte, there is a boundary gas layer in which the temperature is nearly one order of magnitude lower than the plasma temperature and the transition from the plasma to the gas layer may occur abruptly. The fact that the recorded spectrum does not contain additional absorption lines corresponding to the transitions from the resonant level is evidence that the emitted radiation passes through a fairly narrow, weakly absorbing vapor layer in the visible region.

From the time-integrated emission spectrum obtained in our experiments, we could not accurately evaluate the size of the observed transition layer. The thickness of the transition layer was estimated by assuming that it is inversely proportional to the coefficient of continuous absorption in the spectral region near $\lambda = 589$ nm. Using the data on the absorption coefficient calculated for a low-temperature oxygen–hydrogen plasma [11], we found that, for $\chi_{589} \approx 4 \times 10^4$ mm^{-1}, the half-thickness of the gas layer is equal to $\delta \approx 1/\chi \cong 2.5 \times 10^{-2}$ mm.

Figure 3 shows the shadowgraphs of a corona discharge. The inflection of the diaphragm-grid shadows is seen against the background of the time evolution of the plasma bunch. The shadow bands are inflected in three regions: (i) the compression-wave region, (ii) the region where the electrolyte is heated, and (iii) the transition-layer region. At high initial capacitive-storage voltages, a shock is observed in the compression-wave region; as the initial voltage decreases, the shock becomes diffuse. The electrolyte is heated more intensively when the electric parameters ensure a more intense energy release in the discharge. By the transition-layer region, we mean the region adjacent to the plasma where the inflection angle of the grid shadows becomes negative.

From the shadowgraphs of the discharge, the time dependences of the thickness of the layer in which the inflection of shadow bands was observed were calculated for different discharge-circuit parameters (Fig. 4). The dependences were calculated taking into account the inflection of shadows due to the lens effect in the compression-wave region [7]. The layer thickness increases rapidly only during the first microseconds (to 2 μs) of the discharge. For most of operating conditions, the transition layer thickness reaches 0.05 mm and increases only by the end of the energy release in the discharge gap. However, we also observed discharges with high stored energies ($U_0 = 29$ kV, $C = 6$ μF) in which the transition layer thickness reached a fairly large value of 0.2–0.3 mm.

On the whole, the experimental data on the transition layer thickness that were obtained using the shadowgraph technique agree with the above estimate obtained from spectral measurements. This fact is evi-

dence that the plasma–liquid transition layer is correctly identified in the shadowgraphs.

DISCUSSION

The experimental results obtained allow us to estimate the role of the transition layer in the time evolution of a pulsed corona and the influence of the processes of energy conversion in this layer on the total energy balance in a discharge. During the evolution of the plasma bunch in a corona discharge, the intensity of surface evaporation of the liquid is governed by radiation, heat conduction, and Joule heating. To determine the energy that is spent on the formation of the transition layer and is dissipated in it, we consider the conditions at the plasma–liquid interface, i.e., at the phase-transition boundary. The temperature T of the liquid layer adjacent to the transition layer grows due to Joule heating by the current I flowing through the layer. The time-dependent distribution of T can be determined from the one-dimensional heat conduction equation

$$\rho_l c_{\nu_l}\left(\frac{\partial T_l}{\partial t} + \nu_r \frac{\partial T_l}{\partial r}\right) = \frac{\partial}{\partial r}\left(\alpha \frac{\partial T}{\partial r}\right) + w(t), \qquad (1)$$

where ρ_l is the mass density, c_{ν_l} is the specific heat, α is the thermal conductivity of water, ν_r is the radial velocity, and $w(t)$ is the heat-source power density.

In our case, we have

$$w(t) = 2\frac{J^2(t)}{\sigma_1(T_l(r,t))}. \qquad (2)$$

The label l stands for the liquid parameters. Similar relations hold for both the transition-layer region and the coronal plasma. The continuity conditions for the current density J, mass and energy flows, potential φ, and temperature T are satisfied at the phase-transition boundary. A mathematical description of this time-dependent process with nonlinear boundary conditions at the initially unknown phase-transition boundaries, even with some model simplifications, is rather complicated. For this reason, we restrict ourselves to some estimates in our analysis. We will use the following estimation parameters: the transition layer thickness $\delta = 2.5 \times 10^{-4}$ m, the current $I = 2 \times 10^4$ A, the temperature $T = 5 \times 10^3$–2×10^4 K, the plasma-bunch radius (including the electrode radius) $r_p = 4 \times 10^{-3}$ m, and the initial electrical conductivity of the liquid $\sigma_0 = 10$ S/m. This parameters correspond to the regime with high stored energy. The induced convection of liquid by a plasma piston may be ignored. The reason is that, in spite of rather high flow velocities (ν_r can attain 1/3 of the speed of sound), convection does not redistribute the temperature in the local regions of interest. Heat can be transferred through the transition layer from the plasma to the liquid by convective heat conduction (α = 0.683 W/(m deg)) and dissociative heat conduction, which, according to [8], is the main heat-conduction mechanism in such systems and for which we take α_d = 1.4 W/(m deg). For the maximum attainable temperature gradient, this flow can be estimated as

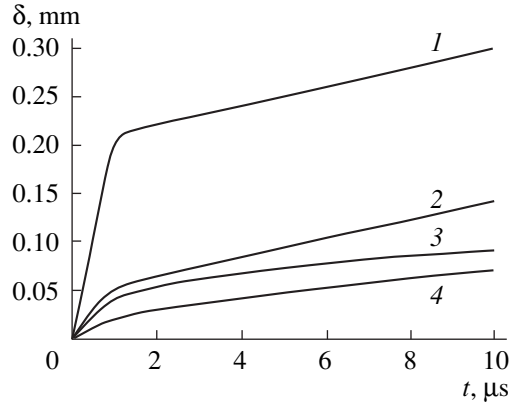

Fig. 4. Thickness of the visualized transition layer at U_0 = (*1*, *3*) 29 and (*2*, *4*) 22 kV for C = (*1*, *2*) 6 and (*3*, *4*) 3 μF.

$$\left.\frac{\alpha \Delta T}{\delta}\right|_{T = 2\times 10^4\ \mathrm{K}} = 5.46 \times 10^7\ \mathrm{W/m^2},$$

$$\left.\frac{\alpha_d \Delta T}{\delta}\right|_{T = 2\times 10^4\ \mathrm{K}} = 1.12 \times 10^8\ \mathrm{W/m^2}.$$

In addition, the gas can be heated due to radiant heat transfer with $g = \sigma_S T^4 = 3.5 \times 10^7$–$9 \times 10^9$ W/m^2, where σ_S is the Stefan–Boltzmann constant. Joule heating in the transition layer can be estimated from Eq. (2) as

$$\frac{I^2 \delta}{4\pi^2 \sigma_{tl} r_p^4} \approx 4 \times 10^{16} \frac{\delta}{\delta_{tl}}\ \mathrm{W/m^2}.$$

Even assuming that σ_{tl} is on the order of the electric conductivity of water, the power of Joule heating in the transition layer is 3–4 orders of magnitude higher than the heat power transferred from the plasma. In fact, the temperature and pressure dependences of the electric conductivity of the vapor of a water solution of NaCl [12] are such that, at temperatures on the order of 400°C, the conductivity reaches its maximum (nearly ten times σ_0 for vapor at 100°C) and then decreases below σ_0 at temperatures on the order of 600°C at 10^8 Pa or 800°C at 2×10^8 Pa. This means that the electric conductivity of the gas in the transition layer is maximum at both boundaries of the layer and is minimum (below σ_0 for vapor at 100°C) in the center of the layer. From here, we estimate the electric conductivity averaged over the layer to be on the order of two to four times σ_0 for 100°C, which increases the power released in the layer by one order of magnitude. In this case, the energy released in the transitional layer for 10 μs is estimated to be on the order of 400 J. Evidently, this is the upper estimate and the actual value may be several times lower; however, this estimate shows that the tran-

Table

t, μs	I, A	r_p, m	δ, m	$d\delta/\delta$	$d\delta/\delta$, calculation
1.75	17 027	0.0070	2.1×10^{-4}	–	0.155
3.50	18 611	0.0080	2.3×10^{-4}	0.087	0.108
5.25	17 027	0.0085	2.5×10^{-4}	0.080	0.071
7.00	13 859	0.0085	2.7×10^{-4}	0.074	0.047

sition layer plays an important role in the process of energy conversion in corona discharges in electrolytes.

Finally, we estimate the role that Joule heating of water adjacent to the transition layer plays in the layer formation. We consider a water layer with an arbitrary thickness x; the only requirement is that x be much less than r_p. The density of the heat flux due to evaporation from a spherical surface of radius R is

$$q = \rho_l \frac{dR}{dt}[c_p(T - T_0) + r_s].$$

Here, r_s is the specific heat of evaporation and c_p is the specific heat of water. We assume that this flux is completely provided by Joule heating at the boundary of the layer. Taking into account that, in our case, the radius is $R = r_p + \delta$ (where only the transition layer thickness changes due to evaporation from the surface) and x and δ are small, using Eqs. (1) and (2) we can write

$$\frac{d\delta}{x} = \frac{I^2 dt}{4\pi^2 r_p^4 \sigma_l \rho_l [c_p(T - T_0) + r_s]}.$$

Since x is chosen arbitrarily and the main scale length of the problem is δ, we can substitute x with δ and integrate this expression using the experimental time dependences of the current and corona radius. Since the water temperature distribution across the layer and its time dependence, as well as the phase-transition dynamics, are unknown, we consider only the limiting case (water heated to the critical temperature). Then, we obtain

$$\frac{d\delta}{\delta} \approx \frac{I^2 dt}{\sigma_l r_p^4 Q},$$

where $Q = 4\pi^2 \rho_c [c_p(T_c - T_0) + r_s]$. Here, the label c stands for the parameters at the critical point. The values of $d\delta/\delta$ calculated using this expression and the experimental values of the transition layer thickness for one of the regimes with a high stored energy are given in the table.

A rather close agreement between the calculated and experimental values confirms that, in the active stage of the corona evolution, its propagation deep inside the gap may be largely related to heating and a phase transition, probably a transition of the second kind (over the lability line), rather than hydrodynamical or other processes.

Thus, optical studies of the plasma–liquid transition layer in a pulsed corona discharge in a strong water electrolyte have shown that this layer plays a significant (or even decisive) role in the formation and evolution of the corona.

REFERENCES

1. G. A. Ostroumov, *Interactions of Electric and Hydrodynamic Fields* (Nauka, Moscow, 1979).
2. L. Z. Boguslavsky, V. V. Kucherenko, and E. V. Krivitskiĭ, Preprint No. 22, IIPT (Institute of Pulse Research and Engineering, National Academy of Sciences of Ukraine, Nikolaev, 1993).
3. E. V. Krivitskiĭ, Zh. Tekh. Fiz. **61** (1), 9 (1991) [Sov. Phys. Tech. Phys. **36**, 4 (1991)].
4. V. A. Pozdeev, N. M. Beskaravaĭnyĭ, and V. K. Sholom, Élektron. Obrab. Mater., No. 3, 33 (1990).
5. L. Z. Boguslavsky and E. V. Krivitskiĭ, Teor., Éksp., Prakt. Élektrorazryad. Tekhnol. **1**, 4 (1993).
6. V. V. Arsent'ev, Prikl. Mekh. Tekh. Fiz., No. 5, 51 (1965).
7. A. L. Kupershtokh, Prikl. Mekh. Tekh. Fiz., No. 6, 64 (1980).
8. É. M. Drobyshevskiĭ, B. G. Zhukov, and B. I. Reznikov, Zh. Tekh. Fiz. **47**, 256 (1977) [Sov. Phys. Tech. Phys. **22**, 148 (1977)].
9. V. M. Sokolov, Zh. Tekh. Fiz. **54**, 1890 (1984) [Sov. Phys. Tech. Phys. **29**, 1112 (1984)].
10. M. M. Skotnikov, *Quantitative Shadowgraphy Methods in Gas Dynamics* (Nauka, Moscow, 1976).
11. A. N. Shcherbak, *Physics and Technology of Electropulse Processing of Materials* (Naukova Dumka, Kiev, 1984), p. 27.
12. A. S. Quist and W. L. Marshall, J. Phys. Chem. **72**, 684 (1968).

Translated by N. Larionova

Technical Physics, Vol. 46, No. 2, 2001, pp. 179–181. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 48–50.

Limiting Operating Pressure in a Plasma Source of Electrons with Hollow-Cathode Discharge

Yu. A. Burachevskiĭ, V. A. Burdovitsin, A. V. Mytnikov, and E. M. Oks

Tomsk State University of Control Systems and Radioelectronics, Tomsk, 634050 Russia

e-mail: oks@fet.tusur.ru

Received March 23, 2000

Abstract—Results of an experimental study of the correlation between the accelerating gap parameters in a plasma source of electrons and the limiting magnitudes of the gas pressure and the voltage applied to the gap are presented. It has been found that the electron beam increases the electrical strength of the gap. © *2001 MAIK "Nauka/Interperiodica".*

The importance of the problem of beam plasma generation in the fore-vacuum range of gas pressures for some technological applications, such as annealing and melting of materials, electron-beam surface treatment, and initiating beam-plasma discharges in plasmochemistry [1], has stimulated a demand for reliable and efficient electron sources capable of producing an electron beam at pressures of about 100 mtorr. This problem can be solved with plasma sources of electrons utilizing a cold (unheated) cathode discharge [2, 3]. It is the capability of producing intense electron beams at elevated pressures that makes plasma electron emitters superior to hot-cathode systems whose lifetime at such pressures is very short.

Earlier [4–7], we developed a plasma source of electrons on the basis of a hollow-cathode discharge capable of producing a stable electron beam with a current up to 1 A and energy of the order of 10 keV.

The results of experimental investigations aimed at determining the maximum pressures at which electron beam generation is still possible are presented in this paper. Basic factors limiting the operating pressure are discussed as well.

The plasma source of electrons used in our experiments is shown schematically in Fig. 1. The electron emitting plasma is formed in the discharge chamber, which consists of a hollow copper cathode *1* and a flat hollow anode *2* with an axial emission hole *3* 16 mm in diameter. To stabilize the plasma boundary and screen the accelerating field in the discharge system, a fine metal grid was placed in the anode hole. The mesh dimension varied from 0.25×0.25 to 1.0×1.0 mm. The geometric transparency of the grid was close to 70%. In some cases, the grid was replaced with a perforated electrode with hole sizes close to the grid mesh dimensions. A beam of electrons was extracted through the emitting anode hole by applying a voltage across the accelerating gap between anode *2* and extractor *4*. Pressure was increased by admitting a gas (air) into the working chamber of the setup. This produced equal pressures in the gas-discharge chamber and in the region of beam acceleration and transport. The design of the source and its parameters have been detailed in [4, 7].

The limiting magnitudes of the gas pressure p_m and voltage U_m were registered at the moment immediately preceding a big increase in current I_e (breakdown) in the circuit of the accelerating voltage source. As could be expected, the values of p_m and U_m varied in inverse proportion. The experiments showed that both quantities increased with decreasing mesh dimension h and the distance d between the anode and the extractor, as illustrated in Figs. 2 and 3. The variation of the breakdown voltage with pressure (Fig. 3) is, in fact, the Paschen curve for a constant distance between the elec-

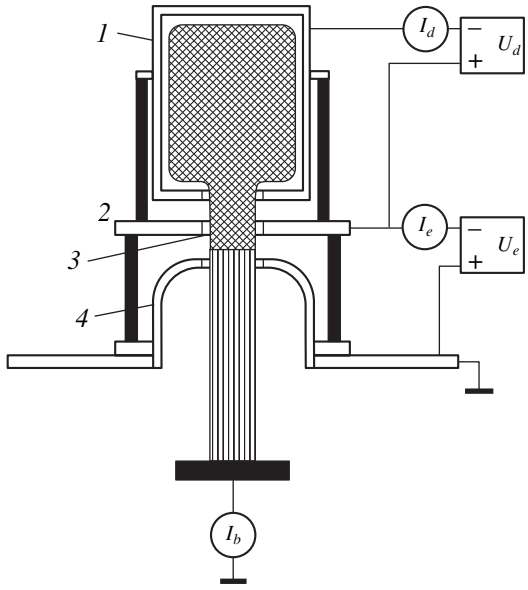

Fig. 1. Diagram of the plasma source of electrons.

1063-7842/01/4602-0179$21.00 © 2001 MAIK "Nauka/Interperiodica"

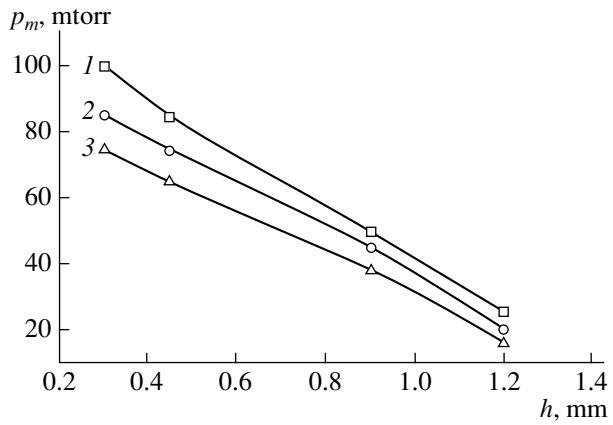

Fig. 2. Variation of the maximum operational pressure with the mesh dimensions of the anode grid. Extracting voltage U_e, kV: (*1*) 4, (*2*) 7, (*3*) 12; discharge current I_d = 500 mA.

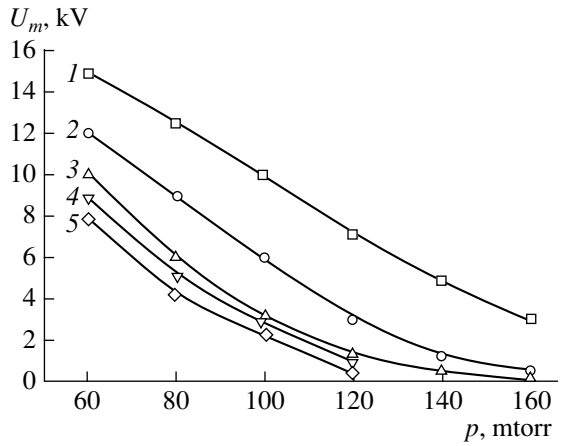

Fig. 3. Variation of the maximum extracting voltage with the gas pressure. Distance d between the anode and the extractor, mm: (*1*) 5, (*2*) 15, (*3*) 25, (*4*) 50, (*5*) 75. The discharge current is 1 A; h = 0.45 mm.

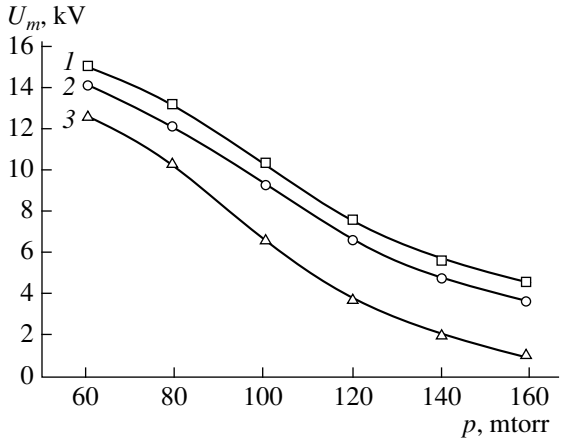

Fig. 4. Same as in Fig. 3 at different values of the discharge current I_d, A: (*1*) 1.0, (*2*) 0.5, (*3*) 0.

trodes. It was expected that the presence of an electron beam in the accelerating gap would facilitate discharge initiation (breakdown) in the gap due to effective ionization of the residual gas by the electron beam. In the experiment, however, the breakdown voltage in the presence of the electron beam in the accelerating gap turned out to be higher compared to the conventional Paschen's breakdown (Fig. 4).

As noted in some papers [8, 9], the basic distinctive feature of the fore-vacuum operating regime of a plasma source of electrons is the high ionization rate of the gas in the accelerating gap producing a backflow of ions, which affects the parameters of the electron emitting plasma in such a way that the plasma density becomes higher. This, in turn, causes an increase in the emission current and further growth of the gas ionization rate. Under certain conditions, an avalanche ionization develops that switches the discharge over from the anode to the extractor. Following this, the accelerating gap voltage drops down to a few tens of volts and the beam collapses.

Growth of the plasma density in the vicinity of the emission hole also shifts the plasma boundary toward the extractor, reducing the extent of the space charge layer in the accelerating gap and, consequently, the effective length of the accelerating gap (the distance between the emission boundary in the plasma and the accelerating electrode). The avalanche growth of the current and the shift of the plasma boundary are prevented from occurring by the grid, which performs its stabilizing function for as long as an increase in the emission current is compensated by the plasma potential growth and corresponding decrease in the emission surface area of the plasma due to expansion of the space charge layer separating the plasma from the grid electrode [10]. Thus, p_m is determined by the combined action of the two processes.

We assume that the basic relaxation mechanism of plasma ions in the accelerating gap is their recharge by gas molecules and that the mechanism of ion escape is diffusion [9]. Then, taking into account the condition of plasma quasi-neutrality and confining ourselves to the one-dimensional case, we obtain a relation between the plasma density $n(0)$ in the emission hole and the current density j_i due to ions exiting the accelerating gap,

$$n(0) = n_0 + \frac{3 j_i Q_i}{e Q_n} \sqrt{\frac{\pi M_i}{8 k T_i}}, \qquad (1)$$

where n_0 is the plasma density in the absence of ion flow; Q_i is the total interaction cross section of slow ions in the plasma; Q_n is the cross section for the recharging of fast ions; and M_i and T_i are the ion mass and the plasma ion component temperature, respectively.

Taking into account known relations for a chaotic electron current from the plasma and for the efficiency

of ionization by fast-moving electrons, we obtain

$$j_e = \frac{en(0)\nu}{4} = \frac{en(0)}{4}\sqrt{\frac{8kT_e}{\pi m}}, \quad (2)$$

$$Z = n_n Q_e d, \quad (3)$$

where j_e is the electron current density at the plasma boundary; ν is the mean thermal velocity of plasma electrons; T_e and m are the electron component temperature and the electron mass, respectively; Z is the number of ions produced by an electron in the accelerating gap; n_n is the concentration of neutral molecules in the accelerating gap; Q_e is the ionization cross section of the gas molecules for interaction with fast electrons; and d is the accelerating gap length.

Then, we can write for the ion current density

$$j_i = n_n Q_e d\frac{en(0)}{4}\sqrt{\frac{8kT_e}{\pi m}}. \quad (4)$$

Substituting Eq. (4) in Eq. (1) and solving the obtained expression with respect to $n(0)$, we get

$$n(0) = n_0\left(1 - n_n Q_e d\frac{3}{4}\frac{Q_i}{Q_n}\sqrt{\frac{MT_e}{mT_i}}\right)^{-1}. \quad (5)$$

Let us determine the thickness of a layer separating the plasma from the grid within a grid opening, assuming that the anode current in the operating regime of the electron source is zero. Under such conditions, the thickness l_l of the layer is

$$l_l = 2\sqrt{\varepsilon_0 U_l^{3/2}(n(0)\sqrt{ekT_e})^{-1}}, \quad (6)$$

where U_l is the potential drop across the layer, which depends on the ratio of the mean thermal velocities of the plasma ions and electrons.

Substitution of Eq. (5) into Eq. (6) gives an expression convenient for qualitative analysis:

$$l_l = 2\sqrt{\varepsilon_0 U_l^{3/2}(n_0\sqrt{ekT_e})^{-1}\left(1 - \frac{3}{4}n_n Q_e d\frac{Q_i}{Q_n}\sqrt{\frac{MT_e}{mT_i}}\right)}. \quad (7)$$

If an approximate equality $l_l \approx h$ is adopted as a criterion of plasma boundary stability, then in the case where $p = n_n kT$ and the dependence of U_i on pressure is weak, the relationships between p_m, h, and d become clear. Increasing either the gas pressure p or distance d causes, according to Eq. (7), a reduction of the layer thickness l_l and, consequently, the plasma boundary stability can be ensured with smaller dimensions h of the grid opening.

To analyze the role of voltage U_e across the accelerating gap, let us turn to expression (1) and note that j_i grows with U_e due to better focusing of the electron and ion beams. Thus, as U_e increases, the plasma density near its boundary increases as well, with all the ensuing consequences. From the decrease in l_l with increasing U_e, we conclude that it is necessary to reduce h to provide the plasma boundary stability.

Enhancement of the electrical strength of the accelerating gap due to the electron beam observed in the experiment (Fig. 4) may be caused by local heating of the gas in the beam region. At constant pressure, heating of the gas reduces the gas density and, consequently, the ion backflow [11]. Another possible reason is overcompensation of the electron beam, which causes the formation of a positive space charge in the accelerating gap [12]. In this case, a substantially nonuniform distribution of the accelerating field potential is realized, which is equivalent to a decrease in the accelerating gap effective length. Under experimental conditions, the operating point in the space of the accelerating gap parameters is on the left branch of Paschen's curve, and the corresponding decrease in the gas density and the accelerating gap length due to the flow of electrons make the electric strength of the gap higher.

Thus, our investigations have demonstrated the possibility of producing an electron beam in a system with a plasma cathode in the fore-vacuum range up to 100 mtorr.

REFERENCES

1. A. A. Ivanov and V. A. Nikiforov, in *Plasma Chemistry* (Atomizdat, Moscow, 1978), Vol. 5, p. 148.
2. Yu. E. Kreĭndel', *Plasma Sources of Electrons* (Énergoatomizdat, Moscow, 1977).
3. E. M. Oks and P. M. Schanin, Phys. Plasmas **7** (5), 1649 (1999).
4. A. V. Mytnikov, E. M. Oks, and A. A. Chagin, Prib. Tekh. Éksp., No. 1, 1 (1998).
5. V. A. Burdovitsin, E. M. Oks, and A. A. Serov, in *Proceedings of the 12th International Conference on High-Power Particle Bearns, Haifa, 1998*, p. 412.
6. V. A. Burdovitsin, A. V. Mytnikov, E. M. Oks, and A. A. Serov, in *Proceedings of the 14th International Symposium on Plasma Chemistry, Prague, 1999*, p. 595.
7. V. A. Burdovitsin and E. M. Oks, Rev. Sci. Instrum. **70** (7), 2975 (1999).
8. Yu. E. Kreĭndel' and V. A. Nikitinskiĭ, Zh. Tekh. Fiz. **41** (11), 2378 (1971) [Sov. Phys. Tech. Phys. **16**, 1888 (1971)].
9. V. A. Gruzdev and Yu. E. Kreĭndel', *Development and Application of Intense Electron Beam Sources* (Nauka, Novosibirsk, 1976), pp. 130–135.
10. V. L. Galanskiĭ, Yu. E. Kreĭndel', E. M. Oks, *et al.*, Zh. Tekh. Fiz. **57** (5), 877 (1987) [Sov. Phys. Tech. Phys. **32**, 533 (1987)].
11. A. Hershcovitch, J. Appl. Phys. **78** (9), 5283 (1995).
12. V. Ya. Martens, Pis'ma Zh. Tekh. Fiz. **12** (13), 769 (1986) [Sov. Tech. Phys. Lett. **12**, 317 (1986)].

Translated by N. Mende

Technical Physics, Vol. 46, No. 2, 2001, pp. 182–185. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 51–54.

Spark Discharges in Condensed Media

N. I. Kuskova

Institute of Pulsed Processes and Technologies, National Academy of Sciences of Ukraine, Nikolaev, 54018 Ukraine
e-mail: kvp@aip.mk.ua

Received March 13, 2000

Abstract—The formalism of phase transition waves is shown to be applicable to the description of both fast and slow discharges in condensed media. The definitions of a streamer and leader are refined, and the mechanisms for their formation and propagation are described. The discharge propagation velocities in low- and high-resistivity media are estimated. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

An analysis of the available experimental data on different types of discharges in condensed media showed that they can be divided into fast and slow discharges. The discharges in low-resistivity media propagate with velocities no higher than the speed of sound and should be classified as slow discharges. Fast discharges are formed in high-resistivity media and propagate with supersonic speeds.

In this paper, we present a unified approach to describing spark discharges as phase transition waves.

PHASE TRANSITION WAVES

The action of strong electric fields on low-conductivity or dielectric condensed media can result in phase transitions such as melting and/or evaporation during the electric or laser breakdown of solid or liquid low-conductivity media such as dielectric–semiconductors or semiconductor–metals in dielectrics or high-resistivity semiconductors. Under certain conditions, the above phase transitions propagate in the form of waves [1–3].

Let us consider phase transition waves whose formation and propagation require both the spatial inhomogeneity of the electric field E and mechanisms providing its propagation in a medium. Let δ_E be the characteristic length of the region in which the electric field is nonuniform and in which a phase transition occurs, and some parameter f of a medium (e.g., the internal energy, density, electric conductivity, or density of current carriers) varies steeply. Then, the velocity of the phase transition wave can be estimated from the expression [2]

$$u \cong \frac{w(E)\delta_E}{\Delta f} + v, \qquad (1)$$

where $\Delta f = |f_{\max} - f_0|$, f_0 is the initial distribution of the parameter f, $w(E)$ is the rate at which this parameter varies, and v is the velocity of thermal expansion or flow of the medium.

When describing the phase transition waves, it is important to know the mechanism for expelling the field from the region where the phase transition occurs and to find the width of the wave front, i.e., the characteristic length of the region in which the field is nonuniform.

STREAMER DISCHARGE IN A CONDENSED MEDIUM

The fact that plasma channels in dielectrics and semiconductors propagate with supersonic velocities exceeding the drift velocities of the current carriers point to the wave nature of the processes in a streamer discharge.

A plasma channel that is formed in a high-resistivity medium in the region with a strong electric field due to impact ionization, photoionization, or generation of current carriers by the electric field and that propagates toward the opposite electrode in the form of an ionization wave will be referred to as a streamer.

Streamer propagation in high-resistivity condensed media exhibits some general features. Discharge is initiated in regions where the electric field is highly nonuniform, e.g., at the interfaces with dielectric inclusions, rough electrode surfaces, and electrodes with small radii. A random distribution of different types of inhomogeneities on the electrode surfaces and in the medium itself results in the stochastic nature of streamer formation and propagation.

In the region where the electric field is strong, the density of current carriers sharply increases. The one-dimensional equation for the current carrier density in Cartesian coordinates has the form

$$\frac{\partial n}{\partial t} - \mu\frac{\partial (nE)}{\partial x} - D\frac{\partial^2 n}{\partial x^2} = w(E) - \frac{n}{\tau_r}, \qquad (2)$$

1063-7842/01/4602-0182$21.00

where $w(E)$ is the rate of electron–hole pair production; τ_r is the recombination time; D is the diffusion coefficient; and μ and n are the electron mobility and density, respectively.

From Eq. (2), we can obtain a rough estimate for the ionization wave velocity,

$$u \cong \left(\frac{w_0}{n_0} - \frac{1}{\tau_r}\right)\delta_E + \frac{D}{\delta_E} + v_0, \tag{3}$$

where $w_0 = w(E_{\max})$, $v_0 = \mu E_{\max}$, and $n_0 = n_{\max}$.

In a strong electric field, the ionization rate substantially exceeds the recombination rate and the drift velocity is much higher than the diffusion velocity. Thus, the ionization wave velocity is determined by the expression

$$u \cong \frac{w_0\delta_E}{n_0} + v_0. \tag{4}$$

The inequality $w(E)/n_0 \gg \tau_r^{-1}$ determines the threshold electric field E^* at which the wave is formed and the length δ_E of the region in which the nonuniform electric field at a given voltage is $E > E^*$. The ionization wave velocity is always higher than the sum of the drift and diffusion velocities of carriers. In the case of intense ionization, it is much higher than this sum. Hence, if a strong electric field is applied to a high-resistivity medium, a streamer propagates with supersonic speed.

The increase in the electron density at the wave front is limited in time by the Maxwellian relaxation time $\tau_m = \varepsilon/\sigma$, where ε is the permittivity and σ is the electric conductivity. Then, the maximum electron density is determined by the expression

$$n_0 = \left(\frac{\varepsilon w_0}{e\mu}\right)^{1/2}, \tag{5}$$

where e is the electron charge.

Correspondingly, the maximum conductivity in the streamer head is $\sigma_0 \cong (e\varepsilon\mu w_0)^{1/2}$. During the same time τ_m, the electric field is expelled to the region in front of the streamer head. Since the conductivity in the streamer is limited and the strong electric field at the streamer front exists for a very short time, the streamer is a weakly ionized low-temperature plasma channel. Consequently, the lattices of solid dielectrics or semiconductors do not melt during the streamer discharge. In gas discharge physics, the low degree of ionization and low temperature of a streamer are used for its definition.

An important property of a streamer is its self-propagation ability, which is associated with the mechanism for expelling a strong electric field from the streamer head. However, the voltage drop along the plasma channel dictates a rapid increase in the applied voltage as the distance passed by the streamer in the divergent field increases. Since the streamer is a weakly ionized plasma channel, the field is not fully expelled from it. Hence, a necessary condition for streamer self-propagation is the formation of a space charge in its head.

Let us consider the processes occurring in the anode region. The electrons are able to leave for the electrode; this effect is related to the so-called field ionization [4]. The higher the electronic work function of the metal and the electric field near the anode, the higher the field ionization probability. The departure of electrons results in the formation of a positive space charge. During the time τ, the space charge will shift by a distance $r = (\mu + \mu^+)E\tau$, where μ^+ is the mobility of the positive current carriers (ions or holes). The space-charge electric field E' is determined from Poisson's equation $\mathrm{div}E' = ep/\varepsilon$, where p is the density of the positive current carriers. The streamer starts propagating if the space-charge electric field reaches the threshold value E^* ($E' \sim E^*$). The time necessary for satisfying this condition can be determined from the above relations by taking into account the rough equality $p \sim \omega_0\tau$. Thus, the time during which the streamer is formed is estimated as

$$\tau = \left(\frac{e}{e(\mu + \mu^+)w(E^*)}\right)^{1/2}. \tag{6}$$

Since no melting or evaporation occurs in a streamer discharge, the slope of the phase trajectory $\Delta P/\Delta T$ on the P–T phase diagram must be larger than the derivative dP/dT taken along the solid–liquid (or liquid–gas for liquid media) phase equilibrium curve:

$$\frac{\Delta P}{\Delta T} < \frac{dP}{dT}. \tag{7}$$

In phase transitions in strong electric fields, the electric field pressure $P = \varepsilon E^2/2$ plays a decisive role. The area on the phase diagram that corresponds to the conditions in the streamer head is characterized by the electric field pressure and the temperature, which increases due to Joule heating. The high local electric field in the streamer head ($E > 10^9$ V/m) corresponds to a pressure of $P > 10^7$ Pa. The rates at which the electric field pressure and the temperature vary are determined by the following expressions:

$$\partial P/\partial t = \varepsilon E\partial E/\partial t \cong \varepsilon E^2/\tau_f, \tag{8}$$

$$\partial T/\partial t = \sigma E^2(\rho c), \tag{9}$$

where τ_f is the time during which the field increases to the amplitude value, ρ is the mass density, and c is the heat capacity.

Since $\Delta P/\Delta T \simeq (\partial P/\partial t)/(\partial T/\partial t)$, we can find an upper limit for the time during which the electric field increases to the threshold value E^*:

$$\tau_f < \frac{\varepsilon\rho c}{\sigma}\frac{dT}{dP}, \tag{10}$$

where the derivative dT/dP characterizes how the phase transition temperature varies with pressure along the phase equilibrium curve.

From the above considerations, it follows that, for a streamer discharge to form, the rise time of the applied voltage pulse must satisfy inequality (10).

One of the most interesting features of a streamer discharge is the crystallographic orientation of plasma channels, whose description sometimes requires rather sophisticated hypotheses. As an example, it was shown that in the most thoroughly studied crystals of zinc selenide and cadmium sulfide, the directions of streamer discharges coincide with the calculated directions of synchronization of microwave and light waves [5, 6]. The microwaves are generated during the propagation of the streamer head as a result of the competition among different processes, which leads to periodic changes (oscillations) in the electric field and plasma density at the streamer front [7].

Although modeling streamer propagation with allowance for the emission of electromagnetic waves is rather complicated, the parameters of a streamer propagating in cadmium sulfide at different voltages can be estimated using formulas (5) and (6). According to the experimental data [1], the characteristic size of the region of a strong electric field is $\delta_E = 5 \times 10^{-6}$ m. Then, the threshold field is $E^* \cong 3.2 \times 10^8$ V/m. For $\delta_E = 2 \times 10^{-6}$ m, we have $E^* \cong 3.5 \times 10^8$ V/m.

The electron drift velocity in cadmium sulfide at $E > 10^8$ V/m is saturated and equal to $v = 10^5$ m/s [1]. For $D \cong 10^{-1}$ m^2/s and $\delta_E = 2 \times 10^{-6}$ m, the calculated minimum velocity is $u_{\min} \cong 1.5 \times 10^5$ m/s. For $\delta_E = 5 \times 10^{-6}$ m, we have $u_{\min} \cong 1.2 \times 10^5$ m/s. These calculated values coincide with the measured values of the minimum velocity [8].

The characteristic features of the streamer discharge in cadmium sulfide deduced from formulas (5) and (6) are as follows. The streamer starts to form at $t \cong 3$ ns at a maximum field of $E \cong 5.4 \times 10^8$ V/m. The strong ionizing field is localized within a small region ($\delta_E = 3 \times 10^{-6}$ m) and affects the crystal during very short time intervals ($t \cong 10^{-13}$ s). The streamer propagation velocity is $u \cong 4 \times 10^6$ m/s, the maximum electron density in the streamer is $n_0 \cong 6 \times 10^{24}$ m^{-3}, and the conductivity in the streamer head is $\sigma \cong 2 \times 10^2$ S/m.

LEADER DISCHARGE IN A CONDENSED MEDIUM

The definition of a streamer allows one to define more accurately the leader discharge. The leader is known to form in a weakly ionized medium during the onset of the ionization–heating instability; i.e., the production of current carriers is governed by Joule heating, whereas the rate at which they are produced is determined by the dependence of the ionization rate on the temperature.

A plasma channel that is formed in a low-conductivity medium due to nonuniform Joule heating of a plasma and propagates toward the opposite electrode in the form of ionization wave will be referred to as a leader. Hence, the leader may propagate along the already formed streamer channel. The leader velocity is also higher than the drift velocity of the current carriers. The leader is a high-temperature plasma channel.

As an illustration, let us consider the breakdown in water in the nanosecond range [9]. For a positive electrode potential, the breakdown in distilled water in the nanosecond range proceeds as follows. Dissociation, which leads to streamer formation within $t < 10$ ns, starts in the local regions with a strong electric field, which appear near microinhomogeneities on the electrode surface. An increase in the temperature in the streamer results in a sharp increase in the probability of thermal ionization $w(T)$; consequently, conductivity continues to rise and the streamer channel transforms into the leader channel.

The minimum time for the streamer formation is $t \cong 4$ ns. There is a narrow peak of the field in the ionization wave propagating from the anode to cathode. This peak is caused by a positive space charge at the streamer head, as well as by its elongated shape. As a result, the field in front of the streamer increases to $E \cong 2 \times 10^9$ V/m for $U = 100$ kV, interelectrode spacing $l = 10^{-2}$ m, and initial conductivity of water $\sigma_0 = 2.76 \times 10^6$ S/m. The field inside the streamer decreases to $E \cong 3 \times 10^5$ V/m as the electric conductivity increases. The streamer propagation velocity is $u \cong 2 \times 10^5$ m/s; for $U = 50$ kV, we have $u \cong 5 \times 10^4$ m/s. These values agree well with estimates obtained using expression (5).

Behind the streamer front, the temperature grows due to intense Joule heating, which causes ionization and further increase in the electric conductivity. The streamer transforms into a leader, whose velocity is less than the propagation velocity of the streamer front. The distribution of the temperature is highly nonuniform both along and across the plasma channel; the temperature of the channel can attain $T \cong 10^4$ K [9].

SLOW SPARK DISCHARGE IN A LOW-CONDUCTIVITY MEDIUM

An electric breakdown in low-resistivity solid or liquid media may develop in a few microseconds after applying the voltage due to the melting and evaporation of the material, which leads to the formation of a gas plasma. In this case, the solid–liquid–gas–plasma phase transitions occur consecutively due to Joule heating in the region where the electric field is nonuniform. The initial scale length on which the electric field varies is determined by the configuration of the discharge gap, which usually ranges from 10^{-6} to 10^{-2} m. The field gra-

dient is limited only by the experimental facilities. The propagation velocity of the phase transition wave can be calculated using the rough formula

$$u \cong \frac{j^2 \delta_E}{\sigma \Delta \varepsilon} + v, \quad (11)$$

where $\Delta\varepsilon$ is the total change in the specific internal energy of a medium at the wave front.

It follows from formula (11) that the propagation velocity of a discharge in a low-resistivity medium can vary over a wide range. However, the relation between the propagation velocity and phase transitions leads to the following restriction that is well known from experiments: the discharge propagation velocity in this case cannot exceed the speed of sound c, because phase transitions in condensed media are accompanied by a change in the volume. Thus, the formation time of a slow discharge is limited from below: $\tau > \delta/c$. The duration of the applied voltage pulse should be at least longer than the discharge formation time.

In slow discharges, the electric field pressure becomes important after the transition of the material to the gaseous state. Since a gas is dielectric, both the electric field strength and electric field pressure inside the produced gas bubble are higher than those in the surrounding low-conductivity medium. The formation of a slow spark discharge is completed by a gas discharge, which results in plasma production. The electric field strength and electric field pressure in the plasma channel decrease, whereas the temperature and gas-kinetic pressure increase.

CONCLUSION

The definitions of a streamer and leader in condensed media are refined. The formation times and propagation velocities of slow and fast discharges are estimated.

A unified approach to describing slow and fast spark discharges in condensed media as phase transition waves is proposed, which allows a quantitative explanation of the difference in the discharge propagation velocities.

It is shown that the medium density changes significantly only in slow discharges, which is related to the liquid–gas phase transition.

REFERENCES

1. N. G. Basov, A. G. Molchanov, A. S. Nasibov, *et al.*, Zh. Éksp. Teor. Fiz. **70**, 1751 (1976) [Sov. Phys. JETP **43**, 912 (1976)].
2. N. I. Kuskova, Pis'ma Zh. Tekh. Fiz. **24** (14), 41 (1998) [Tech. Phys. Lett. **24**, 559 (1998)].
3. V. V. Katin, Yu. V. Martynenko, and Yu. N. Yavlinskiĭ, Pis'ma Zh. Tekh. Fiz. **13**, 665 (1987) [Sov. Tech. Phys. Lett. **13**, 276 (1987)].
4. W. F. Schmidt, IEEE Trans. Electr. Insul. **EI-17**, 478 (1982).
5. V. P. Gribkovskiĭ, A. N. Prokopenya, and K. I. Rusakov, Zh. Prikl. Spektrosk. **60**, 362 (1994).
6. V. P. Gribkovskiĭ, V. V. Parashchuk, and K. I. Rusakov, Zh. Tekh. Fiz. **64** (11), 169 (1994) [Tech. Phys. **39**, 1178 (1994)].
7. V. V. Vladimirov, V. N. Gorshkov, O. V. Konstantinov, *et al.*, Dokl. Akad. Nauk SSSR **305**, 586 (1989) [Sov. Phys. Dokl. **34**, 242 (1989)].
8. A. Z. Obidin, A. N. Pechenov, Yu. M. Popov, *et al.*, Kvantovaya Élektron. (Moscow) **9**, 1530 (1982).
9. V. M. Kosenkov and N. I. Kuskova, Zh. Tekh. Fiz. **57**, 2017 (1987) [Sov. Phys. Tech. Phys. **32**, 1215 (1987)].

Translated by N. Ustinovskiĭ

Technical Physics, Vol. 46, No. 2, 2001, pp. 186–191. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 55–60.

SOLIDS

Effect of the Saturation Conditions and Structure on the Retention of Helium in Structural Materials

A. G. Zaluzhnyĭ and A. L. Suvorov

Institute of Theoretical and Experimental Physics,
Bol'shaya Cheremushkinskaya ul. 25, Moscow, 117259 Russia

Received April 25, 2000

Abstract—A comparison between the kinetics of helium desorption upon linear heating of samples saturated using various regimes is performed, and the effect of dislocations on the retention of helium in materials is estimated. In order to investigate the effect of the conditions of saturation of materials with helium on its retention, samples of austenitic stainless steel 0Kh16N15M3B saturated using various methods were studied, namely, helium irradiation in a cyclotron, in a magnetic mass-separation setup, inside IRT-2000 and BOR-60 reactors, and using the so-called "tritium trick" technique. The investigations show that when saturation of the samples with helium is accompanied by the introduction of radiation defects (in wide limits of helium concentrations and radiation damage), the kinetics of helium evolution from samples of this type is adequate to the kinetics of its evolution from samples irradiated in a reactor. The investigation of the kinetics of helium evolution from the samples of 0Kh16N15M3B steel both after a preliminary deformation and in the process of deformation showed that, in the process of heating, the helium atoms can migrate along dislocation pipes, resulting in a significant effect on the release of helium and its redistribution in the volume of the material. The activation energy for helium pipe diffusion in austenitic steel 0Kh16N15M3B is about 0.7 eV. Mobile dislocations favor the ejection of helium onto the surface of the material, to grain boundaries, interphase interfaces, etc. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

It is known that, under the effect of irradiation, materials can change their physical and mechanical properties. The main factors that restrict the reserve of the material service life are radiation swelling, high-temperature and low-temperature irradiation embrittlement, radiation creep, etc. An important role in these detrimental phenomena belongs to gaseous products of nuclear reactions, including helium.

Since the direct irradiation of materials in a reactor requires much time and material cost, at present, in order to imitate the accumulation of helium, various express methods are used, such as the irradiation of materials with helium ions of various energies, irradiation with high-energy electrons and γ photons, saturation from a plasma, etc. Naturally, the question arises to what extent this or that technique is acceptable for adequately simulating reactor irradiation. In this work, we make an attempt to qualitatively compare the kinetics of helium desorption upon linear heating of samples saturated using various techniques and to estimate the role of dislocations in the retention of helium in materials.

EXPERIMENTAL

The investigation of the kinetics of helium evolution from samples of structural materials in the process of uniform heating was performed in a high-vacuum mass spectrometer setup [1]. To study the effects of the conditions of material saturation with helium on its retention in the material, we investigated samples of an austenitic stainless steel 0Kh16N15M3B saturated with helium using various methods. Upon simulation of the effect of reactor irradiation on the properties of a material, it is insufficient to specify definite values of the helium concentration and its distribution; one should also ensure a specified scaling characteristic of the material damage. This characteristic is determined by the value of a coefficient K, which is expressed as the ratio of the rate of helium formation (He, appm/s) to the rate of production of atomic displacements (dpa/s). Thus, for a number of materials, we have $K = 0.3$ appm/dpa for irradiation in a fast reactor EBR-II and $K = 70$ appm/dpa for irradiation in an HFIR reactor [2].

In order to study the effects of deformation on the retention of helium in the 0Kh16N15M3B steel, we investigated the kinetics of helium evolution from undeformed materials, from materials subjected to a preliminary cold working (to 15 and 50%), and in the process of deformation.

EXPERIMENTAL RESULTS

1. Effect of the Saturation Conditions on the Kinetics of Helium Evolution from Materials

In order to substantiate the methods of express imitation of the accumulation and retention of helium in

structural materials upon reactor irradiation, we performed a comparative analysis of the spectra of gas evolution from samples of 0Kh16N15M3B steel irradiated in a cyclotron [3], in an ILU-100 magnetic mass-separation setup [4], in IRT-2000 and BOR-60 reactors [5], as well as saturated with helium using the "tritium trick" method [6]. Upon irradiation in a cyclotron, the samples were uniformly saturated with helium to concentrations of 1×10^3 at. % throughout the volume. The irradiation temperature did not exceed 400 K.

The saturation of samples in an ILU-100 magnetic mass-separation setup was performed using bombardment of the material studied with α particles with an energy of 70 keV to a fluence of 3×10^{20} ion m^{-2}. The range of α particles with this energy is about 200 nm. The temperature of irradiation did not exceed 400 K. The calculated concentration of helium in the near-surface layer was about 1 at. %.

Irradiation of the samples in an IRT-2000 reactor was performed at temperatures below 420 K to fluences of 1.1×10^{23} neutron m^{-2} for fast neutrons ($E >$ 2.6 MeV) and 2.6×10^{24} neutron m^{-2} for thermal neutrons. The experimentally determined helium concentration in samples irradiated in the IRT-2000 reactor was 3.5×10^{-4} at. %.

As the samples irradiated in the BOR-60 reactor, we used pieces cut out from the can of a fuel element that had been irradiated to a fluence of 7.8×10^{26} neutron m^{-2} ($E > 0.1$ MeV). The irradiation temperature was about 850 K. The samples were studied both in the initial state and after the inner layer that was in contact with the nuclear fuel was etched away. The experimentally determined concentrations of helium in these samples proved to be 2.0×10^{-3} and 1.7×10^{-3} at. %, respectively. The characteristics of the samples are given in the table.

Figure 1 schematically displays the spectra of the rate of helium evolution from the samples studied in the process of uniform heating at a rate of 7 K/min. Figure 1a displays the spectrum of helium evolution from a sample irradiated with α particles in a cyclotron. The identification of the peaks of the helium thermodesorption curve is given in [7], where, along with the study of the kinetics of helium evolution, electron microscopic investigations of the evolution of the dislocation structure and the development of helium-induced porosity were performed. Peak *A* (activation energy $E \approx$ 0.7 eV) corresponds to helium release via pipe diffusion along dislocations that emerge onto the surface; peak *B* ($E \approx 0.8$ eV) corresponds to growth of dislocation loops, their emergence onto the surface, and the evolution of the related helium through dislocation pipes; peak *C* ($E \approx 2.4$ eV) is ascribed to helium diffusion by the vacancy mechanism. The activation energy for the process of helium evolution at temperatures corresponding to peaks *A* and *C*, which are connected with helium diffusion, were calculated using the known method based on the heating of samples at various rates

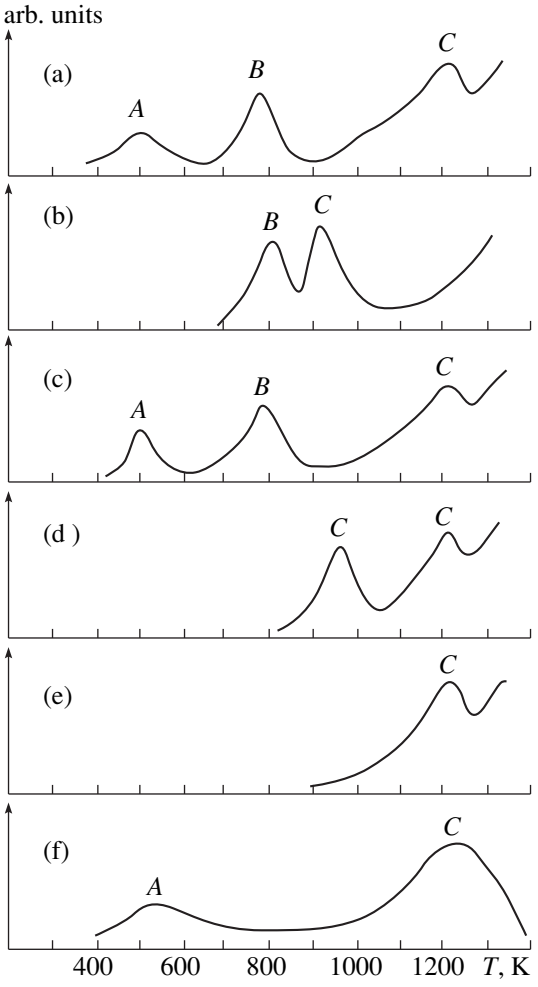

Fig. 1. Spectra of the gas evolution rates upon uniform heating (7 K/min) of samples of 0Kh16N15M3B steel saturated with helium using various techniques: (a) uniform saturation of the entire volume in a cyclotron; (b) irradiation with α particles using a magnetic mass-separation setup (energy of helium atoms 70 keV); (c) irradiation in an IRT-2000 reactor; (d) irradiation in a BOR-60 reactor (as-irradiated state); (e) irradiation in a BOR-60 reactor after etching away a surface layer from the side that was in contact with the fuel; and (f) saturation using the "tritium trick" technique.

[8]. The activation energy of helium evolution at a temperature corresponding to peak *B* was calculated from the rate of growth of interstitial loops [9]. Helium evolution at higher temperatures (at $T > 1200$ K) may be related to the migration of helium–vacancy complexes He_xV_y. These conclusions are confirmed indirectly by the results of recent works concerning the investigation of the evolution and distribution of helium in various materials [10–12].

Note the complete identity of the spectra of the rate of gas evolution from the samples uniformly saturated in a cyclotron (Fig. 1a) and those irradiated in the IRT-2000 reactor (Fig. 1c). The spectrum of gas evolution from the sample irradiated with α particles with an

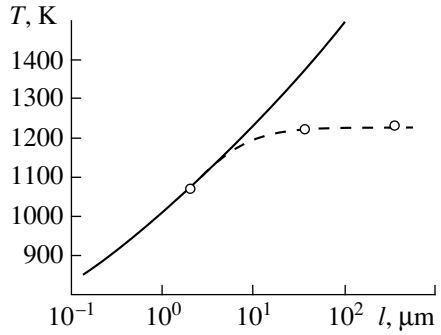

Fig. 2. Variation of the temperature of the maximum rate of helium evolution through the vacancy mechanism upon linear heating of the samples (at a rate of 7 K/min) as functions of the depth of a near-surface layer saturated with helium (solid line, calculated) and the grain size (dashed line, experiment).

energy of 70 keV exhibits peaks *B* and *C* (Fig. 1b). The absence of peak *A* is possibly related to the fact that the related process is already realized during the irradiation due to the near-surface distribution of helium. The shift of peak *C* toward lower temperatures can also be related to the near-surface distribution of helium [13].

The spectra of the rates of helium evolution from the 0Kh16N15M3B steel samples irradiated in the BOR-60 reactor (Figs. 1d, 1e) show that helium is released only at high temperatures exceeding the irradiation temperature. This is likely to be due to the realization of processes responsible for gas evolution at lower temperatures during irradiation. The gas evolution corresponding to the initial sample (Fig. 1d) exhibits two peaks with maxima of the evolution rate at 940 and 1200 K, whereas in the spectrum of helium evolution from the sample with an etched-away inner layer (Fig. 1e), only one peak located at 1200 K is present. This difference in the kinetics of helium evolution from these samples is explained by the different distribution of helium in the material. Thus, in the initial sample the inner layer is enriched in helium by the introduction of α particles from the fuel. In the sample with an etched-away inner surface, helium is distributed uniformly throughout the volume of the sample. The peaks in the curve of helium thermodesorption from the initial sample correspond to the same mechanism of diffusion via vacancies: the first peak is connected with the emergence of helium from the near-surface layer, while the second peak is related to the emergence of helium from the entire volume of the sample. Figure 2 shows the variation of the temperature of the maximum rate of gas evolution (calculated on the basis of [13]) corresponding to helium diffusion by the vacancy mechanism upon linear heating of the sample at a rate of 7 K/min as a function of the thickness *l* of the near-surface layer saturated with helium that determines the gas evolution at this stage. It follows from a comparison of data presented in Fig. 2 with the temperatures of the maxima of the gas-evolution rate (Figs. 1d, 1e) that the thickness of this layer in the initial sample is about 0.5 μm, which agrees in order of magnitude with the range of α particles that are incorporated into the steel from the fuel element [14]. The gas evolution with a maximum at 1200 K, which is controlled by helium evolution from the volume of the sample, is approximately 10 μm, which agrees with the average grain size in the samples studied. It was shown in [15] that grain boundaries are unsaturated traps for inert gases. Therefore, the evolution of inert gases from materials only occurs from near-surface grains of the material. Figure 2 displays the experimentally found temperatures of the maximum rate of helium evolution from samples uniformly saturated with helium in a cyclotron as functions of the depth of evolution and the grain size. This dependence suggests that helium evolution from irradiated materials occurs only from near-surface layers of the samples (in this case, the thickness of this layer is about 10 μm). The rest of the helium is likely to be captured by internal traps such as grain boundaries, dislocations, radiation defects, etc.

Figure 1f displays the curve of the helium evolution rate from a sample saturated using the “tritium trick” technique. This thermodesorption curve has two peaks, which correspond to energies of activation equal to 0.7 and 2.4 eV. This indicates that when saturating the material using a “damage-free” method, such as the “tritium trick” technique, the gas atoms that are retained in the material during the saturation time are

Characteristics of the samples studied

Saturation technique	Energy	Temperature, K	Fluence, m^{-2}	Amount of helium, at. %	Parameter *K* [2]
(*n*, α) [5]	Spectrum of the IRT-2000 reactor	420	2.0×10^{24}	9.5×10^{-4}	400
(*n*, α) [5]	Spectrum of the BOR-60 reactor	850	3.8×10^{26}	2.0×10^{-3} 1.7×10^{-3}	0.6
He* [3]	0–29 MeV	400	1.0×10^{20}	1.0×10^{-3}	4000
He* [4]	70 keV	400	3.0×10^{20}	1.0	
$T \longrightarrow {}^3He$ [6]		300		5.0×10^{-7}	

mainly linked with dislocations and vacancies. The absence of a noticeable helium evolution at temperatures above $0.85T_{mol}$ indicates the absence of stable helium–vacancy complexes.

Thus, we may conclude that, as to the behavior of helium in structural materials, the simulation experiments that ensure the simultaneous introduction into the material lattice of both helium and radiation defects (in wide limits of helium concentrations and radiation damage) sufficiently adequately reproduce reactor irradiation. Most probably, this is due to the stabilization of definite configurations of radiation defects by helium atoms. When helium is introduced by damage-free methods, in each case one should take into account the specific features of its retention and its mobility. The neglect of these features can lead to an incorrect treatment of the results obtained. It has been found that helium evolution upon linear heating of irradiated samples begins only at temperatures exceeding the irradiation temperature. This circumstance can be used to determine the temperature of sample irradiation.

2. Effect of Deformation on the Retention of Helium in Materials

The investigation of the effect of deformation on the retention of helium in materials was performed using samples of 0Kh16N15M3B steel uniformly saturated with the gas as a result of bombardment with helium ions in a cyclotron [3] to a concentration of 1×10^{-3} at. %. The irradiation temperature did not exceed 400 K. The samples were studied in an austenitized state and after a preliminary cold deformation to 15 and 50%. Immediately before the experiment, the samples ($5.0 \times 5.0 \times 0.1$ mm in size) were subjected to electropolishing with subsequent rinsing in alcohol to clean the surface.

Figure 3 displays the curves of helium desorption (in the process of uniform heating at a rate of 7 K/min) from the austenitized and preliminarily deformed samples. It is seen from the figure that, in the case of deformed samples (Figs. 3b, 3c), the low-temperature peak *A*, which is ascribed to helium emergence onto the surface by the mechanism of diffusion along dislocation pipes, becomes higher. The greater the preliminary deformation, the greater the increase in the height of peak *A*. Peak *B*, which is related to the emergence onto the surface of dislocation loops in the process of their growth and to the subsequent evolution of helium by pipe diffusion, becomes somewhat lower. Note that a significant evolution of nitrogen and of a small amount of other chemically reactive gases was also observed in this temperature range, which was likely due to the dissolution of Cottrell atmospheres in the process of the emergence of dislocation loops onto the surface upon their growth. The contribution of this process to helium evolution from steel samples upon heating decreases, as should be expected, with increasing amounts of preliminary deformation (with increasing density of dislocations).

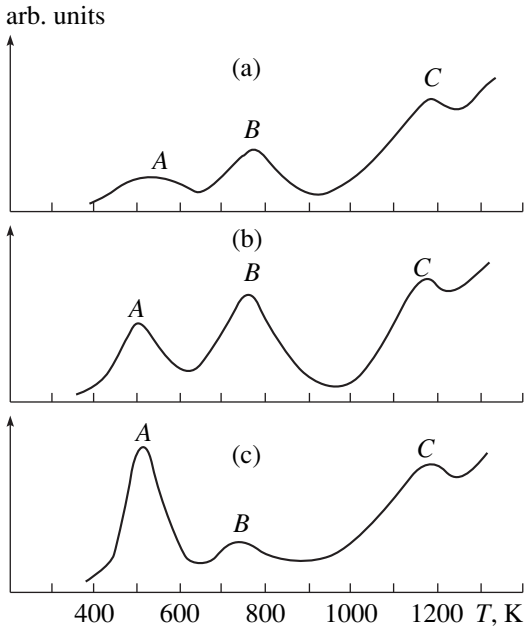

Fig. 3. Spectra of the rates of gas evolution upon a uniform heating of the samples (7 K/min) saturated with helium in a cyclotron: (a) initial state; (b) after a preliminary cold deformation to 15%; and (c) after a preliminary cold deformation to 50%.

Simultaneously, we determined the absolute amounts of helium evolving in the process of isochronous annealings of the samples studied (annealing duration was 0.5 h) using the method of evaporation of the sample in a vacuum with subsequent mass-spectroscopic recording of the helium evolved [5]. These investigations showed that the amount of helium evolved at the low-temperature stage increases with increasing deformation. The total amount of helium evolved in the process of isochronous annealings at temperatures up to 1373 K is approximately the same for all the samples. The results obtained once more confirm the assumption that helium evolution at temperatures of 0.27 to 0.3 T_{mol} is related to the migration of helium atoms along dislocation pipes. The greater the deformation, the greater the number of dislocations that emerge onto the surface of a sample, and the greater the amount of gas, which diffuses along the dislocation pipes, that evolves from the sample.

To further clarify the effect of dislocations on the retention of helium in materials, we studied the release of helium from samples of 0Kh16N15M3B steel in the process of deformation at a temperature of 543 K. This temperature corresponds to the low-temperature peak of helium evolution, which was ascribed by us to the migration of helium atoms along dislocation pipes. To this end, we designed and constructed, at the cover plate of the annealing chamber of the high-vacuum

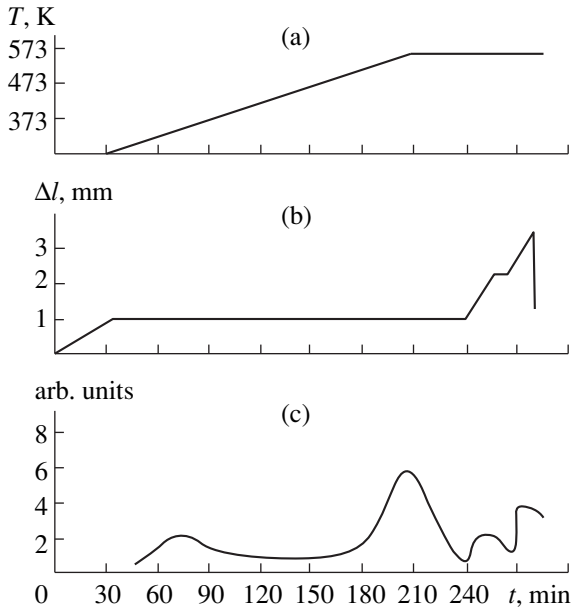

Fig. 4. Deformation-induced helium evolution from samples of 0Kh16N15M3B steel in the process of heating: (a) variation of the temperature; (b) sample elongation; and (c) rate of helium desorption.

mass spectrometric setup for studying the kinetics of gas evolution [1], a special attachment that permitted us to deform the samples studied at various rates in the range from 10^{-5} to 10^{-3} s^{-1}. The samples for the investigation were stamped from a sheet 0.3 mm thick. The samples were 100 mm in length and 4 mm in width. The saturation with helium was performed using irradiation with α particles in a cyclotron [3]. The irradiation temperature did not exceed 400 K. The helium concentration was 10^{-3} at. %.

The experimental procedure was as follows. First, the sample was heated to the test temperature (543 K) and held at this temperature until helium stopped evolving. Then, the sample was deformed at a rate of 10^{-3} s^{-1}. The evolution of helium was observed only at the moment of sample failure. The amount of helium evolved was very small, at the level of the sensitivity of the technique, and was most likely related to its evolution from the fracture surface. The results of this experiment only permit us to make a conclusion on the low probability of interaction of mobile dislocations with helium atoms, which, after irradiation of the material with helium ions in the accelerator and preliminary isothermal annealing at a temperature of 543 K, are most likely located in radiation vacancies, which under these conditions are the main sinks for interstitial helium atoms. Assuming that the interstitial helium atoms interact more efficiently with dislocations, we attempted to "pin" them at dislocations by preliminarily (prior to heating) deforming the irradiated sample at room temperature. In this case the interstitial helium atoms, upon subsequent heating, can be trapped by dislocations, which in turn should manifest itself in the kinetics of its evolution from the sample. Such an experiment was performed in the following way. An irradiated sample was fixed in grips in the annealing chamber and deformed at a rate of 5×10^{-5} s^{-1}. Then, it was heated at a constant rate of 1.5 K/min to a temperature of 543 K, at which the stage of gas evolution was observed (which was ascribed by us to the migration of helium atoms along dislocations), held at this temperature until the gas evolution stopped, and deformed at various rates up to failure. During the experiment, changes in the partial pressure of the evolved helium in the experimental volume of the setup were recorded. The results of the investigation are given in Fig. 4. It is seen that room-temperature deformation is not accompanied by helium evolution. An increase in the temperature to 543 K leads to insignificant gas evolution. The activation energy of this process calculated using the known technique of heating at various rates is E = 0.2 eV [8], which suggests that this peak of helium evolution is related to helium release from the material as a result of its diffusion via the interstitial mechanism [16]. This is related to the fact that after irradiation part of the helium atoms are still located in interstices. After the heating is stopped, gas evolution is observed during 30 min of isothermal holding, which can be ascribed to diffusion of helium atoms along dislocation pipes. Further deformation at a rate of 5×10^{-5} s^{-1} for 20 min was accompanied by helium evolution at a virtually constant rate. A twofold increase in the deformation rate led to an increase in the gas evolution rate. No increase in the gas evolution was observed upon sample fracture.

The results obtained suggest the possibility of both diffusion of helium atoms along dislocation pipes and the entrainment of helium atoms onto the sample surface with mobile dislocations.

Since the deformation of the samples of 0Kh16N15M3B steel leads to an increase in the evolution of helium from the material, we can suggest that deformation at these temperatures may favor the transfer of helium to grain boundaries at temperatures that are lower than the characteristic temperatures of high-temperature radiation embrittlement. This may lead to a degradation of the mechanical characteristics of the material in the case of thermomechanical loading under conditions of generation and incorporation of helium atoms into the surface of materials in combination with pulsed temperature actions, e.g., upon plasma disruption in fusion reactors.

The possibility of pipe diffusion of helium is confirmed by the results of [17], where, in particular, the effect of alternating loading on the kinetics of helium evolution from the A7 alloy (99.7% pure aluminum) was studied. The authors of [17] found that upon the action on the material of an alternating loading (transverse bending of the sample at a frequency of 1 Hz) in the elastic region (the stresses calculated from the mag-

nitude of flexure were ±0.6, ±1.2, ±1.8, and ±2.4 × 10^7 N/m^2); the low-temperature peak of helium evolution (at ~400 K) was shifted upon uniform heating toward lower temperatures in comparison with the state in the absence of a load, and the greater the load, the greater the shift. At the same time, as the stresses increase, the rate of helium evolution increases in this temperature range. The observed changes in the kinetics of gas evolution with increasing mechanical stresses is ascribed to the fact that upon cyclic loads the trapping of helium with dislocations occurs not only during the random walking of helium atoms, but also via their approach to the dislocation due to dislocation bowing. With increasing external load, the sag of the dislocation bowing increases, thereby increasing the probability of trapping of helium atoms by a dislocation and their transfer to the surface. An increase in the load leads to an increase in the absolute amount of helium whose reemission occurs through dislocations.

CONCLUSION

The investigation of the effect of helium introduction into the lattice of a material on the kinetics of gas evolution shows that, as to the behavior of helium in structural materials, simulation experiments that ensure the simultaneous introduction into the material lattice of both helium and radiation defects (in wide limits of helium concentrations and radiation damage) satisfactorily describe the effects of reactor irradiation. Most probably, this is caused by the stabilization of definite configurations of radiation defects by helium atoms. When helium is introduced using damage-free methods, in each case one should take into account the specific features of its retention and its mobility in the material. The neglect of these features can lead to an incorrect treatment of the results obtained. The fact that helium evolution upon heating of irradiated samples begins only at temperatures that exceed the irradiation temperature can be employed to determine the irradiation temperature.

The investigation of the kinetics of helium evolution from the samples of 0Kh16N15M3B steel both after their preliminary deformation and in the course of deformation showed that, in the process of heating, helium atoms can migrate along dislocation pipes, resulting in a significant effect on helium evolution and its redistribution in the volume of the material. The activation energy for pipe diffusion of helium in the austenitic 0Kh16N15M3B steel is about 0.7 eV. Mobile dislocations can favor helium transfer to the surface of the material, grain boundaries, or interphase interfaces.

REFERENCES

1. A. G. Zaluzhnyi, V. P. Kopitin, and M. V. Cherednichenko-Alchevskyi, Fusion Eng. Des., Part B **41**, 129 (1998).
2. V. F. Zelenskiĭ, I. M. Neklyudov, and T. P. Chernyaeva, *Radiation-Induced Defects and Swelling of Materials* (Naukova Dumka, Kiev, 1988).
3. V. F. Reutov and Sh. Sh. Ibragimov, USSR Inventor's Certificate No. 531433, Byull. Izobret., No. 23 (1978).
4. A. G. Zaluzhnyĭ, Yu. N. Sokurskiĭ, N. P. Agapova, *et al.*, At. Énerg. **40** (6), 490 (1976).
5. A. G. Zaluzhnyĭ, M. V. Cherednichenko-Alchevskiĭ, and O. M. Storozhuk, At. Énerg. **52** (6), 398 (1982).
6. V. L. Arbuzov, S. N. Votinov, A. A. Grigor'yan, *et al.*, At. Énerg. **55** (4), 214 (1983).
7. A. G. Zaluzhnyĭ, O. M. Storozhuk, and M. V. Cherednichenko-Alchevskiĭ, Vopr. At. Nauki Tekh., Ser.: Fiz. Radiats. Povrezhdeniĭ Radiats. Materialoved. No. 2, 79 (1988).
8. V. S. Karasev, G. F. Shved, R. V. Grebennikov, *et al.*, At. Énerg. **34** (4), 251 (1973).
9. D. M. Skorov, N. P. Agapova, A. G. Zaluzhnyĭ, *et al.*, At. Énerg. **40** (5), 387 (1976).
10. C. H. Zhang, K. Q. Chen, Y. S. Wang, and J. G. Sun, *Proceedings of the VIII International Conference on Fusion Reactor Materials, 1999*, Part B, p. 1621.
11. A. Ryazanov, H. Matsui, and A. Kazaryan, *Proceedings of the VIII International Conference on Fusion Reactor Materials, 1999*, Part C, p. 356.
12. H. Matsui, M. Tanaka, M. Yamamoto, and M. Tada, J. Nucl. Mater. **191**, 191 (1992).
13. A. G. Zholnin and A. G. Zaluzhnyĭ, Poverkhnost, No. 10, 33 (1986).
14. I. Yu. Abdarashitov, K. V. Botvin, Sh. Sh. Ibragimov, *et al.*, Preprint No. 2-80, IYaF AN KazSSR (Institute of Nuclear Physics, Kazakh Academy of Science, Alma-Ata, 1980).
15. A. G. Zaluzhnyĭ, M. V. Cherednichenko-Alchevskiĭ, O. M. Storozhuk, and A. G. Zholnin, At. Énerg. **56** (5), 314 (1984).
16. F. Charsughi, Dissertation, Jül-2652 (Inst. für Festkörperforschung, Jülich, 1992).
17. A. G. Zaluzhnyĭ, A. P. Komissarov, N. A. Makhlin, *et al.*, Poverkhnost, No. 6, 51 (1983).

Translated by S. Gorin

Technical Physics, Vol. 46, No. 2, 2001, pp. 192–197. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 61–66.

SOLID-STATE ELECTRONICS

Theory of Laterally Nonuniform MOS Transistor under Weak Inversion: A Technique for Determination of Interface Parameters

E. N. Bormontov, M. N. Levin, S. A. Vyalykh, and S. N. Borisov

Voronezh State University, Universitetskaya pl. 1, Voronezh, 394693 Russia

e-mail: levin@lev.csu.ru

Received February 8, 2000

Abstract—The well-known model of current–voltage (*I–V*) characteristics of a MOS transistor (MOST) in weak inversion [1] was modified with regard to lateral nonuniformity of the semiconductor surface potential. A simple technique for determining the fluctuation parameter and the spectral density of interface states from drain-current (output) and drain–gate (transfer) single-threshold *I–V* characteristics is developed. Combined with measurements of the MOST threshold voltage, it makes possible the calculation of the effective oxide charge. The technique is fairly accurate and is useful for IC process control. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

It is known that interface states and lateral nonuniformity (LN) of the semiconductor surface potential strongly affect the drain-current and drain–gate characteristics of a MOST [2]. In an effort to account for the drain–gate curves of a low-voltage CMOS inverter, Swanson and Meindl [3] developed an approximate model of the subthreshold *I–V* characteristics. Overstraeten, Declerck, and Muls proposed a useful method for determination of the interface state density from the MOST subthreshold *I–V* characteristics [1]. They assumed that, under weak inversion conditions, the diffusion component of the current dominates and the semiconductor potential at the interface is uniform. However, most MOS structures are actually characterized by a random distribution of the interface charge density [4–7]. This results in a statistical scatter of the potential y_s. The scatter is usually described by the Gaussian distribution [4]

$$P(y_s) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(y_s-\bar{y}_s)^2}{2\sigma^2}\right], \qquad (1)$$

where $\bar{y}_s$ is the mean value of the potential and σ is its standard deviation.

The nonuniformity of the surface potential substantially increases when the transistor is subjected to ionizing radiation [8–13]. However, a postirradiation decrease in the slope of the drain–gate characteristic is usually associated only with an increase in the interface state density [14–16]. Despite the obvious need for taking into account the lateral nonuniformity of the radiation-induced defect distribution, this has not been done before.

It is the aim of this work to develop a basic model of a weak-inversion-channel MOS transistor [1] with regard for lateral nonuniformity of the device. A solution to this problem would allow designers to correctly treat postirradiation changes in the drain–gate subthreshold characteristics and to elaborate a technique for determining the interface parameters that includes surface potential fluctuations.

MODEL OF A WEAK-INVERSION MOS TRANSISTOR

To be definite, consider a *p*-channel MOST. The bulk potential ϕ_B, surface potential y_s, and drain voltage V_D are negative in this case. The channel potential V varies from 0 at the source to V_D at the drain. The relationship between the gate voltage V_g, surface potential y_s, and potential V at any point of the channel is written as

$$V_g = V_{FB} - \frac{Q_{sc}}{C_{ox}} + \frac{kT}{q}y_s + \frac{qD_{ss}}{C_{ox}}\left(\frac{kT}{q}y_s - V\right). \qquad (2)$$

Here, $V_{FB} = \phi_{ms} - Q_{0t}/C_{ox}$ is the flat-band voltage, ϕ_{ms} is the difference in the work functions between the metal and the semiconductor, Q_{0t} is the fixed oxide charge, C_{ox} is the geometric oxide capacitance, Q_{sc} is the total charge of the space-charge region (SCR) of the semiconductor, and D_{ss} is the spectral density of interface states.

Our fluctuation model of the MOST *I–V* characteristics in the weak inversion range will be based on the model put forward in [1], where the formula for subthreshold current was derived on assumption that,

under weak inversion, the mobile hole charge Q_p in the SCR is much less than the charge of the depletion layer (ionized donor impurity) Q_B; i.e., $Q_{sc} \approx Q_B$. The depletion layer charge Q_B was expanded in a series near the midpoint of the weak inversion range $y_s = 1.5\ln\lambda$, where $\lambda = n_i/N_D$ is the degree of doping of the semiconductor, n_i is the intrinsic concentration of charge carriers, and N_D is the donor concentration. Note that formula (2) incorporates the total space charge Q_{sc}, which includes the charge of mobile minority carriers. It is clear from the physical considerations that the effect of surface potential fluctuations on the inversion layer charge is the strongest. Hence, a more rigorous approach to the problem would be the expansion of Q_{sc} rather than Q_B:

$$Q_{sc} = Q_{sc}^* - \frac{kT}{q}C_{sc}^*(y_s - 1.5\ln\lambda), \tag{3}$$

where Q_{sc}^* and C_{sc}^* are the SCR charge and capacitance, respectively, for the surface potential $y_s = 1.5\ln\lambda$.

Substituting (3) into (2) yields

$$V_g = \left[V_{FB} + 1.5\frac{kT}{q}\ln\lambda - \frac{Q_{sc}^*}{C_{ox}} + \frac{qD_{ss}}{C_{ox}}1.5\frac{kT}{q}\ln\lambda\right] + \frac{kT}{q}(y_s - 1.5\ln\lambda)\left(1 + \frac{C_{sc}^*}{C_{ox}} + \frac{qD_{ss}}{C_{ox}}\right) - \frac{qD_{ss}}{C_{ox}}V. \tag{4}$$

The value in the brackets is the MOST gate voltage V_g^* at which the surface potential equals $1.5\ln\lambda$. Let us introduce the notation [1]

$$n = \frac{C_{ox} + C_{sc}^* + qD_{ss}}{C_{ox}}, \quad m = \frac{C_{ox} + C_{sc}^*}{C_{ox}}. \tag{5}$$

In view of (5), expression (4) can be recast as

$$V_g = V_g^* + \frac{kT}{q}(y_s - 1.5\ln\lambda)n - \frac{qD_{ss}}{C_{ox}}V. \tag{6}$$

Eventually, we obtain the relationship between the surface potential y_s and the gate voltage V_g:

$$y_s = 1.5\ln\lambda + \frac{q}{kT}\frac{V_g - V_g^*}{n} + \frac{q^2}{kT}\frac{D_{ss}}{nC_{ox}}V = 1.5\ln\lambda + \frac{q}{kT}\frac{V_g - V_g^*}{n} + \frac{qV}{kT}\frac{qD_{ss}}{C_{ox} + C_{sc}^* + qD_{ss}}. \tag{7}$$

From Eq. (7), it follows that, at relatively low interface state densities ($qD_{ss} \ll C_{ox}$), the surface potential depends only slightly on V and depends largely on the gate voltage and the semiconductor doping level.

We now proceed to an analysis of the MOST I–V characteristics under weak inversion. In order to find the charge of mobile holes in the channel under weak inversion conditions, we write the expression for SCR total charge Q_{sc} as a function of surface potential y_s and channel potential V [17]:

$$Q_{sc} = \frac{\sqrt{2}\varepsilon_s kT}{qL_D}\left[\exp(y_s) - y_s - 1 + \lambda^2\left(\exp\left(-y_s + \frac{qV}{kT}\right) + y_s - \exp\left(\frac{qV}{kT}\right)\right)\right]^{1/2}, \tag{8}$$

where

$$L_D = \sqrt{\frac{\varepsilon_s kT}{q^2 N_D}}$$

is the Debye screening length.

This expression can be rearranged to the form

$$Q_{sc} = \sqrt{2\varepsilon q N_D \frac{kT}{q}(-y_s - 1)\left[1 + \lambda^2\frac{\exp(-y_s + qV/kT)}{-y_s - 1}\right]} = Q_B\sqrt{1 + \lambda^2\frac{\exp(-y_s + qV/kT)}{-y_s - 1}}. \tag{9}$$

Taking into account that, in the weak inversion range, the charge of mobile holes in the channel is much less than that of ionized donors,

$$\lambda^2\frac{\exp(-y_s + qV/kT)}{-y_s - 1} \ll 1, \tag{10}$$

one can expand expression (9) into a series, leaving only zero- (Q_B) and first-order (Q_p) terms:

$$Q_{sc} = Q_B\left[1 + \frac{\lambda^2}{2}\frac{\exp(-y_s + qV/kT)}{-y_s - 1}\right] = Q_B + Q_p, \tag{11}$$

where the mobile hole charge Q_p is described by the expression [1]

$$Q_p(y_s, V) = \left(\frac{q^2\varepsilon_s N_D}{2kT(-y_s - 1)}\right)^{1/2}\frac{kT}{q} \times \exp\left(2\ln\lambda + \frac{q}{kT}V - y_s\right) = C_d(y_s)\frac{kT}{q}\exp\left(2\ln\lambda + \frac{q}{kT}V - y_s\right). \tag{12}$$

Here, $C_D(y_s)$ is the capacitance of the depletion layer between the inversion layer and the quasi-neutral interior of the semiconductor.

If the surface potential is nonuniform, the SCR (depletion layer) capacitance C_D must be averaged over the surface potential, which is distributed according to (1). As a result, we come to the expression for C_D as a func-

tion of the mean surface potential $\bar{y}_s$ [18]:

$$C_D(\bar{y}_s) = \frac{1}{\sqrt{2\pi}\sigma}\int_{\bar{y}_s-3\sigma}^{\bar{y}_s+3\sigma}\left(\frac{q^2\varepsilon_s N_D}{2kT(-y_s-1)}\right)^{1/2} \times \exp\left[-\frac{(y_s-\bar{y}_s)^2}{2\sigma^2}\right]dy_s. \quad (13)$$

Similarly, the total SCR capacitance, through which the parameters m and n in [18] are defined, is found by averaging the total capacitance over the surface potential:

$$C_{sc}^* = \frac{\varepsilon_s}{\sqrt{2}L_D}\frac{1}{\sqrt{2\pi}\sigma}\int_{1.5\ln\lambda-3\sigma}^{1.5\ln\lambda+3\sigma} \times\left|\frac{\exp(y_s)-1-\lambda^2(\exp(-y_s)-1)}{\sqrt{\exp(y_s)-y_s-1+\lambda^2(\exp(-y_s)+y_s-1)}}\right| \times \exp\left[-\frac{(y_s-1.5\ln\lambda)^2}{2\sigma^2}\right]dy_s. \quad (14)$$

The average surface potential and the gate voltage are related by Eq. (7). This equation contains the voltage V_g^*, which has the following form for an LN MOST:

$$V_g^* = V_{FB} + 1.5\frac{kT}{q}\ln\lambda - \frac{Q_{sc}^*}{C_{ox}} + \frac{qD_{ss}}{C_{ox}}1.5\frac{kT}{q}\ln\lambda, \quad (15)$$

where Q_{sc}^* is calculated from the formula [18]

$$Q_{sc}^* = \frac{\sqrt{2q}N_D L_D}{\sqrt{2\pi}\sigma}\int_{1.5\ln\lambda-3\sigma}^{1.5\ln\lambda+3\sigma}[(\exp y - y - 1) + \lambda^2(\exp(-y)+y-1)]^{1/2}\exp\left[-\frac{(y-1.5\ln\lambda)^2}{2\sigma^2}\right]dy. \quad (16)$$

Thus, with surface potential fluctuations taken into consideration, the charge of mobile holes in the channel under weak inversion can be given in a form similar to (12):

$$Q_p(\bar{y}_s, V) = C_D(\bar{y}_s)\frac{kT}{q}\exp\left(2\ln\lambda + \frac{q}{kT}V - \bar{y}_s\right). \quad (17)$$

In view of (7), formula (17) can be recast as

$$Q_p = C_d(\bar{y}_s)\frac{kT}{q}\exp(0.5\ln\lambda)\exp\left[-\frac{q}{kT}\frac{V_g-V_g^*}{n}\right] \times \exp\left[\frac{qV}{kT}\left(1-\frac{qD_{ss}}{C_{ox}+C_{sc}^*+qD_{ss}}\right)\right] \quad (18)$$

$$= C_D(\bar{y}_s)\frac{kT}{q}\sqrt{\frac{n_i}{N_D}}\exp\left[-\frac{q}{kT}\frac{V_g-V_g^*}{n}\right]\exp\left[\frac{qV}{kT}\frac{m}{n}\right].$$

Although the drain current I_D in the weak inversion range is of a diffusion character, it can be calculated from the general expression

$$I_D = -\frac{Z}{L}\mu_p\int_0^{V_D}Q_p(V)dV, \quad (19)$$

as shown in [19]. Here, Z and L are the channel width and length, respectively, and μ_p is the hole mobility in the channel.

Substituting expression (18) into (19), we obtain the final expression for drain current in the subthreshold operating range of an LN MOS transistor:

$$I_D = \frac{Z}{L}\mu_p\frac{n}{m}\left(\frac{kT}{q}\right)^2\sqrt{\frac{n_i}{N_D}}C_D(\bar{y}_s) \times \exp\left[-\frac{q}{kT}\frac{V_g-V_g^*}{n}\right]\left[1-\exp\left(\frac{qV_D}{kT}\frac{m}{n}\right)\right], \quad (20)$$

where the capacitance $C_D(\bar{y}_s)$ is given by (13) and m and n, by (5).

Under weak inversion and for low interface state densities, the mean surface potential $\bar{y}_s$ corresponding to a given gate voltage V_g is determined by the expression

$$\bar{y}_s = 1.5\ln\lambda + \frac{q}{kT}\frac{V_g-V_g^*}{n}. \quad (21)$$

Thus, formula (20) in combination with (13) and (21) allows one to calculate the MOST *I–V* characteristics under weak inversion with regard for the LN of the surface potential. The form of expression (20) is similar to that of the earlier relationship for subthreshold characteristics [1], but the parameters appearing in these expressions are different. For example, instead of the depletion layer capacitance C_D, the parameters m and n in expression (20) contain the total SCR capacitance C_{sc} including the inversion layer capacitance. In addition, all of the capacitances in (20) are averaged over the surface potential.

The basic static parameter of a MOST, the threshold voltage V_T, is calculated with consideration for surface potential fluctuations by the formula

$$V_T = V_{FB} + 2\frac{kT}{q}\ln\lambda - \frac{Q_{scT}}{C_{ox}} + \frac{qD_{ss}}{C_{ox}}2\frac{kT}{q}\ln\lambda, \quad (22)$$

where the threshold SCR charge Q_{scT} corresponding to

the mean surface potential $2\ln\lambda$ is given by [18]

$$Q_{scT} = \frac{\sqrt{2}qN_DL_D}{\sqrt{2\pi}\sigma}\int_{2\ln\lambda-3\sigma}^{2\ln\lambda+3\sigma}[(\exp y - y - 1) + \lambda^2(\exp(-y) + y - 1)]^{1/2}\exp\left[-\frac{(y-2\ln\lambda)^2}{2\sigma^2}\right]dy. \quad (23)$$

DETERMINATION OF INTERFACE PARAMETERS

Expression (20) may be recast in the more compact form

$$I_D = I_0\exp\left[-\frac{q}{kT}\frac{V_g - V_g^*}{n}\right]\left[1 - \exp\left(\frac{qV_D}{kT}\frac{m}{n}\right)\right], \quad (24)$$

where

$$I_0 = \frac{Z}{L}\mu_p\frac{n}{m}\left(\frac{kT}{q}\right)^2\sqrt{\frac{n_i}{N_D}}C_D(\bar{y}_s). \quad (25)$$

If the MOST drain–gate characteristic is plotted in the semilogarithmic coordinates $\ln I_D - qV_g/kT$, its slope is equal to

$$\frac{kT}{q}\frac{\partial\ln I_D}{\partial V_g} = \tan\alpha_g = -\frac{1}{n}. \quad (26)$$

It should be noted that the parameter n is proportional to the characteristic subthreshold voltage $S = \partial V_g/\partial\log I_D$, which changes the drain current by one order of magnitude. Using formula (26), one obtains

$$S = \frac{\partial V_g}{\partial\log I_D} = \frac{\partial V_g}{\partial\ln I_D}\ln 10 = \frac{kT}{q}n\ln 10. \quad (27)$$

As follows from expression (24), as the drain voltage grows, the drain current tends to the value

$$I_{D\max} = I_0\exp\left[-\frac{q}{kT}\frac{V_g - V_g^*}{n}\right], \quad (28)$$

which is called subthreshold saturation current. Formula (24) can thus be recast as

$$\ln\left(1 - \frac{I_D}{I_{D\max}}\right) = \frac{qV_D}{kT}\frac{m}{n}. \quad (29)$$

If the MOST subthreshold drain-current characteristic is plotted in the coordinates

$$\ln\left(1 - \frac{I_D}{I_{D\max}}\right) - \frac{qV_D}{kT},$$

its slope is equal to

$$\tan\alpha_D = \frac{m}{n}. \quad (30)$$

The parameter m can be found by dividing expression (30) by (26). Using definition (5), one can obtain an experimental value of the SCR capacitance in the midpoint of the weak inversion range by the formula $C_{sc}^* = (m-1)C_{ox}$. For a given dopant concentration N_D, the capacitance C_{sc}^* depends only on the fluctuation parameter. Therefore, using the analytical C_{sc}^* vs. σ dependence [formula (14)] and the experimental C_{sc}^* value, one can calculate the fluctuation parameter σ. Nomograms illustrating the C_{sc}^* vs. σ dependences for different dopant concentrations in the semiconductor are presented in Fig. 1. They allow us to find the fluctuation parameter.

If m and n are known, the spectral density D_{ss} of interface states can be calculated. In fact, from the definitions of m and n [formulas (5)], we have

$$D_{ss} = \frac{C_{ox} + C_{sc}^*}{q}\left(\frac{n}{m} - 1\right) = \frac{C_{ox} + (m-1)C_{ox}}{q} \times\left(\frac{1}{\tan\alpha_D} - 1\right) = \frac{\tan\alpha_D}{|\tan\alpha_g|}\frac{C_{ox}}{q}\left(\frac{1}{\tan\alpha_D} - 1\right). \quad (31)$$

The density of interface states from formula (31) corresponds to the mean surface potential $\bar{y}_s = 1.5\ln\lambda$, i.e., to the midpoint of the weak inversion range. Since the energy dependence of the interface state density is weak in this range [1, 4], the density can be set constant

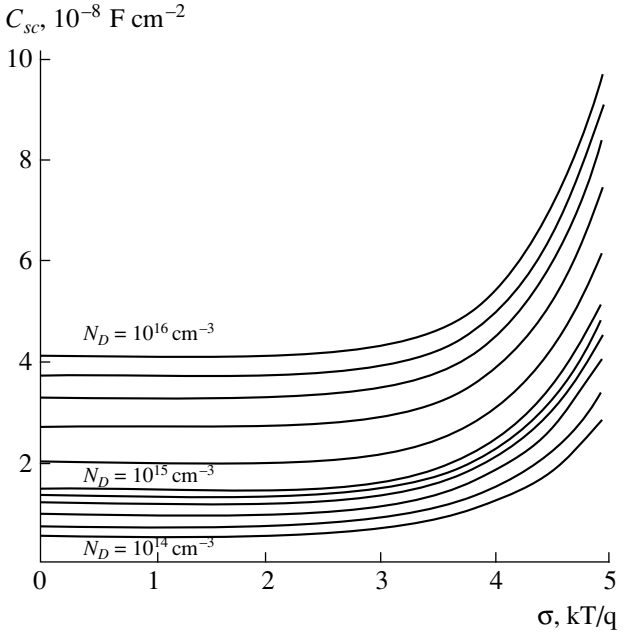

Fig. 1. SCR capacitance in the midpoint of the weak inversion range C_{sc}^* vs. fluctuation parameter σ nomograms. N_D is the dopant concentration in the MOST substrate.

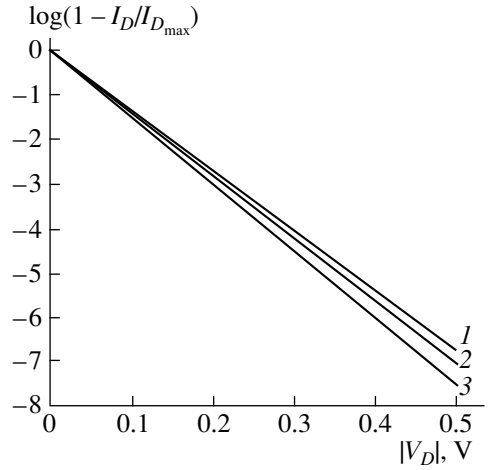

Fig. 2. Output characteristics of the *p*-channel MOST under weak inversion. Gate voltage V_g = –2 V, dopant concentration in the substrate $N_D = 10^{15}$ cm^{-3}, oxide thickness d_{0x} = 200 nm, spectral density of interface states $D_{ss} = 5 \times 10^{10}$ cm^{-2} eV^{-1}, flat-band voltage V_{FB} = –1 V, and channel width-to-length ratio Z/L = 1. The fluctuation parameter σ = (*1*) 3, (*2*) 4, and (*3*) 5.

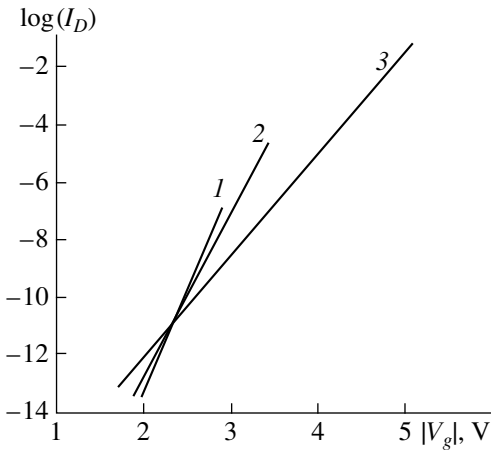

Fig. 3. Transfer characteristics of the *p*-channel MOST under weak inversion. The gate voltage V_D = –1 V. The other parameters are the same as in Fig. 2.

with a high accuracy. Hence, the value of D_{ss} found can be extended to the entire subthreshold range.

Thus, from the slopes of the MOST drain–gate and the drain-current characteristics plotted in semilogarithmic coordinates, the fluctuation parameter σ and the spectral density D_{ss} of interface states are found.

With D_{ss} and σ known and the threshold voltage V_T determined (for example, by extrapolating the $I_D^{1/2}(V_g)$ curve in the strong inversion region to the V_g axis [17]), the effective oxide charge Q_{0t} can be calculated. Indeed, in view of the relationship $V_{FB} = \phi_{ms} - Q_{0t}/C_{ox}$, from formula (22) we obtain

$$Q_{0t} = -Q_{scT} - C_{ox}\left(V_T - \phi_{ms} - 2\frac{kT}{q}\ln\lambda\right) + 2qD_{ss}\frac{kT}{q}\ln\lambda, \quad (32)$$

where the space charge Q_{scT} is given by formula (23).

Note that, in our method, the MOST interface parameters have been obtained without using the assumption [14] that the interface state charge is equal to zero in the "midgap."

RESULTS AND DISCUSSION

Results of simulation of the drain-current and drain–gate characteristics under weak inversion are presented in Figs. 2 and 3. The fluctuation parameter σ was varied from 0 to 5; dopant concentration N_D, from 10^{14} to 10^{16} cm^{-3}; and interface state density D_{ss}, from 0 to 10^{11} cm^{-2} eV^{-1}.

It is seen that the slope of the $\ln(1 - I_D/I_{D\max})$ vs. V_D curve (i.e., the value of m/n) increases with σ, the increase becoming significant at σ > 3 (Fig. 2). At the same time, the slope of the $\ln I_D$ vs. V_g curve, i.e., the value of $1/n$, remains almost constant at σ < 3 but significantly decreases with growing σ for σ > 3 (Fig. 3). The effect of σ on the threshold voltage shift $\Delta V_T = V_T(\sigma) - V_T(\sigma = 0)$ is illustrated in Fig. 4. When σ > 2, the threshold voltage starts to grow markedly. Note also that the higher the dopant concentration in the substrate N_D, the faster the V_T shift increases with σ (Fig. 4). The effect of surface potential fluctuations on the characteristic voltage S is shown in Fig. 5. At σ > 2, S also grows rapidly because of an increase in the SCR capacitance C_{sc}^*.

It should be noted that the slope of the drain–gate characteristics decreases with an increase in both interface state density D_{ss} and σ. In addition, interface states shift the MOST threshold voltage, increasing its absolute value.

Let us apply our technique to determine the interface parameters from experimental drain-current and drain–gate characteristics of a test *p*-channel MOST. The transistor was fabricated on a KÉF-4.5 silicon substrate (resistivity 4.5 Ω cm, phosphorus concentration 10^{15} cm^{-3}). The gate oxide (SiO_2) thickness was d_{ox} = 40 nm. A high LN of the surface potential was achieved by Co_{60} γ irradiation (energy of quanta $E_\gamma \sim 1.2$ MeV). Plotted in semilogarithmic coordinates, the experimental drain-current and drain–gate characteristics had clearly defined linear portions. Their slopes were $\tan\alpha_D = m/n = 0.89$ and $\tan\alpha_g = -1/n = -0.75$. Using these values, we calculated the SCR capacitance in the midpoint of the weak inversion range, $C_{sc}^* = (m-1)C_{0x} =$

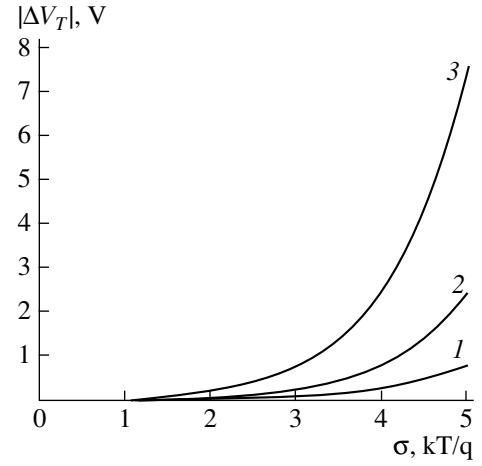

Fig. 4. Threshold voltage shift ΔV_T due to surface potential fluctuations vs. σ. $d_{0x} = 200$ nm; $N_D = (1)\ 10^{14}$, $(2)\ 10^{15}$, and $(3)\ 10^{16}$ cm^{-3}.

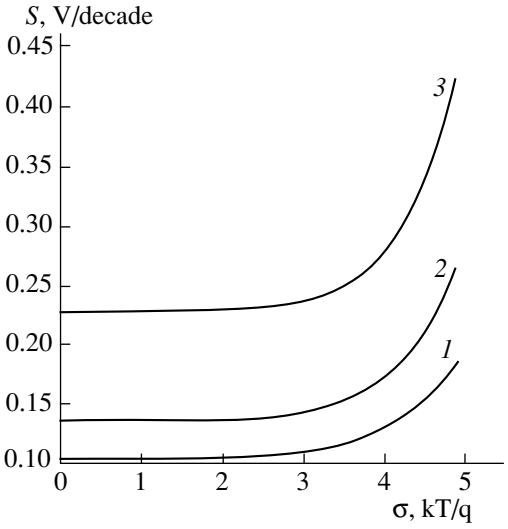

Fig. 5. Characteristic subthreshold voltage S vs. σ. (*1–3*) The same as in Fig. 4.

1.57×10^{-8} F/cm^2, and then (from the corresponding nomogram in Fig. 1) the fluctuation parameter $\sigma = 2.8$. The spectral density of interface states calculated from formula (31) was found to be $D_{ss} = 7.7 \times 10^{10}$ cm^{-2} eV^{-1}.

From the $I_D^{1/2}$ (V_g) curve in the strong inversion range, we also estimated the threshold voltage: $V_T = -1.6$ V. Finally, substituting the values of V_T, D_{ss}, and σ into formulas (23) and (32), we calculated the effective oxide charge $Q_{0t} = 6.15 \times 10^{-8}$ C/cm^2.

Thus, from static subthreshold *I–V* characteristics, one can estimate the MOST charge parameters having regard to the LN of the oxide–semiconductor interface. The method is useful for IC process control.

REFERENCES

1. R. J. van Overstraeten, G. J. Declerck, and P. A. Muls, IEEE Trans. Electron Devices **ED-22** (5), 282 (1975).
2. R. J. van Overstraeten, G. J. Declerck, and G. Broux, IEEE Trans. Electron Devices **ED-20** (12), 1154 (1973).
3. R. M. Swanson and J. D. Meindl, IEEE J. Solid-State Circuits **SG-7** (4), 140 (1972).
4. E. H. Nicollian and A. Goetzberger, Bell Syst. Tech. J. **46** (5), 1055 (1967).
5. J. R. Brews, J. Appl. Phys. **46** (5), 2181 (1975).
6. C. Werner, H. Bernt, and A. Eder, J. Appl. Phys. **50** (11), 7015 (1979).
7. K. Zeigler and E. Klausmann, Appl. Phys. Lett. **28** (11), 678 (1976).
8. *Ionizing Radiation Effects in MOS Devices and Circuits*, Ed. by T. P. Ma and P. V. Dressendorfer (Wiley, New York, 1989).
9. H. Terletzki, A. Boden, F. Wulf, and W. R. Fahrner, Phys. Status Solidi A **86** (8), 789 (1984).
10. R. K. Freitag, E. A. Burke, and D. B. Brown, IEEE Trans. Nucl. Sci. **34** (6), 1172 (1987).
11. N. Saks and M. G. Ancona, IEEE Trans. Nucl. Sci. **34** (6), 1348 (1987).
12. M. A. Xapsos, R. K. Freitag, C. M. Dozier, and D. B. Brown, IEEE Trans. Nucl. Sci. **37** (6), 1671 (1990).
13. R. K. Freitag, E. A. Byrke, C. M. Dozier, and D. B. Brown, IEEE Trans. Nucl. Sci. **35** (6), 1203 (1988).
14. P. J. McWhorter and P. S. Winokur, Appl. Phys. Lett. **48** (2), 133 (1986).
15. D. M. Fleetwood, M. R. Shaneyfelt, J. R. Schwank, *et al.*, IEEE Trans. Nucl. Sci. **36** (6), 1816 (1989).
16. Z. Shanfield and M. M. Moriwaki, IEEE Trans. Nucl. Sci. **34** (6), 1159 (1987).
17. S. Sze, in *Physics of Semiconductor Devices* (Wiley, New York, 1981; Mir, Moscow, 1984), Vol. 2.
18. E. N. Bormontov and S. V. Lukin, in *Proceedings of the 5th International Conference on Simulation of Devices and Technologies, Obninsk, 1996*, p. 35.
19. R. J. van Overstraeten, G. J. Declerck, and G. Broux, IEEE Trans. Electron Devices **ED-20** (12), 1150 (1973).

Translated by V. Isaakyan

Technical Physics, Vol. 46, No. 2, 2001, pp. 198–201. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 67–70.

SOLID-STATE ELECTRONICS

Laser-Induced Implantation and Diffusion of Magnesium into Silicon

V. M. Arutyunyan, A. P. Akhoyan, Z. N. Adamyan, and R. S. Barsegyan

Yerevan State University, ul. A. Manukyana 1, Yerevan, 375049 Armenia

e-mail: vladimir@www.physdep.r.am (Aroutiounian)

Received April 21, 2000

Abstract—Laser-induced diffusion ("implantation") of magnesium atoms into silicon was studied experimentally. Neodymium-glass laser irradiation (λ = 1.06 μm, $\tau \sim 0.4$ ms) was found to increase the diffusion coefficient and solubility of magnesium in silicon. Current–voltage and capacity–voltage characteristics, as well as thermostimulated current spectra of ⟨Si + Mg⟩ crystals, were obtained. © *2001 MAIK "Nauka/Interperiodica".*

Infrared (IR) detectors operating in the atmosphere transmission ranges (3–5 and 8–14 μm) are still a focus of interest today. The characteristics and efficiency of semiconductor IR detectors markedly depend on the properties that an injected impurity imparts to the semiconductor. The most important parameters in this respect are impurity levels in the energy gap, cross section of carrier capture by these levels (hence, electron and hole lifetimes), photoionization cross section, and electroactive impurity concentration [1].

Due to high manufacturability and high concentration of electroactive impurity atoms, deep- and shallow-impurity silicon has recently started to play a significant role as a basic material for photodetectors operating in the range indicated above. However, in spite of the advances in silicon technology, doping of silicon by certain impurities with traditional methods is often a challenge. Therefore, modern semiconductor electronics widely employs photonic technologies. Laser irradiation is one of them [2–5].

It is well known that magnesium is a suitable dopant for silicon infrared detectors operating in the 8–14 μm range. Being an interstitial impurity, magnesium behaves as a helium-like double donor with ionization energies E_c = 0.107 and 0.25 eV for the neutral (Mg^0) and singly ionized (Mg^+) donor states, respectively [6–8].

Hall measurements have revealed another four shallow donor levels with ionization energies of 0.04, 0.055, 0.08, and 0.093 eV [8]. The 0.055 and 0.093 eV levels have also been discovered from photoconductivity spectra [9]. The presence of the latter level, as well as of the deep levels indicated above, has been confirmed by theoretical study [10]. Hall measurements performed in [7] on Mg-doped *n*-silicon have indicated two plausible electrically different states of magnesium ions. The first is the amphoteric state with the acceptor level E_c = 0.115 eV, and the second is the donor level E_c = 0.227 eV.

Reproducibility, uniformity, and degree of doping are strongly dependent on diffusion source, diffusion annealing temperature, and cooling rate. However, the small segregation coefficient and the high vapor pressure of magnesium at the melting temperature of silicon greatly hamper the introduction of Mg into silicon during growth. This has fostered a search for new techniques of doping silicon by magnesium.

It is well known that laser irradiation (with an energy density $W \sim 1$ J/cm^2) of semiconductor materials gives rise to nonequilibrium effects that are characterized by enhanced migration of impurity atoms during short laser pulses [2].

In this paper, we report on laser-induced "implantation" of magnesium into silicon with a view to producting Si⟨Mg⟩-based IR detectors. As a starting material, we took floating-zone *p*-Si(111) wafers of thickness less than 1.0 mm and resistivity $\rho \sim 20$–40 kΩ cm. Chemically cleaned wafers were coated by a magnesium film evaporated at a pressure of 4×10^{-5} mm Hg. Properly selected deposition rates and substrate temperature (T = 250°C) provided for good adhesion of a ~0.2-μm-thick deposited layer.

Then, the covered surface was irradiated by a focused beam from a Nd-glass free-running laser (wavelength λ = 1.06 μm, pulse width $\tau \sim 0.4$ ms). Laser fluence was chosen according to the wafer surface temperature evaluated from the dependence of the temperature on the pulse energy and exposure time [2, 3, 11]. At the fluence $W \sim 1.5$ J/cm^2, the surface is heated up to the MgSi eutectic temperature, ~950°C for 36.61 wt % Si [12]. Reflowing of the wafer surface was clearly observed under a microscope. Note that, at smaller laser fluences, Mg_2Si eutectic with 58 wt % Si (the eutectic temperature is about 637°C [12]) did not form. The beam scanned only half the wafer surface by moving the objective table in two perpendicular directions. The other half of the wafer remained intact. After irradiation, which most likely resulted in the formation

of MgSi eutectic, the wafers were placed into a diffusion chamber with a continuous flow of inert gas and standard thermal diffusion for 10 h at $T = 1200°C$ was carried out.

After diffusion, the wafers were quickly cooled by immersion in water at room temperature. Then, the samples were mechanically and chemically treated. It was found that the Mg-coated side of the wafer changed the conductivity type from *p* to *n* over the entire surface. On the back side of the wafers, the conductivity type remained the same as before diffusion (i.e., *p*-type). To measure the diffusion depth of Mg atoms, a skew metallographic section was made and the conductivity type across the wafer was determined. The diffusion depth of Mg atoms estimated from the *n–p* junction position was found to be $L^* \sim 650$ and $L \leq 500$ μm in the irradiated and unirradiated parts of the wafer, respectively (here, L^* and L are the depths of the *n–p* junction in the respective parts of the surface). Along with these measurements, a four-point probe method was used to determine the sample resistivity, which turned out to be smaller in the irradiated part (Fig. 1).

An extension of the doped layer and an increase in the conductivity of the irradiated part can be explained as follows. (1) Laser irradiation causes magnesium silicides, for example, MgSi, to form on the surface. They may serve as an infinite diffusion source [13]. (2) Laser irradiation produces a liquid phase on the surface, where the impurity diffusion coefficient is considerably larger than in single-crystal (solid) silicon; assuming that the extension of the magnesium-doped layer is fully accounted for by the increased diffusion coefficient in the liquid phase and using the relationship

$$L^* - L = \sqrt{D\tau}$$

(where D is the diffusion coefficient and τ is the pulse duration), one obtains a diffusion coefficient in the melt that exceeds the known values by several orders of magnitude. (3) Irradiation generates defects (for example, silicon vacancies), which not only accelerate diffusion [14, 15], but also increase the ultimate solubility of electroactive magnesium atoms [4]. In addition, athermic, photoinduced diffusion is not ruled out either [16].

To display the advantages and characteristic features of this laser technique, we studied current–voltage (*I–V*) and capacitance–voltage (*C–V*) characteristics, as well as thermostimulated current (TSC) spectra, of laser-irradiated silicon wafers covered by a magnesium film (⟨Si + Mg⟩) for different laser fluences. To perform these measurements, we made alloyed Au + 1% Sb and Al contacts to the *n*- and *p*-type sides of the samples, respectively.

As seen from the *I–V* characteristics shown in Fig. 2, irradiation with a laser-fluence $W = 1.8$ J/cm^2 produces a rectifying structure. With an increase in the laser fluence, the rectifying properties deteriorate, which may be due to either the formation of magnesium silicides (Mg_2Si or MgSi) at $W = 1.8$–2.5 J/cm^2 or magnesium evaporation from the near-surface region at high fluences (at $W > 2.5$ J/cm^2, silicon melts). The scatter in the voltages at which the forward current sharply rises is probably associated with the formation of magnesium silicides with different potential barrier heights. The absence of reverse saturation current in Fig. 2 is probably related to current leak across the junction because of a spread in the beam intensity over the surface area.

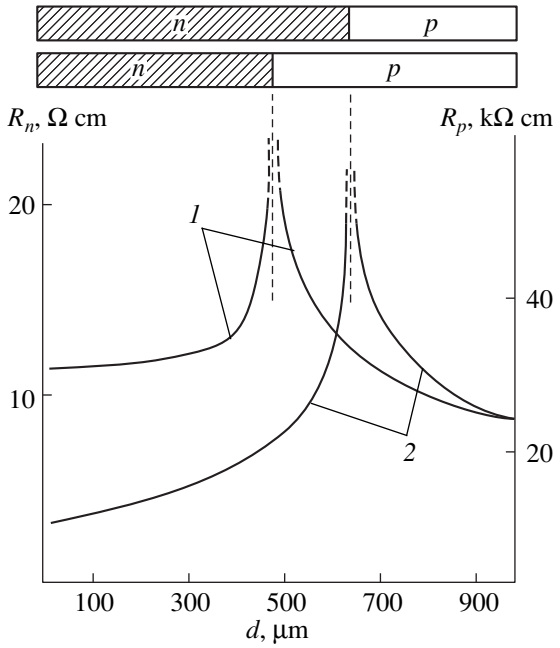

Fig. 1. Resistivity profiles and *n–p* junction depths for (*1*) unirradiated and (*2*) irradiated parts of the wafer.

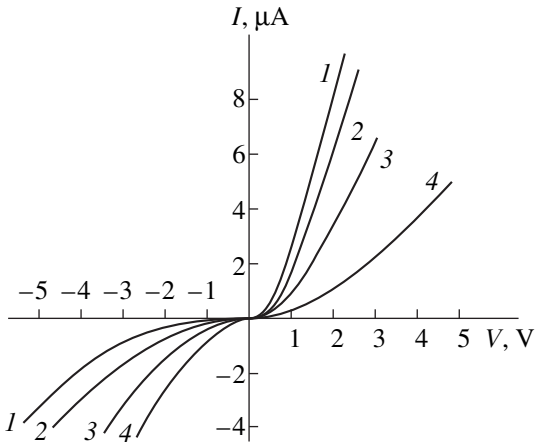

Fig. 2. *I–V* characteristics of ⟨Si + Mg⟩ structures after irradiation with on energy density W = (*1*) 1.8, (*2*) 2.4, (*3*) 2.8, and (*4*) 3.3 J/cm^2.

The *C–V* characteristics for the irradiated samples were taken at a frequency of 1 MHz (Fig. 3). From Figs. 2 and 3, it follows that, at $W = 1.8$ J/cm^2, when the

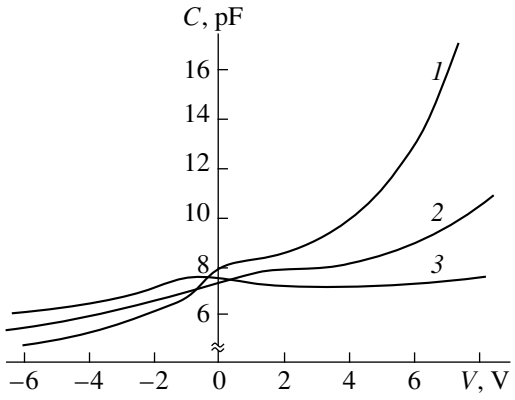

Fig. 3. *C–V* characteristics of ⟨Si + Mg⟩ structures after irradiation with an energy density W = (*1*) 1.8, (*2*) 2.4, and (*3*) 2.8 J/cm^2.

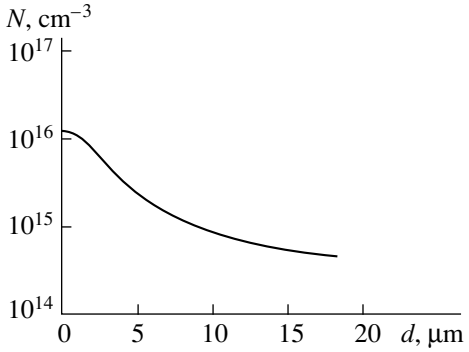

Fig. 4. Concentration profile of magnesium impurity atoms.

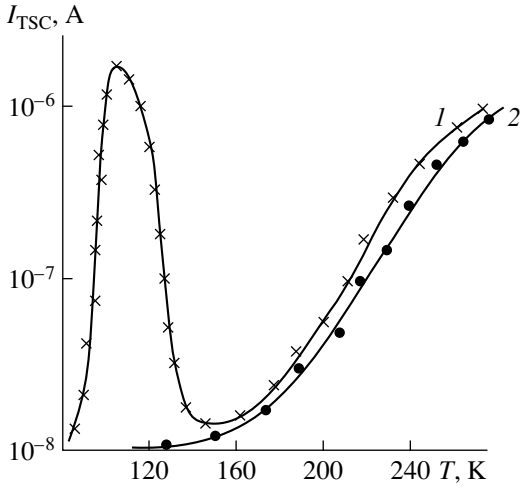

Fig. 5. TSC spectra of irradiated ⟨Si + Mg⟩ structures: (*1*) with and (*2*) without illumination.

rectifying properties are the best, the *C–V* curve (Fig. 3, curve *1*) closely agrees with the theoretical dependence. This curve was used to calculate the concentration profile $N(x)$ of ionized impurity atoms [17]. From the expression

$$\frac{1}{C^2} = \frac{2}{e\varepsilon\varepsilon_0 A^2 N(x)}\left(U - U_0 + \frac{kT}{e}\right),$$

where A is the surface area of a test varactor, U_0 is the barrier height at the metal–semiconductor junction, U is the reverse bias, and $x = (\varepsilon\varepsilon_0 A)/C$, one obtains by differentiation

$$N(x) = -\frac{C^3}{e\varepsilon\varepsilon_0 A^2}\left(\frac{dC}{dU}\right)^{-1}.$$

The calculated dependence $N(x)$ is shown in Fig. 4. It is seen that the impurity atom concentration in the surface layer, $N \sim 10^{16}$ cm^{-3}, exceeds the known value of the ultimate solubility of magnesium in silicon [1].

To record TSC spectra at 80 K and above, the wafers were immersed in a liquid nitrogen cryostat with optical windows. The samples were in contact with a copper unit having a heater to smoothly vary the wafer temperature. The temperature dependence of the current in the range indicated above was measured at a heating rate of ~0.4 K s^{-1}.

The TSC spectra were analyzed according to the following procedure [18]. Placed into a vacuum cryostat, the sample was cooled in liquid nitrogen and then exposed to light from the fundamental band. Then, the sample was uniformly heated in darkness, and the temperature variation of the current was recorded. Comparing the resulting dependences with those obtained without preliminary low-temperature photoexcitation of the sample, one can discover TSC peaks due to trap depletion. These data clarify the energy of the traps in the energy gap. Sometimes, the trap concentrations and capture cross sections can also be determined.

Under certain conditions, the initial stage of TSC growth can be described as

$$I \sim \exp(-E_i/kT), \qquad (1)$$

where E_i is the trap energy reckoned from the edge of the nearest allowed band [18].

TSC spectra of the irradiated samples with the deposited magnesium film (⟨Si + Mg⟩) are presented in Fig. 5. A peak indicating the appearance of a local center is seen at T_{max} = 105 K. Note that laser irradiation of the reference silicon wafers (without the magnesium films) does not give rise to the peak in the TSC spectra, thus pointing to the "magnesium" origin of this peak. The corresponding energy level determined by formula (1) from the initial stage of TSC growth (Fig. 5, curve *1*) was equal to E_c = 0.13 eV, whereas that responsible for the equilibrium conductivity of the irradiated samples (Fig. 5, curve 2) was found to be E_c = 0.28 eV. These values are in good agreement with the literature data for magnesium levels in silicon [6–8].

Thus, we can argue that laser irradiation of ⟨Si + Mg⟩ samples favors magnesium diffusion into silicon. At

particular radiation intensities, the conductivity type changes and rectifying current–voltage characteristics are observed. The magnesium concentration profile constructed from the capacitance–voltage characteristics shows that, in the near-surface region, the magnesium concentration exceeds the concentration of electroactive magnesium atoms reported elsewhere in the literature. As follows from the TSC spectra, laser irradiation ⟨Si + Mg⟩ produces local centers with ionization energies E_c = 0.13 and 0.28 eV.

REFERENCES

1. F. V. Gasparyan, Z. N. Adamyan, and V. M. Arutyunyan, *Silicon Photodetectors* (Yerevansk. Gos. Univ., Yerevan, 1989).
2. V. I. Fistul' and A. M. Pavlov, Fiz. Tekh. Poluprovodn. (Leningrad) **17**, 854 (1983) [Sov. Phys. Semicond. **17**, 535 (1983)].
3. I. B. Khaĭbulin and L. S. Smirnov, Fiz. Tekh. Poluprovodn. (Leningrad) **19**, 569 (1985) [Sov. Phys. Semicond. **19**, 353 (1985)].
4. A. V. Dvurechenskiĭ, G. A. Kachurin, E. V. Nidaev, and L. S. Smirnov, *Pulsed Annealing of Semiconducting Materials* (Moscow, 1982).
5. S. Yu. Karpov and Yu. V. Koval'chuk, Fiz. Tekh. Poluprovodn. (Leningrad) **20**, 1945 (1986) [Sov. Phys. Semicond. **20**, 1221 (1986)].
6. J. W. Chen and A. G. Milnes, Annu. Rev. Mater. Sci. **10**, 157 (1980).
7. E. Ohta and M. Sakata, Solid-State Electron. **22**, 677 (1979).
8. A. L. Lin, J. Appl. Phys. **53**, 6989 (1982).
9. M. Kleverman, K. Bergman, and H. G. Grimmeis, Semicond. Sci. Technol. **1**, 49 (1986).
10. E. Janzen, R. Stedman, G. Grossmann, and H. G. Grimmeiss, Phys. Rev. B **29**, 1907 (1984).
11. J. Wang, R. Wood, and P. Pronko, Appl. Phys. Lett. **33**, 455 (1978).
12. M. Hensen and K. Anderko, *Constitutions of Binary Alloys* (McGraw-Hill, New York, 1958), p. 1488.
13. B. I. Boltaks, *Diffusion in Semiconductors* (Fizmatgiz, Moscow, 1961; Academic, New York, 1963).
14. V. L. Vinetskiĭ and G. E. Chaĭka, Fiz. Tverd. Tela (Leningrad) **24**, 2170 (1982) [Sov. Phys. Solid State **24**, 1236 (1982)].
15. V. N. Strekalov, Fiz. Tekh. Poluprovodn. (Leningrad) **20**, 361 (1986) [Sov. Phys. Semicond. **20**, 225 (1986)].
16. V. S. Vikhnin and M. K. Sheĭnkman, Fiz. Tekh. Poluprovodn. (Leningrad) **19**, 1577 (1985) [Sov. Phys. Semicond. **19**, 970 (1985)].
17. L. V. Pavlov, *Methods for Determination of Parameters of Semiconducting Materials* (Vysshaya Shkola, Moscow, 1987).
18. A. G. Milnes, *Deep Impurities in Semiconductors* (Wiley, New York, 1973; Mir, Moscow, 1977).

Translated by A. Sidorova-Biryukova

Technical Physics, Vol. 46, No. 2, 2001, pp. 202–206. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 71–76.

SOLID-STATE ELECTRONICS

Profiling of Holographic Diffraction Gratings by Using Silver–Arsenic Triselenide Interaction

N. V. Sopinskiĭ, P. F. Romanenko, and I. Z. Indutnyĭ

Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, Kiev, 03028 Ukraine

e-mail: sopinsk@class.semicond.kiev.ua

Received May 10, 2000

Abstract—Rulings of holographic gratings were profiled by interaction of silver with chalcogenide glassy semiconductors. The shape of the rulings was determined with an atomic force microscope. Angular and spectral dependences of the diffraction efficiency of the gratings were found, and a relation between these dependences and the grating surface pattern was analyzed. *© 2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Profiled (blazed) gratings make it possible to concentrate energy in a given spectrum range. Of special interest are profiled holographic gratings (PHGs), since they combine the advantages of ruled gratings with those of conventional (unprofiled) holographic gratings (CHGs), namely, high diffraction efficiency in a given spectrum range and low level of light scattering. One method of fabricating PHGs is making them from CHGs. To do this, an initial symmetric grating is usually etched by an inclined ion beam [1].

The recent trend in CHG technology is the use of chalcogenide glassy semiconductor (CGS) films [2–5]. These films are high-resolution inorganic photoresists. Due to light-induced structure transformations, their solubility, particularly, in organic alkaline solvents, changes. Based on this effect, symmetric CHGs are obtained when an interference pattern is recorded (written) on a CGS film with subsequent selective etching in organic alkaline etchants [2–5].

One of today's problem in this field is the transformation of symmetric CGS-based CHGs into asymmetric ones. Publication [5] reports PHGs prepared by ion etching of symmetric CHGs formed on CGS films. The unique properties of CGS films allow for other, unusual, methods for transforming their surface pattern. In [6, 7], we developed a method of fabricating blazed HGs. In this method, symmetric rulings of an initial grating are made asymmetric by using additional oblique monochromatic or polychromatic irradiation and chemical etching of the irradiated grating.

In this work, we made an attempt to fabricate PHGs through interactions that take place when silver layers are vacuum-deposited onto CGS films. Ag–CGS interaction begins during deposition [8] and continues at a different rate (depending on a specific Ag–CGS system) after the deposition process is terminated [9]. The metal penetrates into the semiconductor to form a metal-enriched (to several tens of atomic percent) phase. This phase differs in properties from both the metal and the semiconductor [10, 11]. Profiling was accomplished by transforming starting symmetric (unprofiled) HGs written on CGS films. Here, we took advantage of the fact that the etching selectivity for Ag-doped CGS films is much higher than the photoinduced selectivity.

In deciding on a CGS material, preference was given to As_2Se_3. First, the use of arsenic triselenide makes feasible writing of a starting symmetric HG by means of a He–Ne laser [12], which was used in this work. Second, the rate of interaction between As_2Se_3 and Ag is one of the highest among CGS–Me systems: intense interaction proceeds both during and after deposition of the metal even at room temperature [9].

PROFILING TECHNIQUE AND STUDY OF SYMMETRIC AND ASYMMETRIC HOLOGRAPHIC GRATINGS

Starting gratings were written on relatively thick (800–1000 nm) As_2Se_3 films deposited onto mirrorlike glass substrates by thermal evaporation of glassy As_2Se_3. The gratings were exposed on a holographic setup based on the wave amplitude division method. The spatial frequency of the gratings was 600 mm^{-1}. Writing was carried out with a He–Ne laser (λ = 632.8 nm). The radiant exposure was ~10^{-1} J/cm^2. After exposure, the samples were etched in an amine-based alkaline solution. During etching, the CGS film was selectively dissolved with the formation of rulings of regular shape.

The next procedure was profiling. A thin (1–10 nm) Ag layer was applied on the gratings at a certain angle φ varying from 10° to 80° with respect to the normal to the grating. The grating was mounted in such a way that the flux of evaporated silver was directed normally to the rulings. Ag penetrates into the As_2Se_3 layer at the instant of deposition and also as a result of subsequent

chemically and thermally stimulated diffusion to form a metal-enriched semiconductor layer (reaction product). The etch rate of Ag-doped CGS films in alkaline etchants is much lower than for undoped ones; therefore, in subsequent etching of the grating, the former served as a protective mask. Unprotected regions of the As_2Se_3 layer were etched off, and the grating (rulings) became asymmetric. In this way, we obtained profiled (blazed) gratings. Profiling was performed in the same amine-based etchant, which was used to pattern the starting gratings.

To estimate the profiling effect, we recorded angular and spectral dependences of the absolute diffraction efficiency η for the starting and transformed gratings in the first diffraction order (η is defined as the ratio of the first-order diffraction intensity to the incident intensity). Prior to optical measurements, the as-prepared and profiled gratings were covered by a 100-nm-thick reflection aluminum film. η vs. β curves (β is the angle of incidence of light) were taken with an LGN 208 A laser (λ = 632.8 nm, β = 0°–80°). Spectral measurements of η were made using Littrow's autocollimation scheme. The angle between the incident and diffracted beams was about 8°; and the spectral range, 400–800 nm. Both the spectral and angular dependences of η were measured for *s*- and *p*-polarized light (*E* is perpendicular and parallel to rulings, respectively), as well as for unpolarized light.

The surface pattern of the gratings was examined with a Dimension 3000 scanning probe microscope (Digital Instruments) in the AFM tapping mode.

RESULTS AND DISCUSSION

The geometric and diffraction properties of the as-prepared and profiled gratings show that the effect of profiling depends, to some extent or another, on many process parameters. Among them are the amount of Ag deposited on the as-prepared grating and the angle of deposition, time of Ag–As_2Se_3 interaction, thickness of the As_2Se_3 film on which the initial symmetric grating is written, exposure and etching conditions for preparation of the starting grating, and etching conditions for profiling. The fact that the desired effect is related to a wide variety of physical and chemical factors implies its significant stability. In addition, this allows us to assume that the properties of PHGs thus obtained can be varied in wide limits. On the other hand, it becomes difficult to estimate the relative contribution of the above factors to the PHG performance. In general, they can be subdivided into three groups: those associated with the CGS film (primarily its thickness and deposition rate), those that govern the fabrication of the symmetric grating, and those closely related to Ag–CGS interaction.

The first two have been much studied. Empiric experimental studies and numerical simulation of HG writing in CGS-based inorganic resists have been performed [2, 3]. These experiments and also the simulation of surface patterning suggest that not only the depth but also the shape of the pattern depends on the initial thickness of the layers, their properties, exposure conditions, etch times, and etchant selectivity. Holographic diffraction gratings with rulings of sinusoidal and cycloidal shapes have been obtained.

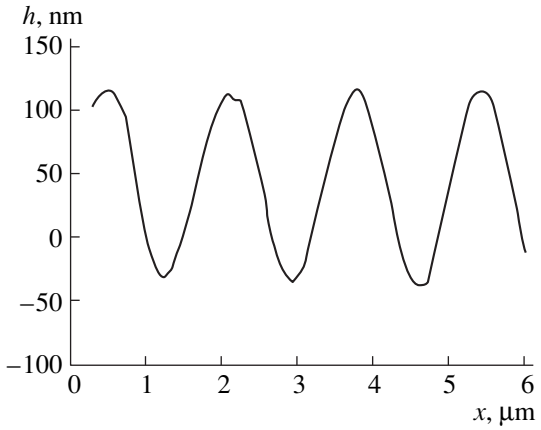

Fig. 1. Ruling profile in the initial symmetric holographic grating.

We performed experiments with gratings that had sinusoidal rulings, which are the most studied. Figure 1 shows a typical AFM image of the starting symmetric grating with a spatial frequency of 600 mm^{-1}. As seen from the image, the shape of the rulings is truly close to sinusoidal. The depth of the grooves is $h_0 \approx 150$ nm; hence, the modulation depth $h_0/d \approx 0.09$ (the spacing of the grating is 1667 nm). The distance between identical points on the tops of the neighboring rulings (surface distance) is 1700 nm, or 2% greater than the horizontal distance between the rulings (which equals the spacing of the grating). The mean angle of inclination of the facets (facet angle) is ~10°, the steepest being 15°. For the wavelength and angle-of-incidence ranges used in the experiments, the diffraction efficiency of such gratings was the same (within the accuracy of measurement: 2% for angular and 5% for spectral measurements) when they were irradiated from opposite directions perpendicular to the rulings.

Figure 2 shows the profile of a typical asymmetric grating made from a symmetric one by additional evaporation of a silver film of thickness $\langle h_{Ag} \rangle$ = 3.6 nm (this value is averaged over the grating area) at 50° to the normal to the substrate surface, and subsequent etching. The ruling depth in the new, asymmetric, grating is more than twice that in the initial one, 355 nm, and the modulation depth is $h_0/d \cong 0.21$. The minimum is 469 nm away from the left peak and 1198 nm from the right one; that is, the projection of the larger facet onto the substrate surface is nearly three-fourths of the grating spacing. The smaller facet has a gradually increasing steepness with an average angle of 37°; the steepest

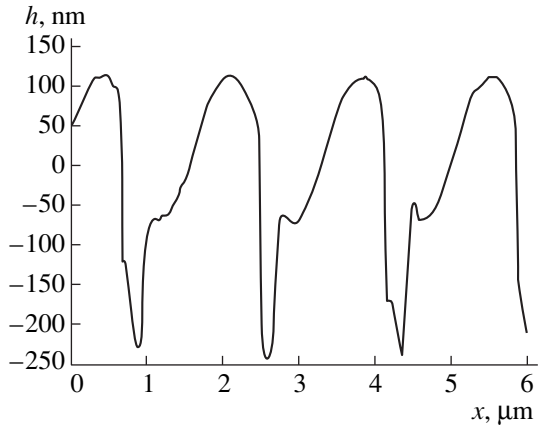

Fig. 2. Ruling profile in the grating transformed from the initial grating shown in Fig. 1.

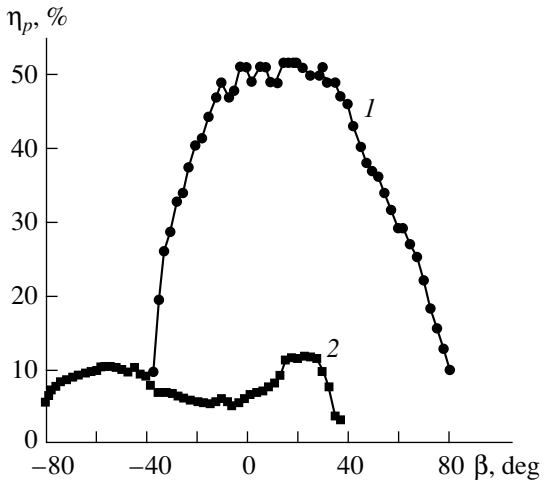

Fig. 3. Angular dependences of the diffraction efficiency of the PHG for light polarized parallel to the ruling direction. λ = 632.8 nm.

portion is inclined 80° with respect to the substrate. With the larger facet, the situation differs. It can be divided into three parts: top (gently sloping), middle (nearly a plateau), and bottom (the steepest portion). The height of the gentle (top) portion somewhat exceeds the ruling half-height; here, the average inclination is about 10°. The inclination of the bottom portion is close to the maximum steepness of the smaller facet. On average, the larger facet angle is about 16°.

From a comparison between the ruling shapes of the initial and transformed gratings, it follows that the front facets of rulings of the initial grating remain practically intact. This means that the reaction product nearly completely protected the CGS layer from dissolution during profiling etching. The first (gentle) portion of the larger slope virtually copies the profile of the initial sinusoidal grating. The back (left-hand in Fig. 2) facets were affected by etching much more noticeably. Here, small plateaus correspond to the partially etched bottom part of the left-hand slope of initial rulings. As we move upward along the left-hand slope, the etch depth of the initial grating increases, the PHG bottom lying almost in the middle of the back slope.

The degree of asymmetry of the profiled grating was studied by taking the angular and spectral dependences of the diffraction efficiency η. For *p* polarization, the angular dependences are plotted in Fig. 3. Here, curves *1* and *2* were obtained when light was incident on the larger and the smaller facet, respectively. For normal incidence (φ = 0, symmetric arrangement of diffraction orders), the diffraction efficiencies measured on the two facets differ by a factor of 8.5. This points to the considerable asymmetry of the ruling shape. For the larger facet, the maximum efficiency was observed when the angle of incidence was close to the mean facet angle. For the smaller one, two peaks appear: the angular position of one of them (–50°) is close to the mean slope of this facet, while that of the other (20°) coincides with the position of the peak from the larger facet and is likely to be associated with rereflections of the light incident on the larger facet. The share of total energy accounted for by conjugate diffraction orders is maximal at 20°; in other words, the maximum amount of light is reflected and rereflected at this angle. Figure 4 shows the angular dependences for *s* polarization. At φ = 0, the diffraction efficiency for the larger and the smaller facet is, respectively, 44 and 8.5%; i.e., the difference is fivefold. Unlike *p* polarization, here distinct anomalies are observed. The anomaly at β = 15° is associated with the disappearance (appearance) of the second order; while that at 38°, with the appearance (disappearance) of the conjugate order. The anomalies are more pronounced for the reflection from the smaller facet possibly because of the greater effect of the larger facet on the smaller than vice versa. Thus, the angular dependences in the case of *p* polarization more adequately depict the ruling shape.

Spectral dependences of the diffraction efficiency for unpolarized light that were taken in the autocollimation regime on the larger facet of the rulings (curve *1*), on their smaller facet (curve *2*), and for the symmetric grating (curve *3*) are shown in Fig. 5. Throughout the spectral range (400–800 nm), the diffraction efficiency for the larger facet exceeds that for the smaller one. For wavelengths of 620, 660, and 700 nm, these values differ by a factor of 5.7, 4.5, and 4.5, respectively. Also, the efficiency for the larger facet exceeds that for the symmetric grating, except for the short-wave part of the interval. This means the profiling has a significant effect between 500 and 800 nm, and at 620 nm, the effect is the greatest. If the wavelength 620 nm is taken as the blaze wavelength, one can determine the blaze angle φ_b (the larger-facet angle). This angle is found from the formula [13]

$$2d\sin\varphi_b = m\lambda_b, \qquad (1)$$

where λ_b is the blaze wavelength, d is the lattice spacing, and m is a diffraction order.

The blaze angle equals 11°. Comparing the value of the blaze angle obtained from spectral measurements with the facet angles, we see that λ_b nearly coincides with the slope of the gentle portion of the larger facet. The peak at the blaze angle is not very pronounced because the three portions on the larger facet greatly differ in slope.

Simple geometric simulation of Ag deposition will allow us to describe Ag–As_2Se_3 interaction in quantitative terms. Let the x axis be directed normally to the rulings and the y axis, normally to the substrate (Fig. 6). The density of Ag deposited on the ruling surface (and, hence, the thickness of the Ag film) depends on the amount of the metal evaporated and the angle between the Ag flux and each specific deposition area on the ruling surface. Then, the thickness of the metal at a point x in the cross section perpendicular to the grating relief is given by

$$h_{Ag}(x) = K\cos\Theta(x), \quad (2)$$

where K is a proportionality coefficient, which depends on the amount of Ag deposited, and $\Theta(x)$ is the angle of incidence of the Ag flux on the ruling surface at the point x.

We also have

$$\Theta(x) = \varphi - \alpha(x), \quad (3)$$

where φ is the angle of incidence of the Ag flux on the substrate that is reckoned from the normal to the substrate and $\alpha(x)$ is the ruling inclination to the flat surface of the substrate at the point x.

Equation (2) is valid if the sticking coefficient of the metal is independent of the angle of incidence on the CGS film surface and on the amount of the metal deposited. The profile of the initial sinusoidal grating is described by the expression

$$h(x) = (h_0/2)(1 + \cos(2\pi x/d)), \quad (4)$$

where d is the grating spacing and h_0 is the height of the profile.

The inclination of this grating to the substrate surface is then expressed as

$$\alpha(x) = \arctan\{[\pi h_0/d][-\sin(2\pi x/d)]\}, \quad (5)$$

where h_0/d is the modulation depth for the initial grating.

Figure 6 illustrates the distribution of the Ag film thickness along the ruling of the initial grating. The thickness was calculated by formulas (2), (3), and (5). Comparing this distribution with the distortions of the initial grating due to the metal deposition and subsequent etching, one can conclude that etching is virtually absent for $h_{Ag}(x) > 3$ nm. When $h_{Ag}(x)$ is less than 3 nm, the etch rate sharply grows. Since the Ag films (of thickness less than 10 nm) deposited on the noninter-

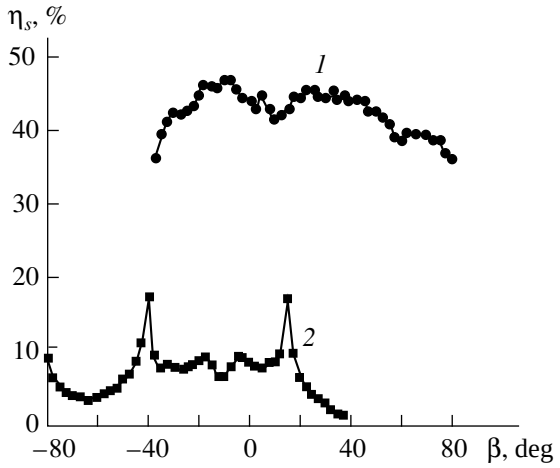

Fig. 4. The same as in Fig. 3 for light polarized perpendicular to the ruling direction.

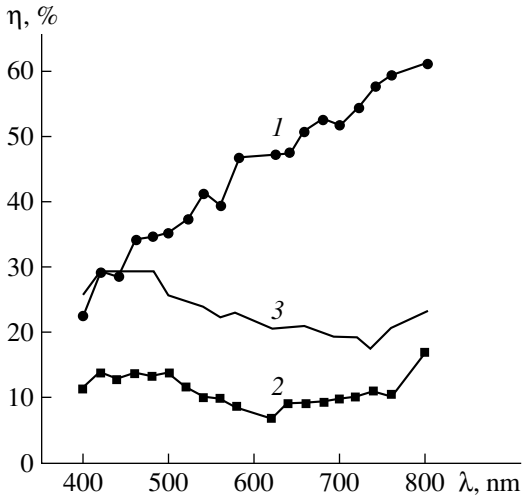

Fig. 5. Spectral dependences of the diffraction efficiency for unpolarized light.

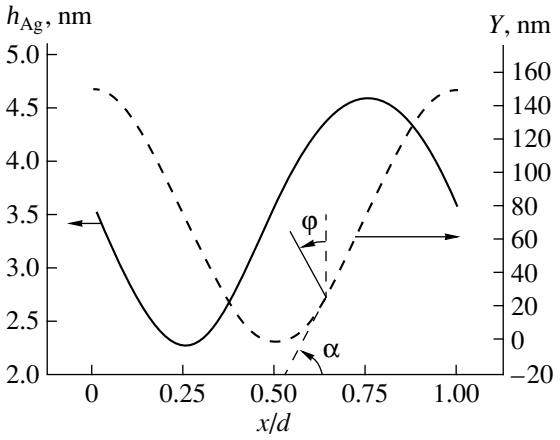

Fig. 6. Profile of a ruling of the sinusoidal grating with h_0 = 150 nm and d = 1667 nm (Y, dashed line) and the distribution of the Ag deposit along this ruling (h_{Ag}, solid line). The mean thickness of the Ag layer is $\langle h_{Ag}\rangle$ = 3.6 nm.

acting substrate are discontinuous (form islands), they could not play the role of a mask during profiling etching. It appears that protection is provided by the top metal-doped CGS layer. The Ag thickness critical for the formation of the mask is seen to be about 3 nm. This value is close to that obtained earlier upon calculating the Ag amount incorporated into the As_2Se_3 film when the metal is thermally evaporated on this film [8]. The etch rate is maximal at the point one-fourth of the spacing away from the top of the ruling back facet, where the Ag thickness estimated is minimal, 2.3 nm.

From the measurements and simulation of the shapes of the initial and profiled gratings, an empiric relationship $\Delta h(h_{Ag})$ can be derived, where Δh is the thickness of that part of the As_2Se_3 film dissolved during etching and h_{Ag} is the amount of the Ag deposit. As follows from correlation analysis data, this empiric relationship is fairly accurately approximated by the sum of two exponentials:

$$\Delta h(h_{Ag}) = 204.93\exp\{-(h_{Ag} - 2.412)/0.125\} + 61.51\exp\{-(h_{Ag} - 2.412)/1.553\}. \quad (6)$$

With (6), we simulated the transformation of the initial symmetric gratings into asymmetric ones. The initial grating was assumed to be sinusoidal with a modulation depth of $h_0/d = 0.09$. Variable parameters were angle of deposition of Ag particles and the thickness of the metal deposited.

The simulation confirms experimental data suggesting that the profiling effect takes place in a wide range of these parameters. For example, it was shown that, as the Ag film gets thinner (at a given angle of Ag deposition) and etching of the Ag-doped grating becomes selective, the initially symmetric profile transforms into an asymmetric one because of a slight (~0.01*d*) shift of its maxima and a more considerable (~0.05*d*) shift of its minima to the left of the initial grating rulings. The modulation depth remains practically unchanged in this case, since the minima and maxima lower roughly equally. With further decreasing $\langle h_{Ag}\rangle$, the ruling shape can be approximated by an asymmetric trapezoid. The modulation depth in such a grating is somewhat larger than in the initial symmetric one. As $\langle h_{Ag}\rangle$ continues to decrease, the shape becomes nearly triangular, the projections of the ruling facets on the *x* axis being about 0.7*d* for the larger and 0.3*d* for the smaller. For an extremely thin Ag film, the larger facet has two drastically differing slopes because of a sharp increase in the etch rate. The maximum ratio of the facet projections is ~0.72*d*/0.28*d* in this case.

CONCLUSION

Thus, our results indicate that the profile of a holographic grating can be changed by using Ag–As_2Se_3 interaction. The blaze angle in the PHGs depends on the modulation depth of the initial sinusoidal gratings and also on the angle of deposition and the Ag layer thickness. By controlling the Ag distribution along the ruling cross section through a change in the deposition angle and the Ag layer thickness, one can substantially change the grating profile and, hence, the blaze angle, which specifies the reflection properties of the grating.

REFERENCES

1. J. Flammand, F. Bonnemason, A. Thevenon, and J. X. Lerner, Proc. SPIE **1055**, 288 (1989).
2. I. Z. Indutnyi, I. I. Robur, P. F. Romanenko, and A. V. Stronski, Proc. SPIE **1555**, 248 (1991).
3. I. Z. Indutnyi, A. V. Stronski, S. A. Kostioukevich, *et al.*, Opt. Eng. **34**, 1030 (1995).
4. R. R. Gerke, T. G. Dubrovina, P. A. Dmitrikov, and M. D. Mikhaĭlov, Opt. Zh. **64** (11), 26 (1997) [J. Opt. Technol. **64**, 1008 (1997)].
5. A. V. Lukin, A. S. Makarov, F. A. Sattarov, *et al.*, Opt. Zh. **66** (12), 73 (1999) [J. Opt. Technol. **66**, 1071 (1999)].
6. P. F. Romanenko, M. V. Sopinski, and I. Z. Indytnyi, Proc. SPIE **3573**, 457 (1998).
7. P. F. Romanenko, N. V. Sopinskiĭ, I. Z. Indutnyĭ, *et al.*, Zh. Prikl. Spektrosk. **66** (4), 587 (1999).
8. M. T. Kostyshin, V. L. Gromashevskiĭ, N. V. Sopinskiĭ, *et al.*, Zh. Tekh. Fiz. **54** (6), 1231 (1984) [Sov. Phys. Tech. Phys. **29**, 709 (1984)].
9. I. Z. Indutnyĭ, M. T. Kostyshin, O. P. Kasyarum, *et al.*, *Photostimulated Interactions in Metal-Semiconductor Structures* (Naukova Dumka, Kiev, 1992).
10. A. G. Fitzgerald and C. P. McHardy, Surf. Sci. **152/153**, 1255 (1985).
11. V. Honig, V. Fedorov, G. Liebmann, and P. Suptitz, Phys. Status Solidi A **96** (2), 611 (1986).
12. P. F. Romanenko, I. I. Robur, and A. V. Stronskiĭ, Optoélektron. Poluprovodn. Tekh. **27**, 47 (1994).
13. M. C. Hutley, *Diffraction Gratings* (Academic, London, 1982).

Translated by V. Isaakyan

Technical Physics, Vol. 46, No. 2, 2001, pp. 207–211. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 77–81.

An Electric Discharge Emitter Operating Simultaneously in the 308 [XeCl(*B*–*X*)], 258 [Cl_2(*D*'–*A*')], 236 [XeCl(*D*–*X*)], 222 [KrCl(*B*–*X*)], 175 [ArCl(*B*–*X*)], and 160 [H_2(*B*–*X*)] nm Bands

A. K. Shuaibov, L. L. Shimon, A. I. Dashchenko, and I. V. Shevera

Uzhgorod State University, Uzhgorod, 88000 Ukraine

e-mail: ishev@univ.uzhgorod.ua

Received March 31, 2000

Abstract—An experiment to optimize an ultraviolet (UV)–vacuum ultraviolet (VUV) multiwave emitter using chlorine molecules and chlorides of heavy inert gases is reported. The emitting medium was an Ar–Kr–Cl_2 or an Ar–Kr–Xe–Cl_2 (HCl) mixture kept at a pressure ranging from 1 to 30 kPa. Excitation was effected by means of a transverse volume discharge with spark preionization. Emission spectra were examined. The dependences of the emission intensity on the total pressure of the medium, partial pressures of its components, charging voltage, and number of discharge pulses were studied. It is demonstrated that such a discharge emits simultaneously in the 308, 258, 236, 222, 175, and 160 nm bands due to the transitions XeCl(*B*–*X*), Cl_2(*D*'–*A*'), XeCl(*D*–*X*), KrCl(*B*–*X*), ArCl(*B*–*X*), and H_2(*B*–*X*), respectively. It was established that the respective intensities are close to each other if the partial pressures are as follows: P_{Ar} = 10–20 kPa; $P_{Kr,\ Xe}$ = 0.4–0.6 kPa; P_{Cl_2} = 0.2–0.4 kPa, P_{HCl} = 0.08 kPa, and P_{H_2} = 0.5–1.0 kPa. It was found that the addition of H_2 to the medium decreases the intensities of the excimer bands, increases the emission resource (to 10^4 pulses or higher), and expands the operating wavelength range. The last-named effect is due to Lyman H_2 bands (at 158–161 nm). © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Today, electric discharge excimer lamps offer the highest power and the best selectivity among sources of spontaneous emission in the wavelength range 200–350 nm [1–3]. For this reason, they are widely employed in microelectronics, high-energy chemistry, biology, and quantum electronics [4–6]. On the other hand, the majority of them operate at a single wavelength determined by the emitting medium. It seems worthwhile to explore the feasibility of multiwave lamps, namely, those using particular *B*–*X* transitions of *RX** molecules, where *R* is Ar, Kr, or Xe and *X* is F or Cl. These sources could find application in pulse photometry, biochemistry, biophysics, and medicine. For example, they could serve to calibrate detectors of short-wave light pulses or to simultaneously treat selected molecular bonds in compounds having high chemical or biological activity. We designed such emitters for the range 353–222 nm. Some of them operate at 222 (KrCl), 249 (KrF), 308 (XeCl), and 353 (XeF) nm, with F and Cl being provided by CF_2Cl_2 molecules [7, 8]. Other emitters use multicomponent media (consisting of He, Kr, Xe, HCl, and SF_6), which include two heavy inert gases and differing halogen-containing molecules [9, 10]. To generate photons with higher energies, we started looking for ways to expand the operating range into the VUV region (≤190 nm). Early results were obtained with a source emitting at 258 (Cl_2^*) and 175 (ArCl*) nm, the medium consisting of He (Ne), Ar, and Cl_2 (HCl) [11].

This paper reports on an experiment to optimize a UV–VUV emitter operating simultaneously at 308, 258, 236, 222, 175, and 160 nm. Operating mixtures were composed of Ar, Kr, Xe, Cl_2 (HCl), and H_2.

EXPERIMENTAL SETUP

The medium was excited by transverse volume discharge (TVD). Spark preionization was effected automatically in the interelectrode space. A sketch of the emitter is shown in [8]. A TVD plasma occupied a volume of 18 × 2.2 × 1.0 cm^3, and the electrodes were spaced 2.2 cm apart. TVD was initiated by a two-loop *LC* circuit including a 30-nF storage ceramic capacitor and 20 pulse-sharpening ceramic capacitors (KVI-3, 470 pF, 20 kV), the latter providing a total capacitance of 9.4 nF. The capacitors were combined into two banks sealed with an insulating compound. The banks were placed near the TVD electrodes and preionization spark

gaps inside a discharge chamber. This excitation circuit closely resembles the system used for generating fast ionization waves in pulsed longitudinal discharge [12–14], since the main part of the circuit has a very small inductance (less than 10 nH), the discharge plasma is separated from the grounded shields by a high-permittivity insulator (the compound), and the pulses of TVD current are narrower than 30 ns.

Spectra were examined with a 0.5-m vacuum monochromator including a diffraction grating with 1200 lines per millimeter. The reciprocal linear dispersion of the spectrophotometer was 1.4 nm/mm. The emission emerged from the hermetically sealed discharge chamber via a CaF_2 window. The photodetector was built around an FÉU-142 photomultiplier with a LiF window. The chambers containing the photomultiplier and the grating, respectively, were evacuated to a residual pressure $P \le 10^{-3}$ Pa. The monochromator–photomultiplier system operates in the spectral band 130–350 nm. The spectrophotometer was calibrated against the continuous emission of H_2 in the region 165–350 nm.

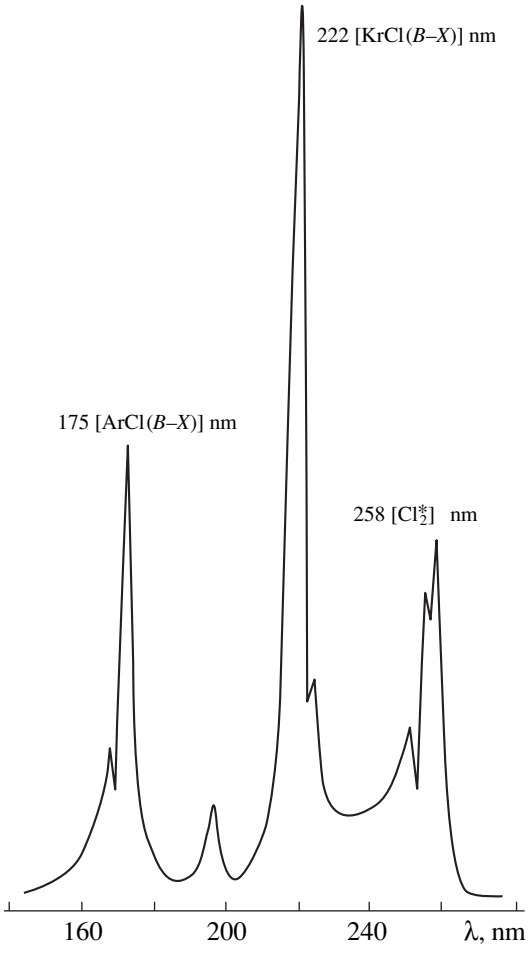

Fig. 1. TVD plasma emission spectrum of the Ar–Kr–Cl_2 mixture with P_{Ar} = 13.3, P_{Kr} = 0.6, and P_{Cl_2} = 0.24 kPa.

EMISSION AND RESOURCE CHARACTERISTICS

Figure 1 shows an emission spectrum of the TVD plasma in an Ar–Kr–Cl_2 mixture. Note that the UV–VUV power is emitted largely at wavelengths of 175, 222, and 258 nm, which, respectively, correspond to the transitions ArCl(*B*–*X*), KrCl(*B*–*X*), and Cl_2(*D*'–*A*'). If small amounts of Xe and H_2 are added to the medium, the spectrum also exhibits bands at 308, 236, and 158–161 nm, representing the transitions XeCl(*B*–*X*), XeCl(*D*–*X*), and H_2(*B*–*X*), respectively. (The third transition produces Lyman H_2 bands.) With the Ar–Cl_2 (HCl) mixture, only two bands are pronounced, at 258 and 175 nm. If Cl_2 is replaced with HCl, the emission due to Cl_2(*D*'–*A*') and ArCl(*B*–*X*) is less intense for all of the mixtures. It was found that the intensity at 175 nm is maximal if the partial pressure of Cl_2 is 0.2–0.4 kPa and that of HCl, 0.08 kPa. This is because HCl and Cl_2 have strong and zero VUV absorption, respectively. With the Ar–Cl_2 mixture, the optimal value of *P*(Ar) is within 8–15 kPa at a reasonable level of the charging voltage *U* across the storage capacitor (*U* = 4–15 kV). If the total pressure *P* in the mixture is raised above 30 kPa, the TVD becomes inhomogeneous. As *U* is increased from 5 to 15 kV, the intensity linearly rises by a factor of 3–5 for all of the bands observed.

Figures 2 and 3 show the intensities emitted in the excimer bands and the Cl_2(*D*'– *A*') band against P_{Kr} for the Ar–Kr–Cl_2 mixture and the Ar–Kr–Xe–Cl_2 mixture, respectively. In the former case, a rise in P_{Kr} to ≥0.05 kPa leads to an increase in the intensity for KrCl(*B*–*X*) and a decrease in that for ArCl(*B*–*X*). For Cl_2(*D*'–*A*'), the radiation yield is maximal at P_{Kr} = 0.3 kPa. If P_{Kr} = 0.6 kPa, the intensities of the three bands are close to each other. At P_{Kr} < 0.05 kPa, a rise in the Kr concentration leads to an increase in the intensity for ArCl(*B*–*X*) (probably, due to changes in disharge parameters such as n_e and T_e). At $P_{Kr} \ge 0.05$ kPa, the decrease for ArCl(*B*–*X*) and the attending increase for KrCl(*B*–*X*), which proceed at equal rates, stem from the fact that Ar atoms are replaced by Kr atoms in the formation of excimer molecules. It has been demonstrated [15] that the largest rate constant *k* in this process is that of energy transfer from Ar($^3P_2$) to Kr with the formation of Kr[5*p*(3/2)$_2$]: $k = 5.6 \times 10^{-12}$ cm^3/s. The process runs vigorously in an Ar–Kr plasma at P_{Kr} = 0.1–1.0 kPa, which agrees with Fig. 2.

With the Ar–Kr–Xe–Cl_2 mixture (Fig. 3), the intensity curves run in a more complicated fashion. As P_{Kr} is increased, the intensity for KrCl(*B*–*X*) rises again, whereas the intensities for XeCl(*B*–*X*) and ArCl(*B*–*X*) exhibit a peak if *P*(Xe) = 0.4 kPa or a decrease if P_{Xe} is larger. For Cl_2(*D*'–*A*'), the curve has a valley at P_{Kr} = 0.4 kPa. Such behavior is mainly due to energy transfer from Ar($^3P_2$) to Kr and Xe atoms [15–17] and to specific features of interaction between excited Kr and Xe

atoms in a Kr–Xe plasma [18, 19]. Compared with the emitting media of electric discharge excimer lasers and high-pressure ($P \geq 100$ kPa) excimer lamps, the recombination reaction $Ar^+ + Cl^- + (Ar) \longrightarrow ArCl^* + (Ar)$ plays a less important part in TVD at P = 1–30 kPa, since its rate constant may be one order of magnitude smaller at such low buffer gas pressures (e.g., 10 kPa rather than 100 kPa) [20]. Instead, the initial TVD stage is governed by the "harpoon" reaction $Ar(m) + Cl_2 \longrightarrow ArCl^* + Cl$ [21, 22], which requires that the plasma contain excited (metastable) atoms of heavy inert gases. This reaction to produce ArCl(B) may proceed at a rate as high as 7×10^{-10} cm^3/s [23]. For P_{Ar} = 13.3, P_{Xe} = 0.4, and P_{Cl_2} = 0.24 kPa, the respective intensities emitted in the excimer bands and the Cl_2(D'–A') band are comparable to each other if P_{Kr} = 0.4–1.2 kPa. This result allows using such an emitter in UV–VUV pulse photometry.

From Fig. 4, where the emission intensity is plotted against the number of discharge pulses N for the KrCl(B–X) band of the Ar–Kr–Cl_2 mixture, one can evaluate the emission resource (the number of pulses after which the intensity is halved). In the experiment, the emission resource was found to be as high as $(1–2) \times 10^4$ pulses. The resource characteristic was obtained under stationary conditions (in the absence of gas flow), the discharge chamber having a passive volume of 10 l. The emission resource is determined mainly by the purity of the buffer gas (we used commercial-grade Ar) and the materials of the electrodes and the discharge chamber (stainless steel, aluminum, acrylic plastic, fluoroplastic, and quartz). Intensity fluctuations could be suppressed by pumping fresh portions of the gas through the emitter at a small rate (≤0.01 m/s). Alternatively, one could employ a solid-state source of Cl_2 and a regenerator of the waste emitting mixture; the former would be switched on if the intensity falls below a given level. Also, note that the resource of an electric discharge laser operating at 308 (XeCl) and 222 (KrCl) nm and using a He–R–HCl mixture at $P \geq 100$ kPa can be increased by two orders of magnitude by introducing a small amount of H_2 ($P_{H_2} \leq 130$ Pa) into the lasing medium [24].

To expand the operating band of the emitter by means of Lyman H_2 bands, we investigated the effect of H_2 on TVD emission. Figure 5 shows the respective intensities in the ArCl, H_2^*, and Cl_2^* bands as functions of P_{H_2} for the Ar–HCl–H_2 mixture. It is seen that the ArCl* and H_2 intensities are reduced considerably if P_{H_2} is raised to a sufficiently high level (≥0.1 kPa). The spectrum exhibits Lyman bands, at 158–161 nm, if $P_{H_2} \geq 0.3$ kPa. At P_{H_2} = 0.57 kPa, the ArCl* and H_2 intensities are the same. As P_{H_2} is increased further, the Lyman bands become dominant. If a small amount of H_2 is added to the Ar–Kr–HCl mixture, the intensity in the ArCl(B–X) band slightly decreases and the resource

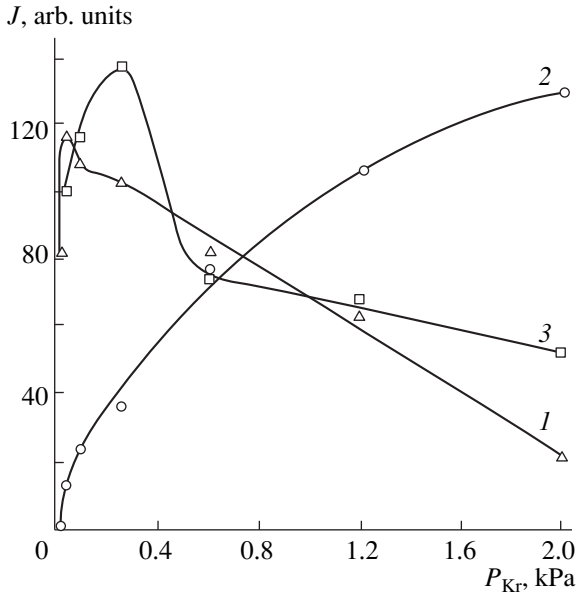

Fig. 2. TVD emission intensity vs. Kr partial pressure at a wavelength of (*1*) 175 (ArCl), (*2*) 222 (KrCl), and (*3*) 258 (Cl_2^*) nm for the Ar–Kr–Cl_2 mixture with P_{Ar} = 13.3 kPa and P_{Cl_2} = 0.24 kPa (U = 12.5 kV).

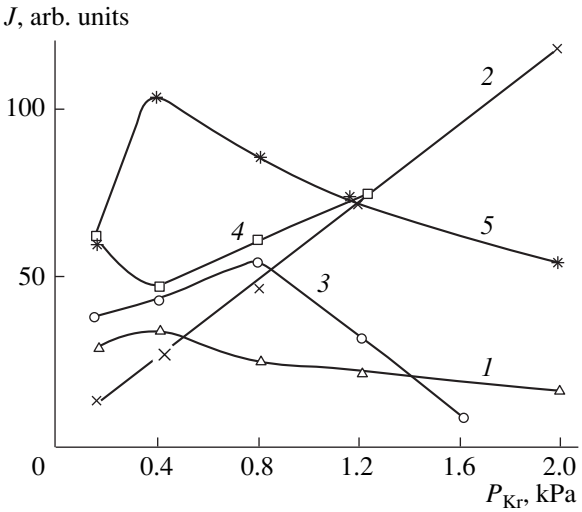

Fig. 3. TVD emission intensity vs. Kr partial pressure at a wavelength of (*1*) 175, (*2*) 222, (*3*) 236, (*4*) 258, and (*5*) 308 nm for the Ar–Kr–Xe–Cl_2 mixture with P_{Ar} = 13.3, P_{Xe} = 0.4, and P_{Cl_2} = 0.24 kPa (U = 12.5 kV).

characteristic develops a plateau (curve *2* in Fig. 4). Consequently, the emission resource increases considerably ($N > 10^4$ pulses). The intensities of the RX^* bands fall mainly because of intense energy transfer from Ar(m) to H_2 [25] and the reduction of HCl when the decay products of the former process interact with H_2. With the mixtures containing H_2, strong continuous emission was also observed in the region 200–400 nm

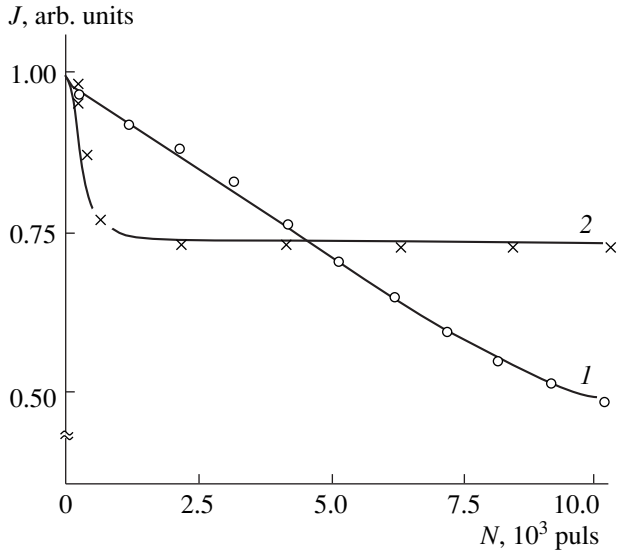

Fig. 4. TVD emission intensity vs. the number of discharge pulses at a wavelength of (*1*) 222 (KrCl) and (*2*) 175 (ArCl) nm for the (*1*) Ar–Kr–Cl_2 and (*2*) Ar–Kr–HCl–H_2 mixtures with P_{Ar} = 13.3, P_{Kr} = 0.4, P_{Cl_2} = 0.24, P_{HCl} = 0.12, and P_{H_2} = 0.13 kPa.

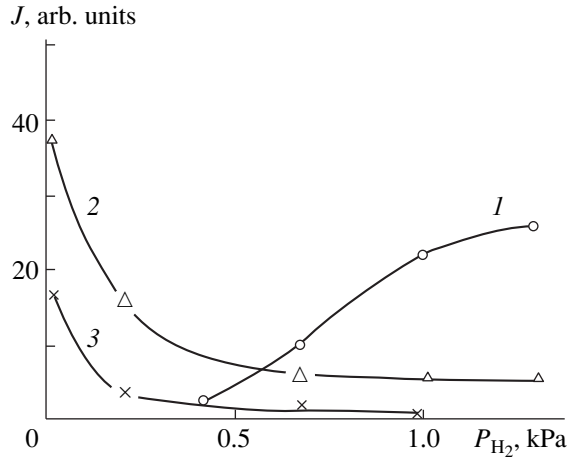

Fig. 5. TVD emission intensity vs. H_2 partial pressure at a wavelength of (*1*) 160 [$H_2(B–X)$], (*2*) 175 (ArCl), and (*3*) 258 (Cl_2^*) nm for the Ar–HCl–H_2 mixture with P_{Ar} = 13.3 and P_{HCl} = 0.12 kPa (U = 12 kV).

at $P_{H_2} \geq 0.5$ kPa. This may be attributed to the formation of ArH* molecules [26] or to continuous emission from H_2 ($a^3\Sigma_g^+ - b^3\Sigma_u^+$). In this experiment, we were unable to distinguish between these types of continuous spectrum. For longitudinal pulse discharge, which develops in the form of a fast ionization wave, it has been demonstrated that this system offers a strong continuous hydrogen spectrum in the UV range if $P_{H_2} \geq 1.33$ kPa and $E/N \geq 100$ Td, the excitation rate of the $a^3\Sigma_g^+$ state of H_2 being 1.3×10^{20} cm^3/s [27]. However, in [27], no data concerning the efficiency with which VUV emission is excited are available. With Ar–Kr–HCl–H_2 mixtures, these processes lead to a decrease in the intensities of the excimer bands when the H_2 concentration is raised.

CONCLUSION

We have established that a TVD plasma in Ar–Kr–Cl_2 and Ar–Kr–Xe–Cl_2 mixtures is an efficient multiwave source of pulsed emission at 308, 258, 236, 222, and 175 nm due to the transitions XeCl($B–X$), $Cl_2(D'–A')$, XeCl($D–X$), KrCl($B–X$), and ArCl($B–X$), respectively. For a reasonable charging voltage (4–15 kV), the respective intensities are close to each other if P_{Ar} = 10–20 kPa, $P_{Kr,\,Xe}$ = 0.4–0.6 kPa, P_{Cl_2} = 0.2–0.4 kPa, and P_{HCl} = 0.08 kPa. With a small amount of H_2 added to the mixture, $P_{H_2} \leq 0.1$–0.2 kPa, the resource characteristic acquires a plateau and the emission resource increases. If P_{H_2} is raised above 0.5–1.0 kPa, the intensities of the bands corresponding to the excimer molecules and Cl_2^* fall by one order of magnitude and Lyman H_2 bands appear at wavelengths of 158–161 nm. The optimal composition of the emitting medium is determined by the energy transfer processes Ar(m)–Kr, Xe, H_2, and Kr(m)–Xe and by the formation of the ArH* and $H_2(a^3\Sigma_g^+)$ molecules.

REFERENCES

1. T. Gerber, W. Luthy, and P. Burkhard, Opt. Commun. **35**, 242 (1980).
2. B. A. Koval', V. S. Skakun, V. F. Tarasenko, and E. A. Fomin, Prib. Tekh. Éksp., No. 4, 244 (1992).
3. V. M. Borisov, V. A. Vodchits, *et al.*, Kvantovaya Élektron. (Moscow) **25**, 308 (1998).
4. V. Yu. Baranov, V. M. Borisov, and Yu. Yu. Stepanov, *Electric Discharge Excimer Lasers on Inert Gas Halogenides* (Énergoatomizdat, Moscow, 1988).
5. U. Kogelshanz and H. Esrom, Laser Optoelectron. **22**, 55 (1990).
6. Yu. G. Basov, *Pumping Sources for Microsecond Lasers* (Énergoatomizdat, Moscow, 1990).
7. A. K. Shuaibov, L. L. Shimon, and I. V. Shevera, Prib. Tekh. Éksp., No. 3, 142 (1998).
8. A. K. Shuaibov, L. L. Shimon, I. V. Shevera, and O. J. Minya, J. Phys. Stud. **3** (2), 157 (1999).
9. A. K. Shuaibov, Teplofiz. Vys. Temp. **36**, 508 (1998).
10. A. K. Shuaibov, Zh. Tekh. Fiz. **68** (12), 64 (1998) [Tech. Phys. **43**, 1459 (1998)].
11. A. K. Shuaibov, L. L. Shimon, A. I. Dashchenko, *et al.*, Pis'ma Zh. Tekh. Fiz. **25** (11), 29 (1999) [Tech. Phys. Lett. **25**, 433 (1999)].
12. É. I. Asinovskiĭ, L. M. Vasilyak, and V. V. Markovets, Teplofiz. Vys. Temp. **21**, 371 (1983).

13. A. G. Abramov, É. I. Asinovskiĭ, and L. M. Vasilyak, Fiz. Plazmy **14**, 979 (1988) [Sov. J. Plasma Phys. **14**, 575 (1988)].
14. B. B. Slavin and P. I. Sopin, Teplofiz. Vys. Temp. **30**, 1 (1992).
15. L. G. Piper and D. W. Setser, J. Chem. Phys. **63**, 5018 (1975).
16. C. H. Chen, J. P. Judisch, and M. G. Payne, J. Phys. B **11**, 2189 (1978).
17. J. Galy, K. Aoume, A. Birot, *et al.*, J. Phys. B **26**, 477 (1993).
18. J. D. Cook and P. K. Leichner, Phys. Rev. A **31**, 90 (1985).
19. J. D. Cook and P. K. Leichner, Phys. Rev. A **43**, 1614 (1991).
20. M. R. Flannery, in *Applied Atomic Collision Physics*, Ed. by E. W. McDaniel and W. Nigan (Academic, New York, 1982; Mir, Moscow, 1986), Vol. 3.
21. E. B. Gordon, V. G. Egorov, V. T. Mikhkel'lson, *et al.*, Kvantovaya Élektron. (Moscow) **15**, 285 (1988).
22. V. É. Peét, E. V. Slivinskiĭ, and A. B. Treshchalov, Kvantovaya Élektron. (Moscow) **17**, 438 (1990).
23. É. M. Vrublevskiĭ, A. V. Gusev, A. G. Zhidkov, *et al.*, Khim. Vys. Énerg. **24**, 356 (1990).
24. T. J. McKee, D. J. James, W. S. Nip, *et al.*, Appl. Phys. Lett. **36**, 943 (1980).
25. B. P. Lavrov and A. S. Mel'nikov, Opt. Spektrosk. **85**, 729 (1998) [Opt. Spectrosc. **85**, 666 (1998)].
26. F. V. Bunkin, V. I. Derzhiev, V. A. Yurovskiĭ, and S. I. Yakovlenko, Kvantovaya Élektron. (Moscow) **13**, 1828 (1986).
27. S. V. Pancheshnyĭ, S. M. Starikovskaya, and A. Yu. Starikovskiĭ, Fiz. Plazmy **25**, 435 (1999) [Plasma Phys. Rep. **25**, 393 (1999)].

Translated by A. Sharshakov

Technical Physics, Vol. 46, No. 2, 2001, pp. 212–218. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 82–88.

RADIOPHYSICS

Synthesis of Current on a Strip from a Given Realizable Radiation Pattern

S. I. Éminov

Mudryĭ State University, Sankt-Peterburgskaya ul. 41, Novgorod, 173003 Russia

Received March 14, 2000

Abstract—A theory of synthesis of current on a strip from a given realizable radiation pattern is developed. The theory chooses the space of currents from the condition that the near-zone field is limited. The space of patterns is defined as the image of the space of currents due to current-to-pattern mapping. For these spaces, the Hilbert structure is introduced and the basis is constructed. As a result, the problem of synthesizing current from a given pattern is reduced to expansion over the basis. A numerical example is considered. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION AND STATEMENT OF THE PROBLEM

A relationship between the density of surface currents induced on both sides of a strip of electrical width $2a$, $j(t)$, and a radiation pattern $F(\chi)$ is given by the integral equation [1–3]

$$F(\chi) = K_a j = a\int_{-1}^{1} \exp(ia\chi t) j(t) dt. \qquad (1)$$

Real angles correspond to $|\chi| \le 1$. However, it is common practice to consider the function $F(\chi)$ at any real χ.

Equation (1) has been studied in many investigations. Analytical methods of synthesizing currents from a given realizable radiation pattern have been developed in [2, 3]. Variational methods that calculate currents producing a radiation pattern close to a given one (not necessarily realizable) are addressed in [4]. In these publications, it is assumed that the current $j(\tau)$ belongs to the space L_2; i.e.,

$$\int_{-1}^{1} |j(t)|^2 dt < +\infty \qquad (2)$$

or, by virtue of Parseval's identity,

$$\int_{-\infty}^{\infty} |F(\chi)|^2 d\chi < +\infty. \qquad (3)$$

However, the selection of the space L_2 has not been substantiated in the publications cited above.

Synthesis of current on an unclosed surface has also been considered in [5], where the current is found with regard for the Meixner conditions imposed on its edge behavior. However, to date, methods developed in [5] have been implemented for closed surfaces, but none for unclosed surfaces.

Whereas the problem of synthesizing current on a closed surface from a given radiation pattern has been covered adequately [6], many problems concerning unclosed surfaces, in particular, the problem of choosing the space of currents, remain unsolved.

The purpose of this paper is to develop a criterion for choosing the space of currents, to construct the space of currents for a strip and the space of radiation patterns, and to study radiation described by formula (1) in appropriate spaces.

ENERGY INTEGRAL: *H*-POLARIZATION PROBLEM

In this section, we consider the *H*-polarization problem: the strip is located in the plane $y = 0$, the generatrix of the strip is parallel to the z axis, and the currents $j_x(\tau)$ pass parallel to the x axis (perpendicularly to the strip edge) and vanish at the edge. In our opinion, the space of currents should be chosen from physical considerations. Namely, we must assume that not only the far-field power but also the near-field power is limited. It is the condition that ensures the uniqueness of solutions to the Maxwell equations [7].

Let us find a condition that provides the finiteness of the field power in the near zone. To this end, we integrate the Poynting vector along a closed line l encircling the cross section of the strip and contract this line into a segment:

$$P = \frac{1}{2}\int_l [\mathbf{E}, \mathbf{H}^*]\mathbf{n} dl = -\frac{\tilde{a}}{2}\int_{-1}^{1} E_x(\tau) j_x^*(\tau) d\tau, \qquad (4)$$

where $\mathbf{E}$ and $\mathbf{H}$ are the fields produced by the currents $j_x(\tau)$, $2\tilde{a}$ is the width of the strip, the asterisk means

complex conjugation, and the normal **n** is parallel to the *y* axis.

Next, we express the electric field E_x in terms of the current density $j_x(\tau)$ [1] as

$$E_x(\tau) = \sqrt{\frac{\mu}{\varepsilon}}\frac{1}{4a}\frac{\partial}{\partial\tau}\int_{-1}^{1} j_x(t)\frac{\partial}{\partial t}H_0^{(2)}(a|\tau-t|)dt - \frac{a}{4}\sqrt{\frac{\mu}{\varepsilon}}\int_{-1}^{1} j_x(t)H_0^{(2)}(a|\tau-t|)dt \quad (5)$$

and represent the Hankel function in formula (5) as the Fourier integral

$$H_0^{(2)}(a|\tau-t|) = \frac{i}{\pi}\int_{-\infty}^{+\infty}\frac{\exp(-ia\chi(\tau-t))}{\sqrt{\chi^2-1}}d\chi, \quad \sqrt{\chi^2-1} = i\sqrt{1-\chi^2}, \quad |\chi|\le 1. \quad (6)$$

Substituting (5) into (4) in view of (1) and (6), we obtain

$$P = -i\sqrt{\frac{\mu}{\varepsilon}}\frac{\tilde{a}\pi}{8a}\int_{-\infty}^{+\infty}\sqrt{\chi^2-1}|F(\chi)|^2 d\chi. \quad (7)$$

Energy integral (7) converges if

$$\int_{-\infty}^{+\infty}\left|\sqrt{\chi^2-1}\right||F(\chi)|^2 d\chi < +\infty. \quad (8)$$

As follows from (8) and (3), the fact that the currents belong to the space $L_2[-1, 1]$ does not suffice for the field power to be finite. To meet this condition, it is necessary to narrow the space $L_2[-1, 1]$.

ENERGY INTEGRAL: *E*-POLARIZATION PROBLEM

Consider the *E*-polarization problem: the currents $j_z(\tau)$ pass parallel to the *z* axis (parallel to the edge of the strip) and tend to infinity at the edges. In this case, integration of the Poynting vector yields

$$P = \frac{1}{2}\int_t [\mathbf{E}, \mathbf{H}^*]\mathbf{n}dl = -\frac{\tilde{a}}{2}\int_{-1}^{1} E_z(\tau)j_z^*(\tau)d\tau. \quad (9)$$

We write the electric field E_z in terms of the current density $j_z(\tau)$,

$$E_z(\tau) = -\frac{a}{4}\sqrt{\frac{\mu}{\varepsilon}}\int_{-1}^{1} j_z(t)H_0^{(2)}(a|\tau-t|)dt, \quad (10)$$

and substitute this expression into (9) to obtain

$$P = i\sqrt{\frac{\mu}{\varepsilon}}\frac{\tilde{a}\pi}{8a}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{\chi^2-1}}|F(\chi)|^2 d\chi. \quad (11)$$

For energy integral (12) to be finite, it is sufficient that

$$\int_{-\infty}^{+\infty}\frac{1}{\left|\sqrt{\chi^2-1}\right|}|F(\chi)|^2 d\chi < +\infty. \quad (12)$$

A comparison of (12) and (3) shows that the space $L_2[-1, 1]$ has become narrower: there exist currents that satisfy (12) but do not belong to $L_2[-1, 1]$.

SPACE OF CURRENTS IN THE *H*-POLARIZATION PROBLEM

We introduce the space of currents by means of the operator

$$(Au)(\tau) = \frac{1}{\pi}\int_0^{+\infty}\chi\int_{-1}^{1}\cos[\chi(\tau-t)]u(t)dtd\chi \equiv \frac{1}{2\pi}\int_{-\infty}^{+\infty}|\chi|\int_{-1}^{1}\exp(-i\chi(\tau-t))dtd\chi. \quad (13)$$

It is known [8] that the operator *A* is symmetric and positive definite and its domain of definition *D*(*A*) is everywhere dense in $L_2[-1, 1]$:

$$(Au, u) \ge \gamma^2(u, u), \quad \forall u \in D(A), \quad \gamma^2 > 0. \quad (14)$$

Hereafter, (. , .) means a scalar product in $L_2[-1, 1]$. The positive definiteness of the operator *A* allows one to introduce the energy space H_A [9], which is defined as a completion of *D*(*A*) on the norm

$$[u]^2 = (Au, u), \quad (15)$$

and the scalar product in this space is defined as

$$[u, v] = (Au, v).$$

Next, let us establish a relationship between H_A and the Sobolev space $H_{\frac{1}{2}}([-1, 1])$, which can be regarded [10] as a completion of $C_0^\infty([-1, 1])$ (the set of infinitely differentiable finite functions with the support [–1, 1]) on the norm

$$\|u\|_{\frac{1}{2}}^2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty}(1+|\chi|)|\tilde{u}(\chi)|^2 d\chi, \quad (16)$$

where

$$\tilde{u}(\chi) = \int_{-1}^{1} u(t)\exp(i\chi t)dt.$$

The positive definiteness of the operator A immediately yields the equivalence of norms (15) and (16). Consequently, the energy space H_A coincides with the Sobolev space $H_{\frac{1}{2}}([-1, 1])$. A significant advantage of the norm and the scalar product introduced above is that they allow one to analytically introduce an orthonormal basis of H_A [8] in the form

$$\varphi_n(\tau) = \sqrt{\frac{2}{\pi n}}\sin[n\arccos(\tau)]; \quad n = 1, 2, \ldots,$$
$$(A\varphi_m, \varphi_n) = \begin{cases} 0, & m \neq n \\ 1, & m = n. \end{cases} \tag{17}$$

To conclude this section, write the operator A in the coordinate form:

$$(Au)(\tau) = \frac{1}{\pi}\frac{\partial}{\partial\tau}\int_{-1}^{1} u(t)\frac{\partial}{\partial t}\ln\frac{1}{|\tau - t|}dt. \tag{18}$$

The equivalence of representations (13) and (18) for the operator A on a dense set is validated by the well-known expression

$$\ln\frac{1}{|\tau - t|} = C + \int_{0}^{1}\frac{\cos[\chi(\tau - t)] - 1}{\chi}d\chi + \int_{1}^{+\infty}\frac{\cos[\chi(\tau - t)]}{\chi}d\chi, \tag{19}$$

where C is the Euler constant.

SPACE OF PATTERNS IN THE *H*-POLARIZATION PROBLEM: BASIS

We introduce Sobolev spaces [11] on the straight line $H_s(-\infty, +\infty)$ as a completion of the set $C_0^\infty(-\infty, +\infty)$ of infinitely differentiable finite functions on the norm

$$\|\tilde{u}\|_s^2 = \frac{1}{2\pi}\int_{-\infty}^{+\infty}(1 + |\chi|)^{2s}|\tilde{u}(\chi)|^2 d\chi < +\infty, \tag{20}$$

where

$$\tilde{u}(\chi) = \int_{-\infty}^{+\infty} u(t)\exp(i\chi t)dt.$$

The constant factor preceding the integral in the right-hand side of (20) is introduced for convenience. Designate the Fourier transform for the space $H_s(-\infty, +\infty)$ as $\tilde{H}_s(-\infty, +\infty)$. In the latter, the norm is also defined by relationship (20). The spaces $H_s(-\infty, +\infty)$ and $\tilde{H}_s(-\infty, +\infty)$ are Hilbert spaces with the scalar product

$$\langle u, v\rangle_s = \frac{1}{2\pi}\int_{-\infty}^{+\infty}(1 + |\chi|)^{2s}\tilde{u}(\chi)\overline{\tilde{v}(\chi)}d\chi. \tag{21}$$

The operator K_1, or the Fourier transform as an operator, which maps from the space $H_{\frac{1}{2}}([-1, 1])$ into the space $\tilde{H}_{\frac{1}{2}}(-\infty, +\infty)$, is an isomorphism; this mapping does not change the norm. Therefore, the image of $H_{\frac{1}{2}}([-1, 1])$ is a closed set on which the inverse operator K_1^{-1} is defined and limited. At the same time, the Sobolev space $H_{\frac{1}{2}}([-1, 1])$ coincides with the energy space H_A, in which the basis is analytically specified. Therefore, we will assume that the operators act in the energy space H_A.

We consider again Eq. (1),

$$F(\chi) = K_a j = a\int_{-1}^{1}\exp(ia\chi t)j(t)dt, \tag{22}$$

and map it from the energy space $j \in H_A$ into the Sobolev space $F \in \tilde{H}_{\frac{1}{2}}(-\infty, +\infty)$. Define the space of patterns as the image of H_A due to the mapping K_a : Im(K_a). On the set Im(K_a), the scalar product is given by

$$(K_a u, K_a v)_{\frac{1}{2}} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} K_a u\overline{K_a v}|\chi|d\chi, \tag{23}$$

which is equivalent to scalar product (21). This statement follows from the positive definiteness of the operator A. The set Im(K_a), being a closed set, is a Hilbert space with scalar product (23).

We define the basis functions f_n^a of this space through the basis functions φ_n of the space of currents:

$$f_n^a = K_a\varphi_n. \tag{24}$$

With definition (17) of the basis functions φ_n (17) and the definition of the operator K_a, the basis functions

f_n^a are found as

$$f_n^a(\chi) = \begin{cases} (-1)^{k-1}\sqrt{2\pi(2k-1)}\dfrac{J_{2k-1}(a\chi)}{\chi}, & n = 2k-1 \\ i(-1)^{k-1}\sqrt{4\pi k}\dfrac{J_{2k}(a\chi)}{\chi}, & n = 2k, \end{cases} \quad (25)$$

where J_n are Bessel functions.

Functions (25) were derived using the integrals of the products of trigonometric functions and Chebyshev polynomials of the second kind. The basis functions f_n^a are orthonormal by construction:

$$(f_n^a, f_m^a)_{\frac{1}{2}} = \begin{cases} 0, & m \neq n \\ 1, & m = n. \end{cases} \quad (26)$$

Relationship (26) can also be verified using the tabulated integral

$$\int_0^{+\infty} \frac{J_m(\chi)J_n(\chi)}{\chi} d\chi = \frac{2}{\pi(m^2-n^2)} \sin\frac{m-n}{2}\pi. \quad (27)$$

An arbitrary function F belonging to the space of patterns $\mathrm{Im}(K_a)$ can be expanded over the orthonormal basis f_n^a:

$$F(\chi) = \sum_{n=1}^{+\infty} (F, f_n^a)_{\frac{1}{2}} f_n^a(\chi). \quad (28)$$

From an expansion for the pattern, we immediately obtain an expansion for the current:

$$j(\tau) = \sum_{n=1}^{+\infty} (F, f_n^a)_{\frac{1}{2}} \varphi_n(\tau). \quad (29)$$

As a result, the problem of finding the current from a given realizable radiation pattern is reduced to the problem of expanding over a given orthonormal basis. Thus, the above-stated problem of finding the current from a given realizable radiation pattern in the case of H polarization has been completely solved.

SPACE OF CURRENTS IN THE E-POLARIZATION PROBLEM

We introduce the space of currents by means of the operator

$$(Lu)(\tau) = \frac{1}{\pi}\int_{-1}^{1} u(t)\ln\frac{1}{|\tau - t|} dt. \quad (30)$$

We also need the weight spaces $L_{2,q}$ with the scalar product defined as

$$(u, v)_{2,q} = \int_{-1}^{1} u(t)\overline{v(t)}q(t)dt. \quad (31)$$

Consider an operator L that maps the weight space $L_{2,\rho}$ into the weight space $L_{2,\rho^{-1}}$, where $\rho(t) = \sqrt{1-t^2}$. Chebyshev polynomials of the first kind $T_n(t) = \cos[(n-1)\arccos(t)]$ $(n = 1, 2, \ldots)$ form a basis of the space $L_{2,\rho^{-1}}$, while weighted Chebyshev polynomials $T_n(t)/\rho$ form a basis of the space $L_{2,\rho}$. The following relationship is valid for the operator L [12]:

$$L(T_n/\rho)(\tau) = \begin{cases} \ln 2, & n = 1 \\ \dfrac{1}{n-1}T_n(\tau), & n \neq 1. \end{cases} \quad (32)$$

Let I denote an identity operator that maps from the space $L_{2,\rho^{-1}}$ into the space $L_{2,\rho}$ and relates a function $u(\tau)$ to a function $u(\tau)/\rho(\tau)$. We consider the operator

$$IL: L_{2,\rho} \longrightarrow L_{2,\rho}. \quad (33)$$

As follows from relationship (32), this operator is positive. The positiveness (this property alone) of the operator IL also allows the introduction of the energy space H_L [9], which is defined as a completion of $L_{2,\rho}$ on the norm

$$[u]^2 = (ILu, u)_{2,\rho}. \quad (34)$$

The space H_L is a Hilbert space with the scalar product

$$[u, v] = (ILu, v)_{2,\rho}. \quad (35)$$

According to (32), the orthonormal basis of this space has the form

$$\psi_n(\tau) = \begin{cases} \sqrt{\dfrac{1}{\pi\ln 2}}\dfrac{1}{\sqrt{1-\tau^2}}, & n = 1 \\ \sqrt{\dfrac{2(n-1)}{\pi}}\dfrac{\cos[(n-1)\arccos(\tau)]}{\sqrt{1-\tau^2}}, & n > 1, \end{cases}$$

$$(IL\psi_m, \psi_n) = \begin{cases} 0, & m \neq n \\ 1, & m = n. \end{cases} \quad (36)$$

The power space H_L can be shown to coincide with the Sobolev space $H_{\frac{1}{2}}([-1, 1])$, which is defined as a completion of $C_0^\infty([-1, 1])$ (the set of infinitely differ-

entiable finite functions with the support [–1, 1]) on the norm

$$\|u\|^2_{H_{\frac{1}{2}}} = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{|\tilde{u}(\chi)|^2}{1+|\chi|}d\chi, \tag{37}$$

where

$$\tilde{u}(\chi) = \int_{-1}^{1} u(t)\exp(i\chi t)dt.$$

Calculate the Fourier transform of the basis functions:

$$K_1\psi_n = \int_{-1}^{1}\exp(i\chi t)\psi_n(t)dt$$

$$= \begin{cases} \sqrt{\dfrac{\pi}{\ln 2}}J_0(\chi), & n = 1 \\ (-1)^{k-1}\sqrt{2\pi(2k-1)}J_{2k-2}(\chi), & n = 2k-1 \\ i(-1)^{k-1}\sqrt{4\pi k}J_{2k}(\chi), & n = 2k. \end{cases} \tag{38}$$

Using formula (38) and methods presented in [8], one can easily prove that norms (34) and (37) are equivalent. Note in conclusion that the energy space H_L corresponds to the *E*-polarization problem: if the function describing the current belongs to H_L, then energy integral (12) is finite.

SPACE OF PATTERNS IN THE *E*-POLARIZATION PROBLEM: BASIS

The operator K_1, or the Fourier transform as an operator, mapping from the space $H_{\frac{1}{2}}([-1, 1])$ into the space $\tilde{H}_{\frac{1}{2}}(-\infty, +\infty)$, also is an isomorphism; this mapping does not change the norm. Therefore, the image of $H_{\frac{1}{2}}([-1, 1])$ is a closed set on which the inverse operator K_1^{-1} is defined and limited. The Sobolev space $H_{\frac{1}{2}}([-1, 1])$ coincides with the energy space H_L, in which the basis is given analytically. Therefore, we consider Eq. (1) once more,

$$F(\chi) = K_a j = \int_{-1}^{1}\exp(ia\chi t)j(t)dt, \tag{39}$$

and map it from the energy space $j \in H_L$ into the Sobolev space $F \in \tilde{H}_{\frac{1}{2}}(-\infty, +\infty)$. Equation (39) differs from Eq. (1) by a constant factor.

We define the space of radiation patterns as the image of H_L due to the mapping K_a : $\mathrm{Im}(K_a)$. In this space, we introduce a scalar product equivalent to (21) by the formula

$$(f, g)_{\frac{1}{2}} = \frac{C}{\pi}f(0)\overline{g(0)} + \frac{1}{2\pi}\int_{|\chi|<1}\frac{f(\chi)\overline{g(\chi)} - f(0)\overline{g(0)}}{|\chi|}d\chi + \frac{1}{2\pi}\int_{|\chi|>1}\frac{f(\chi)\overline{g(\chi)}}{|\chi|}d\chi, \tag{40}$$

where $f = K_a u$, $g = K_a v$, and C is the Euler constant.

If f or g is zero at $\chi = 0$ (all functions $K_1\psi_n$ possess this property at $n > 1$), scalar product (40) simplifies to

$$(f, g)_{\frac{1}{2}} = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{f(\chi)\overline{g(\chi)}}{|\chi|}d\chi.$$

The set $\mathrm{Im}(K_a)$, being a closed set, is a Hilbert space with scalar product (40). As follows from (32), the basis functions g_n^a of this space have the form

$$g_n^a = \int_{-1}^{1}\exp(ia\chi t)\psi_n(t)dt$$

$$= \begin{cases} \sqrt{\dfrac{\pi}{\ln 2}}J_0(a\chi), & n = 1 \\ (-1)^{k-1}\sqrt{2\pi(2k-1)}J_{2k-2}(a\chi), & n = 2k-1 \\ i(-1)^{k-1}\sqrt{4\pi k}J_{2k}(a\chi), & n = 2k. \end{cases} \tag{41}$$

An arbitrary function F belonging to the class of realizable patterns $\mathrm{Im}(K_a)$ can be expanded over the orthogonal basis g_n^a:

$$F(\chi) = \sum_{n=1}^{+\infty} c_n g_n^a(\chi), \tag{42}$$

where

$$c_n = \begin{cases} \dfrac{(F, g_n^a)_{\frac{1}{2}}}{(g_n^a, g_n^a)_{\frac{1}{2}}}, & n = 1 \\ (F, g_n^a)_{\frac{1}{2}}, & n > 1. \end{cases}$$

It should be noted that the basis g_n^a is orthonormal at $n > 1$. From an expansion for the pattern, we immediately obtain an expansion for the current:

$$j(\tau) = \sum_{n=1}^{+\infty} c_n\psi_n(\tau). \tag{43}$$

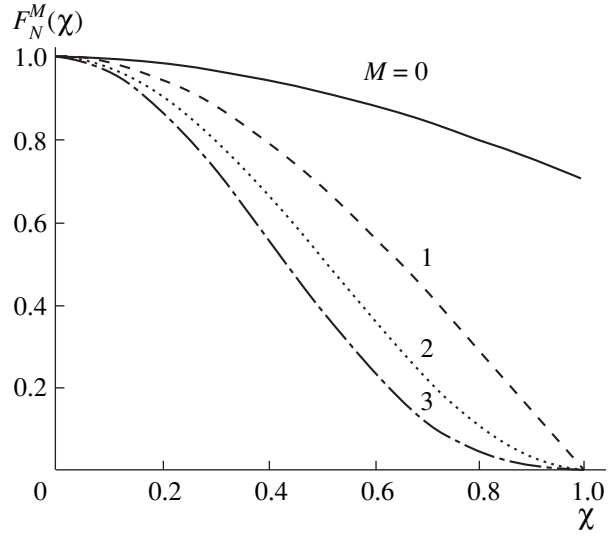

Fig. 1.

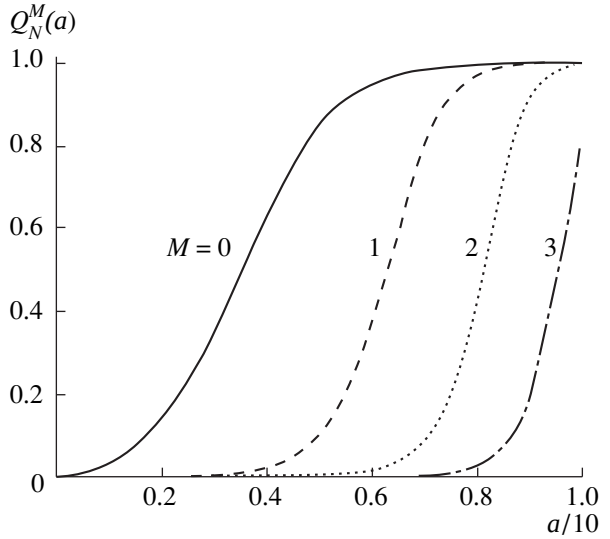

Fig. 2.

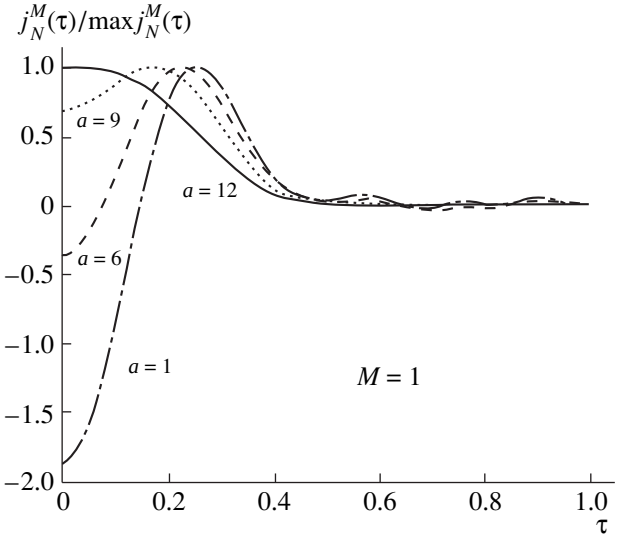

Fig. 3.

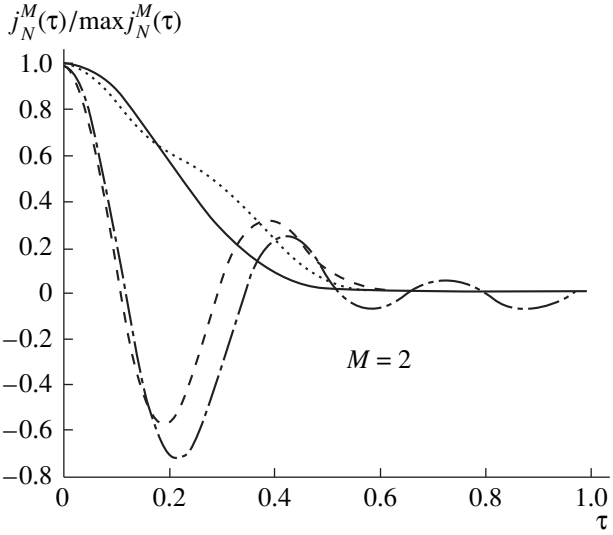

Fig. 4.

The problem of finding the current from a given realizable pattern in the case of *E* polarization has also been solved completely.

NUMERICAL RESULTS

We consider an example of calculating the current from a given radiation pattern, e.g., in the *H*-polarization problem. Let the pattern have the form

$$F_N^M(\chi) = (1-\chi^2)^M \sin(a\chi)\left[a\chi\prod_{n=1}^{N}\left(1-\frac{a^2\chi^2}{\pi^2 n}\right)\right]^{-1}.$$

This radiation pattern is realizable [2]. By selecting appropriate *M* and *N*, one can obtain various narrow-beam patterns without secondary lobes (Fig. 1). However, in this case, the inequality $M < N$ must be satisfied (in Figs. 1–4, $N = 10$). The radiation pattern $F_N^M(\chi)$ has only a weak dependence on the length of a radiator *a*. However, the radiator length strongly affects other parameters. Figure 2 plots the radiated-to-total power ratio against *a*. At small *a*, the radiated power is much smaller than the total power. As *a* increases, the ratio grows. Figures 3 and 4 illustrate curves for the current. At small *a*, the current oscillates and, as calculations show, is large. Though narrow-beam secondary-lobe-free patterns can be produced at small *a*'s, such an approach is extremely inefficient. As *a* increases, the current distribution becomes smoother. The narrower the pattern, the larger *a* must be to increase the power ratio and, accordingly, to smooth out the current.

REFERENCES

1. E. V. Zakharov and Yu. V. Pimenov, *Numerical Analysis of Radio-Wave Diffraction* (Radio i Svyaz', Moscow, 1982).
2. E. G. Zelkin, *Construction of Radiating Systems from Given Radiation Pattern* (Énergoizdat, Moscow, 1963).
3. B. M. Minkovich and V. P. Yakovlev, *Theory of Synthesis of Antennas* (Sov. Radio, Moscow, 1969).
4. L. D. Bakhrakh and S. D. Kremenetskiĭ, *Synthesis of Radiating Systems: Theory and Computing Methods* (Sov. Radio, Moscow, 1974).
5. Ya. N. Fel'd, Radiotekh. Élektron. (Moscow) **32**, 1137 (1987).
6. B. Z. Katsenelenbaum, *Approximability of Electromagnetic Field* (Nauka, Moscow, 1996).
7. H. Hönl, A. W. Maue, and K. Westpfahl, in *Handbuch der Physik*, Vol. 25: *Kristalloptik. Beugung*, Ed. by S. Flügge (Springer-Verlag, Berlin, 1961; Mir, Moscow, 1964), Part 1, p. 218.
8. V. N. Plotnikov, Yu. Yu. Radtsig, and S. I. Éminov, Zh. Vychisl. Mat. Mat. Fiz. **34** (1), 68 (1994).
9. S. G. Mikhlin, *Variational Methods in Mathematical Physics* (Nauka, Moscow, 1970; Pergamon, Oxford, 1964).
10. J.-L. Lions and E. Magenes, *Problemes aux limites non homogénes et applications* (Dunod, Paris, 1968; Mir, Moscow, 1971).
11. G. I. Éskin, *Boundary-Value Problems for Elliptic Pseudodifferential Equations* (Nauka, Moscow, 1973).
12. I. I. Vorovich, V. M. Aleksandrov, and V. A. Babeshko, *Nonclassical Mixed Problems in Theory of Elasticity* (Nauka, Moscow, 1974).

Translated by A. Khzmalyan

Technical Physics, Vol. 46, No. 2, 2001, pp. 219–222. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 89–93.

RADIOPHYSICS

Spectra of Exchange Dipole Electromagnetic–Spin Waves in Asymmetric Metal–Insulator–Ferromagnetic–Insulator–Metal Systems

V. E. Demidov and B. A. Kalinikos

St. Petersburg State University of Electrical Engineering, St. Petersburg, 197376 Russia

e-mail: eivt@eltech.ru

Received May 29, 2000

Abstract—Spectra of exchange dipole electromagnetic–spin waves in tangentially magnetized asymmetric planar-layer metal–insulator–ferromagnetic–insulator–metal (MIFIM) systems were studied theoretically. It is shown that symmetry breakdown due to a difference in the permittivities of the insulating layers may substantially enhance interaction between the spin and electromagnetic waves. This improves the electrical controllability of the dispersion properties of the spin waves, for example, by varying the permittivity of one of the layers that are in contact with the ferromagnetic film. Optimal geometries of the layer structures that provide the best electrical control are suggested. © *2001 MAIK "Nauka/Interperiodica".*

Spin waves in ferromagnetic films and layer structures are used to advantage in various microwave devices. The basic advantage of spin-wave devices is the possibility of their electrical tuning by varying the permanent magnetic field applied to the ferromagnetic film. However, this way of tuning suffers from drawbacks, such as large dimensions of the magnetic systems, low speed, and high power consumption. Tuning of spin-wave devices can be improved by using layer structures containing both ferromagnetic and ferroelectric layers. In such structures, a new mechanism of controlling the dispersion characteristics of spin waves propagating in layer waveguide systems appears, since ferroelectrics can change their permittivity under the action of a permanent magnetic field. This mechanism offers high speed and is not power-hungry.

It seems therefore logical to study the wave spectra in layer structures in which ferroelectric and ferromagnetic layers are in contact. Since ferroelectrics are of high permittivity, a theory of spin waves that includes electromagnetic delay (which is usually neglected) should be elaborated.

In [1], we developed a theory of exchange dipole electromagnetic and spin waves propagating in symmetric layer MIFIM systems. It was shown that such systems allow the control of the spin wave spectrum by varying the permittivity of the ferroelectric layers. Based on this effect, a variety of spin-wave devices for processing microwave signals tuned through a change in the constant electric field can be designed. However, the geometry of the symmetric layer structure is hardly feasible.

This work is aimed at investigating the spectrum of exchange dipole electromagnetic and spin waves in asymmetric tangentially magnetized MIFIM systems and at finding the optimum geometry that provides the best controllability and is easy to implement in practice.

Consider an unbounded plane-parallel layer structure in the $y0z$ plane (Fig. 1). It includes an isotropic ferromagnetic film of thickness L that has a saturated magnetization M_s and a permittivity ϵ_L. The film is separated from two perfectly conducting metal screens by insulating layers of thicknesses a and b and permittivities ϵ_a and ϵ_b, respectively. The origin is placed at the center of the ferromagnetic film. It is assumed that the film is magnetized to saturation by a permanent magnetic field of strength H_0 applied along the z axis. The spins on the surfaces of the film are assumed to be noninteracting.

For convenience, we introduce the second coordinate system $\xi\eta\zeta$ rotated through an angle φ about the x axis. The ζ axis coincides with the direction of wave propagation.

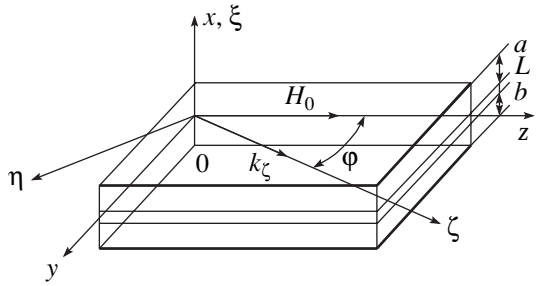

Fig. 1. Layer structure geometry.

1063-7842/01/4602-0219$21.00 © 2001 MAIK "Nauka/Interperiodica"

In the structure considered, a dispersion relation for exchange dipole spin and electromagnetic waves is derived by jointly integrating the complete set of the Maxwell equations and equations for magnetization motion using the Green tensor function formalism for planar-layer structures [2, 3]. The result is a transcendental equation that relates the eigenfrequency ω of the nth mode of spin waves to the longitudinal wave number k_ζ:

$$(\Omega_{nk} - \omega_M A_n^{xx})[\Omega_{nk} - \omega_M(A_n^{yy}\cos^2\varphi + A_n^{zz}\sin^2\varphi)] - (\omega + \omega_M A_n^{xz}\sin\varphi)(\omega - \omega_M A_n^{zx}\sin\varphi) = 0, \quad (1)$$

where

$$A_n^{xx} = -1 + \frac{k_\zeta^2}{\gamma_L^2 + \kappa_n^2} - \frac{1}{\sinh(\gamma_L d)N}\frac{k_\zeta^2\gamma_L}{(\gamma_L^2 + \kappa_n^2)}\frac{2}{L(1+\delta_{0n})} \times [C_n^1(D_b(-1)^n + T_b) - C_n^2(D_a - T_a(-1)^n)],$$

$$A_n^{xz} = \frac{1}{\sinh(\gamma_L d)}\frac{k_\zeta\gamma_L^2}{(\gamma_L^2 + \kappa_n^2)}\frac{2}{L(1+\delta_{0n})} \times [C_n^3(D_b(-1)^n + T_b) + C_n^4(D_a - T_a(-1)^n)],$$

$$A_n^{yy} = \frac{k_{0L}^2}{\gamma_L^2 + \kappa_n^2} - \frac{1}{\sinh(\gamma_L d)N^y}\frac{k_{0L}^2\gamma_L}{(\gamma_L^2 + \kappa_n^2)^2}\frac{2}{L(1+\delta_{0n})} \times\left[C_n^3\frac{\epsilon_L\gamma_a}{\epsilon_a\gamma_L}\tanh(\gamma_a a)D_b^y(-1)^n - \frac{\epsilon_L\gamma_b}{\epsilon_b\gamma_L}\tanh(\gamma_b b)T_b^y\right] - C_n^4\left(\frac{\epsilon_L\gamma_b}{\epsilon_b\gamma_L}\tanh(\gamma_b b)D_a^y + \frac{\epsilon_L\gamma_a}{\epsilon_a\gamma_L}\tanh(\gamma_a a)T_a^y(-1)^n\right),$$

$$A_n^{zx} = \frac{1}{\sinh(\gamma_L d)N}\frac{k_\zeta\gamma_L^2}{(\gamma_L^2 + \kappa_n^2)^2}\frac{2}{L(1+\delta_{0n})} \times\left[C_n^1\left(\frac{\gamma_L}{\gamma_a}\tanh(\gamma_a a)D_b(-1)^n - \frac{\gamma_L}{\gamma_b}\tanh(\gamma_b b)T_b\right) + C_n^2\left(\frac{\gamma_L}{\gamma_b}\tanh(\gamma_b b)D_a + \frac{\gamma_L}{\gamma_a}\tanh(\gamma_a a)T_a(-1)^n\right)\right],$$

$$A_n^{zz} = -\frac{\gamma_L^2}{\gamma_L^2 + \kappa_n^2} + \frac{1}{\sinh(\gamma_L d)N}\frac{\gamma_L^3}{(\gamma_L^2 - \kappa_n^2)^2}\frac{2}{L(1+\delta_{0n})} \times\left[C_n^3\left(\frac{\gamma_L}{\gamma_a}\tanh(\gamma_a a)D_b(-1)^n - \frac{\gamma_L}{\gamma_b}\tanh(\gamma_b b)T_b\right) - C_n^4\left(\frac{\gamma_L}{\gamma_b}\tanh(\gamma_b b)D_a + \frac{\gamma_L}{\gamma_a}\tanh(\gamma_a a)T_a(-1)^n\right)\right],$$

$$C_n^1 = \cosh(\gamma_L(a+L))(-1)^n - \cosh(\gamma_L a),$$

$$C_n^2 = \cosh(\gamma_L b)(-1)^n - \cosh(\gamma_L(b+L)),$$

$$C_n^3 = \sinh(\gamma_L(a+L))(-1)^n - \sinh(\gamma_L a),$$

$$C_4^n = \sinh(\gamma_L b)(-1)^n - \sinh(\gamma_L(b+L)),$$

$$D_a = \sinh(\gamma_L(a+L)) + \frac{\gamma_L}{\gamma_a}\tanh(\gamma_a a)\cosh(\gamma_L(a+L)),$$

$$D_b = \sinh(\gamma_L(b+L)) + \frac{\gamma_L}{\gamma_b}\tanh(\gamma_b b)\cosh(\gamma_L(b+L)),$$

$$T_a = \sinh(\gamma_L a) - \frac{\gamma_L}{\gamma_b}\tanh(\gamma_b b)\cosh(\gamma_L a),$$

$$T_b = -\sinh(\gamma_L b) + \frac{\gamma_L}{\gamma_a}\tanh(\gamma_a a)\cosh(\gamma_L b),$$

$$N = \sinh(\gamma_L L)\left(1 + \frac{\gamma_L^2}{\gamma_a\gamma_b}\tanh(\gamma_a a)\tanh(\gamma_b b)\right) + \cosh(\gamma_L L)\left(\frac{\gamma_L}{\gamma_b}\tanh(\gamma_b b) + \frac{\gamma_L}{\gamma_a}\tanh(\gamma_a a)\right),$$

$$d = a + b + L.$$

The elements N^y, T_i^y, and D_i^y are obtained from N, T_i, and D_i by the replacement

$$\frac{\gamma_L}{\gamma_a} \longrightarrow \frac{\epsilon_L\gamma_a}{\epsilon_a\gamma_L}, \quad \frac{\gamma_L}{\gamma_b} \longrightarrow \frac{\epsilon_L\gamma_b}{\epsilon_b\gamma_L}.$$

The other designations are the same as in [1].

In parallel with dispersion relation (1) for the exchange dipole spin waves, one can deduce a dispersion relation for the natural waves of the structure in the absence of magnetization in the layer L. For the *TE* mode,

$$\frac{N}{\gamma_L} = 0, \quad (2)$$

and for the *TM* mode,

$$\gamma_L N^y = 0. \quad (3)$$

Note that each of the factors in Eqs. (2) and (3) depends both on wave number k_ζ and on frequency ω.

Practical spin-wave devices usually employ two types of spin waves: longitudinal waves, propagating in the direction of the permanent magnetic field ($\varphi = 0$), and transverse waves, traveling at a right angle to the field ($\varphi = \pi/2$).

Longitudinal spin waves interact with lower order TM_0 electromagnetic waves, which are free of cutoff [1]. For longitudinal spin waves, a dispersion relation

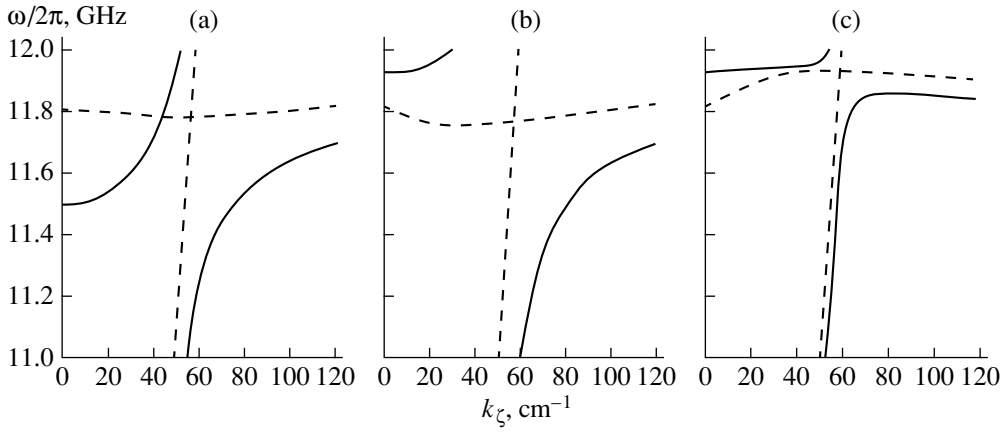

Fig. 2. Change in the spectra of electromagnetic–spin waves when symmetry breakdown is due to different thicknesses of the *a* and *b* layers.

may relate the slope of the dispersion curve for the TM_0 waves with ϵ_a and ϵ_b. In symmetric MIFIM structures, such a control mechanism was shown to be inefficient [1]. As follows from Eq. (1), if the symmetry breaks down, this mechanism becomes still less efficient, because any asymmetry weakens interaction between longitudinal spin waves and electromagnetic TM_0 waves.

Transverse spin waves extensively interact with TE_1 electromagnetic waves. They have a cutoff frequency that varies with the permittivity of the ferroelectric layer. If the cutoff frequency lies close to the frequency range of long (low-frequency) spin waves, the dispersion curve for the spin waves considerably changes.

Thus, transverse spin waves appear to be the most promising for control of the dispersion by varying the permittivity of ferroelectrics. Therefore, in subsequent analysis of asymmetric layer structures, we will consider transverse spin waves alone.

Magnetostatic studies of the dispersion characteristics of spin waves in asymmetric layer MIFIM structures [2, 3] show that transverse spin waves have the property of nonreciprocity; that is, their dispersion characteristics depend on the sign of the longitudinal wave number k_ζ. One more nonreciprocity mechanism emerges when electromagnetic delay is taken into consideration. In this case, transverse spin waves propagating in the opposite directions along the ζ axis interact with electromagnetic waves with various efficiencies.

Figure 2 shows the spectrum of transverse electromagnetic–spin waves when symmetry breakdown is due to various thicknesses of the insulating layers. In the calculations, the parameter values were as follows: L = 20 μm, ϵ_L = 14, $\epsilon_a = \epsilon_b$ = 1000, and H_0 = 3400 Oe. In Fig. 2a, $a = b$ = 300 μm; in Figs. 2b and 2c, a = 100 μm and b = 500 μm. The thicknesses of the ferroelectric layers were taken so as not to change the cutoff frequency of the TE_1 electromagnetic wave. In Figs. 2b and 2c, the waves propagate, respectively, in the positive and negative directions along the ζ axis. Dashed lines are the dispersion curves for TE_1 electromagnetic waves and transverse spin waves. In the former case, the curves were calculated for the unmagnetized L layer (fast waves), and in the latter, the magnetostatic approximation (slow waves) was used. Solid lines depict the spectrum of the electromagnetic and spin waves that follows from Eq. (1). It is seen that, in the asymmetric layer structure, the transverse spin waves propagating in the opposite directions of the ζ axis variously interact with the electromagnetic waves. Compared with the case of symmetric MIFIM structures, for the waves traveling in the positive and negative directions, the interaction efficiency grows and diminishes, respectively. This can easily be explained by considering the cross-sectional distribution of the magnetic fields h_ξ and h_ζ, which provide interaction between transverse spin and TE_1 electromagnetic waves. It turns out that the polarizations of these fields for both waves mainly coincide in one case and differ in the other.

Figure 3 shows the variation of the spectrum of the transverse spin waves when symmetry breakdown is associated with different permittivities of the layers. Here, L = 20 μm, ϵ_L = 14, $a = b$ = 300 μm, and H_0 = 3400 Oe. In Fig. 3a, $\epsilon_a = \epsilon_b$ = 1000; in Figs. 3b and 3c, ϵ_a = 1630 and ϵ_b = 14. As in the case when the layers a and b are heterogeneous in thickness, a difference in their permittivities causes the nonreciprocity of the waves traveling in the opposite directions of the ζ axis. For those traveling in the negative direction, interaction is stronger; while for those propagating in the negative direction, weaker, as compared with the symmetric structure.

The above analysis implies that the asymmetric structure can provide stronger interaction between transverse spin and electromagnetic waves. Note also that the layer structure shown in Fig. 3c is quite feasible: the properties of the layers L and b are close to those of epitaxial YIG films grown on GGG substrates.

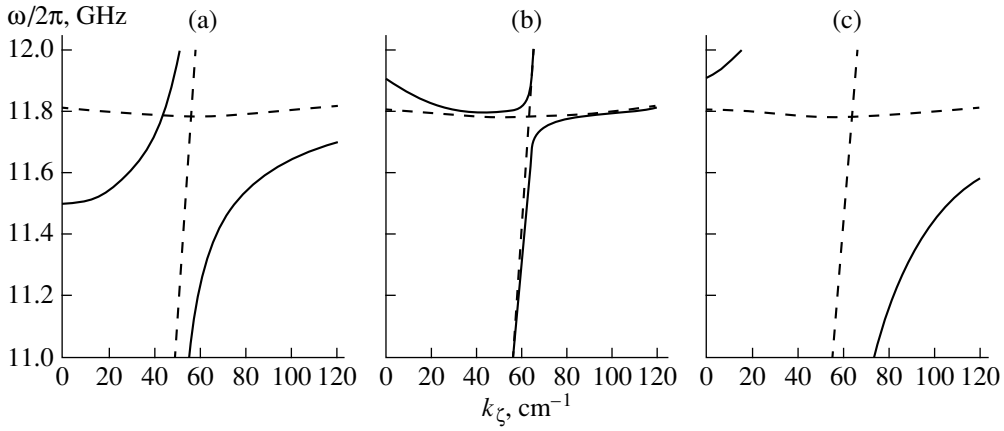

Fig. 3. Change in the spectra of electromagnetic–spin waves when symmetry breakdown is due to different permittivities of the *a* and *b* layers.

Enhanced interaction between electromagnetic and spin waves in asymmetric layer structures makes it possible to improve the efficiency of controlling the spectrum of exchange dipole spin waves by varying the permittivities of the layers. Thus, the asymmetric layer structures offer greater promise for devices with electrically tunable properties.

Figure 4 exemplifies the control of the dispersion characteristics of transverse spin waves traveling in the negative ζ-axis direction. Here, the permittivity of the layer *a* in the asymmetric layer structure (L = 20 μm, ϵ_L = 14, $a = b$ = 300 μm, and ϵ_b = 14) varies from 1400 to 800. The calculation is made for the permanent magnetic field H_0 = 3000 Oe. The dashed line is the dispersion curve of the spin waves in the magnetostatic approximation, and solid lines show dispersion curves obtained from Eq. (1) at various ϵ_a (the associated values are indicated by figures).

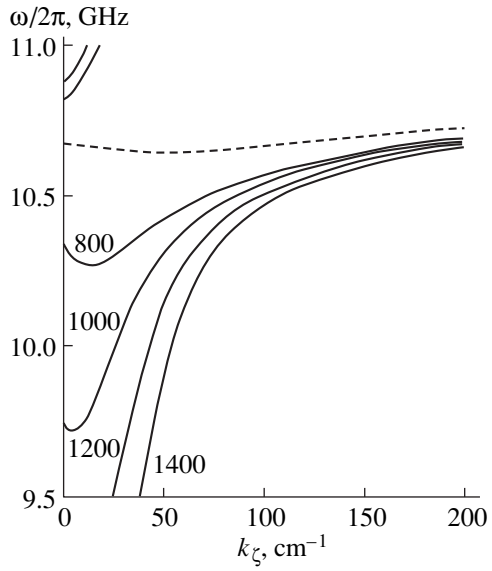

Fig. 4. Control of spin wave dispersion characteristics in the asymmetric layer structure.

A comparison of our Fig. 4 with Fig. 4 in [1], showing the control of the dispersion characteristics for transverse spin waves in the symmetric structure, clearly points to a better controllability of the spin wave spectrum in the asymmetric structure. As ϵ_a varies, the spin wave dispersion curve changes in a much wider range of longitudinal wave numbers. Moreover, in the asymmetric structure, the frequency range of the upper branch of hybridized dispersion curves is higher than in the symmetric one. This provides single-mode propagation of transverse spin waves throughout the range of ϵ_a.

Thus, the asymmetric structures are candidates for spin-wave devices intended for processing microwave signals. They are easier to fabricate and offer better tuning capabilities than the symmetric ones.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (grant no. 99-02-16370) and by the Ministry of Education of the Russian Federation (grant no. 97-8.3-13).

REFERENCES

1. V. E. Demidov and B. A. Kalinikos, Pis'ma Zh. Tekh. Fiz. **26** (7), 8 (2000) [Tech. Phys. Lett. **26**, 272 (2000)].
2. V. F. Dmitriev and B. A. Kalinikos, Izv. Vyssh. Uchebn. Zaved., Fiz. **31** (11), 24 (1988).
3. B. A. Kalinikos, Izv. Vyssh. Uchebn. Zaved., Fiz. **24** (8), 42 (1981).

Translated by V. Isaakyan

Technical Physics, Vol. 46, No. 2, 2001, pp. 223–226. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 94–97.

A Charged Ellipsoidal Bunch in a Magnetic Field

A. S. Chikhachev

Khrunichev State Space Research and Production Center, Moscow, 121087 Russia

Received November 9, 1999; in final form, March 13, 2000

Abstract—The states of a long rotating charged ellipsoidal bunch in a longitudinal uniform magnetic field are studied. The states are described using two integrals of motion that couple the transverse velocities $\dot{x}$ and $\dot{y}$ with the x and y coordinates; the frequency $\omega_H = eH/mc$ (where H is the total magnetic field); and the quantities ω_1 and ω_2, which characterize the Coulomb repulsion in the x and y directions. It is shown that equilibrium states with a high charge density per unit length ($\nu \gtrsim 1$) can exist. © *2001 MAIK "Nauka/Interperiodica".*

Rapid progress in accelerator technology during the last few decades has required the development of methods for solving problems related to the dynamics of charged particles under different conditions. Of special interest are situations in which charged particles interact with the self-fields of dense bunches. In this paper, a charged bunch rotating in a magnetic field is studied with allowance for the interaction of the particles with the self-field of the bunch.

Note that the behavior of long charged-particle bunches (or beams) with elliptic cross sections in a longitudinal magnetic field is as yet poorly investigated. In [1], a beam with an elliptic cross section in a magnetic field in the presence of an external quadrupole structure was studied. A nonsteady ellipsoidal bunch in the absence of a magnetic field was investigated in [2].

Diamagnetic properties of beams with circular cross sections were studied within the so-called rigid rotator model [3]. In this paper, the ellipsoidal bunch is described using another model that differs from the rigid rotator model in that the average angular velocity is assumed to depend on the angle. The effect of the interaction delay, which may be important for dense bunches, is also studied.

1. We consider the motion of charged particles in a magnetic field aligned with the z-axis with allowance for repulsion caused by the particle space charge. Let the x- and y-axes be directed along the principal axes of the elliptic cross section of a rotating bunch. In this case, the equations of motion have the form

$$\begin{aligned}\ddot{x} - \dot{y}(2\dot{\Theta} + \omega_H) - x\left(\dot{\Theta} + \frac{\omega_H}{2}\right)^2 &= \left(\omega_1^2 - \frac{\omega_H^4}{4}\right)x,\\ \ddot{y} + \dot{x}(2\dot{\Theta} + \omega_H) - y\left(\dot{\Theta} + \frac{\omega_H}{2}\right)^2 &= \left(\omega_2^2 - \frac{\omega_H^4}{4}\right)y,\end{aligned} \tag{1}$$

where $\dot{\Theta}$ is the angular rotational velocity of the bunch, $\omega_H = eH/mc$, H is the total magnetic field, and the parameters ω_1 and ω_2 describe the Coulomb repulsion of the particles in the ellipsoidal bunch.

In the frame of reference related to the principal axes, we have

$$\omega_1^2 = \frac{\nu c^2}{R_x(R_x + R_y)}, \quad \omega_2^2 = \frac{\nu c^2}{R_y(R_x + R_y)},$$

where

$$\nu = \frac{3e^2N}{mc^2R_z}$$

is the charge density per unit length and N is the total number of particles in the bunch. It is assumed that longitudinal size of the bunch is much larger than its transverse dimensions ($R_z \gg R_x, R_y$).

Equations (1) have periodic solutions $x \sim e^{i\Omega t}$ and $y \sim e^{i\Omega t}$ with

$$\begin{aligned}\Omega_{1,2}^2 &= \frac{\omega_H^2 + 2\dot{\Theta}^2 + 2\dot{\Theta}\omega_H - \omega_1^2 - \omega_2^2}{2}\\ &\pm \sqrt{\left(\dot{\Theta} + \frac{\omega_H}{2}\right)^2(\omega_H^2 - 2(\omega_1^2 + \omega_2^2)) + \left(\frac{\omega_1^2 - \omega_2^2}{2}\right)^2}.\end{aligned} \tag{2}$$

From Eq. (8), which will be derived below, it follows that physically allowable (real) values of R_x/R_y can be obtained when

$$\frac{\omega_1^2 + \omega_2^2}{2} - \frac{\omega_H^2}{4} = \omega_0^2 - \frac{\omega_H^2}{4} = K_0^2 > 0.$$

This inequality substantially limits the domain of parameters in which nonsteady equilibrium states can exist. Taking into account the relationship

$$\delta_0^4 = -4K_0^2\left(\frac{\omega_H}{2} + \dot{\Theta}\right)^2 + \frac{\Delta^2}{4} \longrightarrow 0$$

$$(\Delta = \omega_1^2 - \omega_2^2),$$

1063-7842/01/4602-0223$21.00

it also follows from Eq. (8) that $\dot{\Theta} + \omega_H/2 \simeq \omega_0^2/\omega_H$, which results in real values of Ω. If $\dot{\Theta}$ and ω_H are constant, then Eqs. (1) have the following invariants:

$$I_1 = \left(\dot{x} + (\omega_H + 2\dot{\Theta})\frac{\omega_2^2 + \dot{\Theta}^2 + \dot{\Theta}\omega_H}{\omega_2^2 + \dot{\Theta}^2 + \dot{\Theta}\omega_H + \Omega_1^2}y\right)^2$$
$$+\left(\frac{(\omega_H + 2\dot{\Theta})\Omega_1}{\omega_2^2 + \dot{\Theta}^2 + \dot{\Theta}\omega_H + \Omega_1^2}\dot{y} + -\frac{\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_1^2}{\Omega_1}x\right)^2,$$

$$I_2 = \left(\dot{y} - (\omega_H + 2\dot{\Theta})\frac{\omega_1^2 + \dot{\Theta}^2 + \dot{\Theta}\omega_H}{\omega_1^2 + \dot{\Theta}^2 + \dot{\Theta}\omega_H + \Omega_2^2}x\right)^2$$
$$+\left(\frac{(\omega_H + 2\dot{\Theta})\Omega_2}{\omega_1^2 + \dot{\Theta}^2 + \dot{\Theta}\omega_H + \Omega_3^2}\dot{x} + -\frac{\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2}{\Omega_2}y\right)^2.$$

Setting

$$I = \alpha_1 I_1 + \alpha_2 I_2,$$

where $\alpha_{1,2}$ are positive constants representing the distribution function in the form

$$f = \kappa\delta(I - v_0^2),$$

and integrating this function with respect to velocities, we obtain that the density is nonzero in an elliptic region in the x and y coordinates. The complete expression for the integral of motion I is

$$I = A\dot{x}^2 + B\dot{y}^2 + 2C_1\dot{x}y + 2C_2x\dot{y} + Dx^2 + Ey^2, \quad (3)$$

where

$$A = \alpha_1 + \alpha_2\frac{(\omega_H + 2\dot{\Theta})^2\Omega_2^2}{(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_1^2 + \Omega_2^2)^2},$$

$$B = \alpha_1\frac{(\omega_H + 2\dot{\Theta})^2\Omega_1^2}{(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2 + \Omega_1^2)^2} + \alpha_2,$$

$$C_1 = (\omega_H + 2\dot{\Theta})(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2)$$
$$\times\left(\frac{\alpha_1}{\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2 + \Omega_1^2} + \frac{\alpha_2}{\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_1^2 + \Omega_2^2}\right),$$

$$C_2 = -\frac{\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_1^2}{\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2}C_1, \quad (4)$$

$$D = (\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_1^2)^2$$
$$\times\left\{\frac{\alpha_1}{\Omega_1^2} + \alpha_2\frac{(\omega_H + 2\dot{\Theta})^2}{(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_1^2 + \Omega_2^2)^2}\right\},$$

$$E = (\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2)^2$$
$$\times\left\{\frac{\alpha_1(\omega_H + 2\dot{\Theta})^2}{(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2 + \Omega_1^2)^2} + \frac{\alpha_2}{\Omega_2^2}\right\}.$$

For the distribution function specified above, the semiaxes of the elliptic cross section of the bunch are determined by the formulas

$$\frac{R_x^2}{B} = \frac{v_0^2\Omega_1^2(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_1^2 + \Omega_2^2)^2}{\alpha_1\alpha_2(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_1^2)^2(\Omega_1^2 - \Omega_2^2)^2},$$
$$\frac{R_y^2}{A} = \frac{v_0^2\Omega_2^2(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2 + \Omega_1^2)^2}{\alpha_1\alpha_2(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2)^2(\Omega_1^2 - \Omega_2^2)^2}. \quad (5)$$

The expressions for ω_1^2 and ω_2^2 and Eq. (5) yield

$$R_x = \frac{v_0\Omega_1(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_1^2 + \Omega_2^2)^2}{\sqrt{\alpha_1\alpha_2}(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_1^2)(\Omega_1^2 - \Omega_2^2)}$$
$$\times\sqrt{\alpha_1 + \alpha_2\frac{(\omega_H + 2\dot{\Theta})^2\Omega_1^2}{(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2 + \Omega_1^2)^2}} \quad (6)$$
$$= \sqrt{\nu c}\frac{\omega_2}{\omega_1\sqrt{\omega_1^2 + \omega_2^2}},$$

$$R_y = \frac{v_0\Omega_2(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2 + \Omega_1^2)^2}{\sqrt{\alpha_1\alpha_2}(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_2^2)(\Omega_1^2 - \Omega_2^2)}$$
$$\times\sqrt{\alpha_1\frac{(\omega_H + 2\dot{\Theta})^2\Omega_2^2}{(\dot{\Theta}^2 + \dot{\Theta}\omega_H + \omega_1^2 + \Omega_2^2)^2} + \alpha_2} \quad (7)$$
$$= \sqrt{\nu c}\frac{\omega_1}{\omega_2\sqrt{\omega_1^2 + \omega_2^2}}.$$

Below, we consider the case $\delta_0 \longrightarrow 0$; i.e.,

$$\Omega_1 \simeq \Omega_2 \simeq \Omega_0, \quad |\delta_0/\Omega_0| \ll 1.$$

Then, from Eqs. (6) and (7), it follows that $v_0/\sqrt{\nu c} \ll 1$ in a rotating frame of reference. Dividing Eq. (6) by Eq. (7) results in the equation

$$\frac{R_x}{R_y} = \frac{\omega_2^2}{\omega_1^2} = \frac{\sqrt{\dot{\Theta} + \frac{\omega_H}{2} - K_0}}{\sqrt{\dot{\Theta} + \frac{\omega_H}{2} + K_0}}. \quad (8)$$

2. Let us calculate the longitudinal magnetic self-field of the bunch. The coordinates x_1 and y_1 in the rest frame are related to the coordinates x and y in the rotating frame of reference with the axes aligned with the

principal axes of the elliptic cross section of the bunch as follows:

$$x_1 = x\cos\Theta - y\sin\Theta, \quad y_1 = x\sin\Theta + y\cos\Theta.$$

Integrating the expressions

$$A_{x_1} = \frac{1}{c}\int\frac{j_{x_1}(\mathbf{r}')(\mathbf{r}_{1\perp}\mathbf{r}'_{\perp})d\mathbf{r}'}{r'^3},$$

$$A_{y_1} = \frac{1}{c}\int\frac{j_{y_1}(\mathbf{r}')(\mathbf{r}_{1\perp}\mathbf{r}'_{\perp})d\mathbf{r}'}{r'^3}$$

and taking into account the equalities

$$\bar{\dot{x}} = -\frac{C_1}{A}y, \quad \bar{\dot{y}} = -\frac{C_2}{B}x, \quad \rho_0 = \frac{eN}{\frac{4\pi}{3}R_xR_yR_z},$$

we obtain

$$A_{x_1} = \frac{2\pi R_xR_y\rho_0}{c(R_x+R_y)}\left\{x_1\left[R_x\left(\frac{C_2}{B}-\dot\Theta\right)\right.\right.$$

$$\left.+R_y\left(\frac{C_1}{A}+\dot\Theta\right)\right]\frac{\sin2\Theta}{2} + y_1\left[R_x\left(\frac{C_2}{B}-\dot\Theta\right)\frac{1-\cos2\Theta}{2}\right. \quad (9)$$

$$\left.\left.-R_y\left(\frac{C_1}{A}+\dot\Theta\right)\frac{1+\cos2\Theta}{2}\right]\right\},$$

$$A_{y_1} = \frac{2\pi R_xR_y\rho_0}{c(R_x+R_y)}\left\{x_1\left[-\frac{1+\cos2\Theta}{2}\left(\frac{C_2}{B}-\dot\Theta\right)R_x\right.\right.$$

$$\left.+\frac{1-\cos2\Theta}{2}\left(\frac{C_1}{A}+\dot\Theta\right)R_y\right] \quad (10)$$

$$\left.-y_1\left[\left(\frac{C_2}{B}-\dot\Theta\right)R_x+\left(\frac{C_1}{A}+\dot\Theta\right)R_y\right]\frac{\sin2\Theta}{2}\right\}.$$

From Eqs. (9) and (10), we find the magnetic self-field of the bunch

$$H_z^{(\mathrm{in})} = \frac{\partial A_{y_1}}{\partial x_1} - \frac{\partial A_{x_1}}{\partial y_1}.$$

The external field is characterized by $\Omega_H = eH^{(\mathrm{ext})}/mc$, and the self-field is characterized by the difference $\omega_H - \Omega_H$, which can be expressed as follows:

$$\omega_H - \Omega_H = \nu\dot\Theta + \nu\left(\frac{C_1}{A}\frac{R_y}{R_x+R_y} - \frac{C_2}{B}\frac{R_x}{R_x+R_y}\right). \quad (11)$$

According to Eqs. (4), for $\delta_0 \longrightarrow 0$, we have

$$\frac{C_1}{A} = -\left(\dot\Theta + \frac{\omega_H}{2} - K_0\right), \quad \frac{C_2}{B} = \dot\Theta + \frac{\omega_H}{2} + K_0.$$

Using expressions (9), Eq. (11) can be transformed into

$$\omega_H - \Omega_H = \nu\dot\Theta - \nu|\Omega_0|. \quad (12)$$

Taking into account the interaction delay in the lower order approximation leads to the correction to the bunch potential

$$\Phi^{(1)}(x_1, y_1) = \frac{\dot\Theta^2\rho_0}{2c^2}\int\frac{d\mathbf{r}'}{r'^3}(x'^2 - y'^2)$$

$$\times[(x_1^2 - y_1^2)\cos2\Theta + 2x_1y_1\sin2\Theta]$$

$$= \frac{\dot\Theta^2\rho_0}{2c^2}\pi R_xR_y\frac{R_x - R_y}{R_x + R_y} \quad (13)$$

$$\times[(x_1^2 - y_1^2)\cos2\Theta + 2x_1y_1\sin2\Theta].$$

3. Let us analyze Eqs. (8) and (12) taking into account the relationship

$$2K_0\left(\dot\Theta + \frac{\omega_H}{2}\right) - \frac{\Delta}{2} \simeq 0, \quad (14)$$

which follows from the condition

$$\delta_0 \simeq 0 \quad \left(\dot\Theta + \frac{\omega_H}{2} > 0\right).$$

From Eq. (8), it follows that

$$\dot\Theta + \frac{\omega_H}{2} = \frac{\omega_0^2}{\omega_H}$$

and

$$\Omega_0 = \left|\frac{\omega_0^2}{\omega_H} - \frac{\omega_H}{2}\right|.$$

When $\omega_H^2 > 2\omega_0^2$, Eq. (12) yields

$$\Omega_H = \omega_H(1+\nu) - 2\nu\frac{\omega_0^2}{\omega_H}. \quad (15)$$

The external field is attenuated by rotating particles so that $\Omega_H > \omega_H$. For $\omega_H^2 < 2\omega_0^2$, the sets of Eqs. (8) and (12) have no real solutions. When the delay of the electric field is taken into account, Eqs. (9), (10), and (12) and the equality

$$\mathbf{E}^{(1)} = -\frac{1}{c}\frac{\partial\mathbf{A}}{\partial t} - \nabla\Phi^{(1)}$$

yield the following expressions for ω_1^2 and ω_2^2:

$$\omega_1^2 = \frac{\nu}{R_x+R_y}\left(\frac{c^2}{R_x} - \dot\Theta\left(R_x\frac{C_2}{B} + R_y\frac{C_1}{A}\right)\right), \quad (16)$$

$$\omega_2^2 = \frac{\nu}{R_x + R_y}\left(\frac{c^2}{R_y} + \dot{\Theta}\left(R_x\frac{C_2}{B} + R_y\frac{C_1}{A}\right)\right). \quad (17)$$

Accordingly, instead of Eq. (8), we obtain

$$\frac{R_x}{R_y} = \frac{\omega_2^2 + \nu\dot{\Theta}\left(\dot{\Theta} + \frac{\omega_H}{2} - K_0\right)}{\omega_1^2 + \nu\dot{\Theta}\left(\dot{\Theta} + \frac{\omega_H}{2} + K_0\right)} = \frac{\sqrt{\dot{\Theta} + \frac{\omega_H}{2} - K_0}}{\sqrt{\dot{\Theta} + \frac{\omega_H}{2} + K_0}}. \quad (18)$$

It follows from here that

$$2\left(\dot{\Theta} + \frac{\omega_H}{2}\right)\left(\dot{\Theta} + \frac{\omega_H}{2} + |\Omega_0|\right) = \omega_0^2 - \nu\dot{\Theta}|\Omega_0|. \quad (19)$$

The external field can be found from the following set:

$$(4+\nu)\dot{\Theta}^2 + \frac{2\dot{\Theta}}{\nu}[\omega_H(\nu-1) + \Omega_H(\nu+2)] + \frac{\omega_H}{\nu}\left(\omega_H\frac{\nu-2}{2} + \Omega_H\right) - \omega_0^2 = 0, \quad (20)$$

$$\dot{\Theta} = \frac{\omega_0^2 - \frac{\omega_H^2}{2} + \left(\frac{\omega_H - \Omega_H}{\nu}\right)^2}{\omega_H + 2\frac{\omega_H - \Omega_H}{2}}. \quad (21)$$

Assuming that the term $\nu\dot{\Theta}|\Omega_0|$ in Eqs. (18) and (19) is small, for the external field we obtain

$$\Omega_H \simeq \omega_H + \nu\left(\omega_H - 2\frac{\omega_0^2}{\omega_H}\right) + \frac{4\nu^2\left(\frac{\omega_H}{2} - \frac{\omega_0^2}{\omega_H}\right)^3}{\omega_H^2}, \quad (22)$$

where the term proportional to ν_2 is always positive; i.e., when the delay in the potential is taken into account, the attenuation of the external field is stronger than in the case when the delay is ignored.

4. The particle velocities in the rest frame are

$$\dot{x}_1 = (\dot{x} - y\dot{\Theta})\cos\Theta - (\dot{y} + x\dot{\Theta})\sin\Theta,$$

$$\dot{y}_1 = (\dot{x} - y\dot{\Theta})\sin\Theta + (\dot{y} + x\dot{\Theta})\cos\Theta.$$

Since the average quantities satisfy the relationships

$$\bar{\dot{x}} = -\frac{C_1}{A}y, \quad \bar{\dot{y}} = -\frac{C_2}{B}x,$$

the maximum average particle velocities in the rest frame are

$$(\bar{\dot{x}}_1)_{\max} = \left(\frac{\omega_H}{2} - K_0\right)R_y, \quad (\bar{\dot{y}}_1)_{\max} = \left(\frac{\omega_H}{2} + K_0\right)R_x,$$

where

$$R_y = \frac{\sqrt{\nu}c\omega_1}{\omega_2\omega_0\sqrt{2}}, \quad R_x = \frac{\sqrt{\nu}c\omega_2}{\omega_1\omega_0\sqrt{2}}.$$

When $\omega_1 \gg \omega_2$ and $\omega_0^2 \simeq \omega_H^2/2$, the inequalities $(\bar{\dot{x}}_1)_{\max} \ll c$ and $(\bar{\dot{y}}_1)_{\max} \ll c$ are valid for $\nu \gtrsim 1$.

The mean-square velocities are on the order of v_0 (note that $v_0 \ll c$ at any ν). This means that dense bunches can be confined by the external field. The particles in a long bunch expand rather slowly along the field lines.

Thus, in this paper, the states of a long charged high-density bunch rotating in a magnetic field have been studied and the effect of the interaction delay on these states has been estimated.

REFERENCES

1. A. S. Chikhachev, Zh. Tekh. Fiz. **55**, 1179 (1985) [Sov. Phys. Tech. Phys. **30**, 673 (1985)].
2. A. S. Chikhachev, Radiotekh. Élektron. (Moscow) **39**, 453 (1994).
3. R. C. Davidson, *Theory of Nonneutral Plasmas* (Benjamin, New York, 1974; Mir, Moscow, 1978).

Translated by I. Efimova

Technical Physics, Vol. 46, No. 2, 2001, pp. 227–233. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 98–104.

Electron Beam Generation in a Magnetron Diode with Metal Secondary-Emission Cathode

Yu. Ya. Volkolupov, A. N. Dovbnya, V. V. Zakutin, M. A. Krasnogolovets, N. G. Reshetnyak, V. V. Mitrochenko, V. P. Romas'ko, and G. I. Churyumov

National Research Center, Kharkov Institute of Physics and Technology, Kharkov, 61108 Ukraine

Received August 24, 1999; in final form, February 15, 2000

Abstract—The effects of electric and magnetic field intensities, the triggering-pulse droop rate, and the electrode diameter on the processes of electron beam formation and generation were studied experimentally. The results of mathematical simulation of the secondary-emission multiplication of the electron flow are presented. Tubular electron beams with a wall thickness of 1.5–2 mm, a current density of 1–70 A/cm^2, and a particle energy of 5–100 keV were obtained. It was shown that several electron bunches could be obtained during a single voltage pulse. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The pulse and mean power of microwave sources, as well as their service life, depend on the type of the cathode [1–4]. It is well known that magnetron diodes based on cold secondary-emission metal cathodes have a prolonged service life (~100 000 h [2]) and provide high density electron emissions (~50 A/cm^2 [4]). These diodes have a rather simple design and can be effectively used as microwave sources. However, certain effects relevant to beam generation in magnetron diodes are insufficiently understood, both theoretically and experimentally. In particular, the process of space charge formation in the anode–cathode gap in crossed electric and magnetic fields is virtually obscure [5]. To solve this problem, physical processes in the magnetron diode in the magnetic field exceeding the Hull cutoff field ($H_{\text{Hull}} = 6.72(U)^{1/2}[r_a(1 - r_k^2/r_a^2)]^{-1}$, where H is the longitudinal magnetic field (Oe); U is the voltage across the diode (V); and r_c and r_a are the cathode and anode radii (cm), respectively) should be studied both theoretically and experimentally.

The goal of this work was to study the processes of electron layer formation near the secondary-emission metal cathode and electron beam generation in a magnetron diode.

EXPERIMENTAL SETUP AND METHODS

Parameters of the electron beams generated by magnetron diodes were studied using the experimental setup shown in Fig. 1. The setup consisted of a high-voltage pulse modulator (*1*) (voltage amplitude, 5–200 kV; pulse duration, 2–10 μs; triggering pulse amplitude, up to 15 kV; triggering pulse duration, ~70 ns); a high-voltage generator (*2*) for inducing secondary emission (amplitude, up to 3.5 kV; duration, 1–10 ns); a focusing solenoid (*4*) (magnetic field intensity, up to 3500 Oe; longitudinal nonuniformity, ~8%); a vacuum chamber (*3*) containing a secondary-emission diode with a central copper cathode (*5*) and a tubular stainless-steel anode (*6*) (electrode length, 100–140 mm; pressure in the chamber is maintained at a level of $\leq 10^{-6}$ torr using an ion pump); an indication system consisting of current and voltage detectors and a Faraday cup (*7*) with a calorimetric power meter; and a synchronization system. The electron energy was determined by measuring the electron absorption by aluminum foil. The beam spot size was measured using the beam spot image on X-ray film and molybdenum foil. The electron beam produced by the magnetron diode powered by a pulse modulator at a pulse repetition rate of 10–50 Hz was studied.

Secondary emission was induced both by an external voltage pulse applied to the anode from separate pulse sources (U = 1–15 kV, t_{droop} ~ 2–100 ns) and by a voltage surge droop in the initial part of the cathode voltage pulse (self-triggering). The amplitude of this

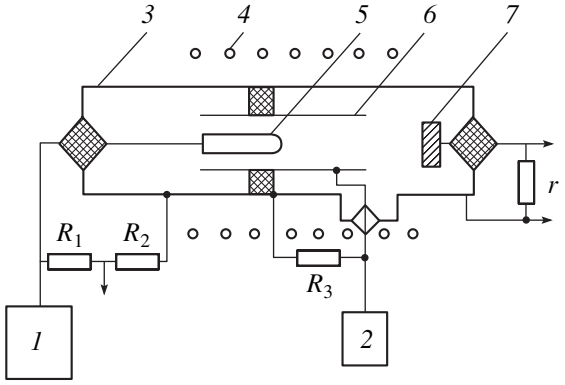

Fig. 1. Diagram of the experimental setup.

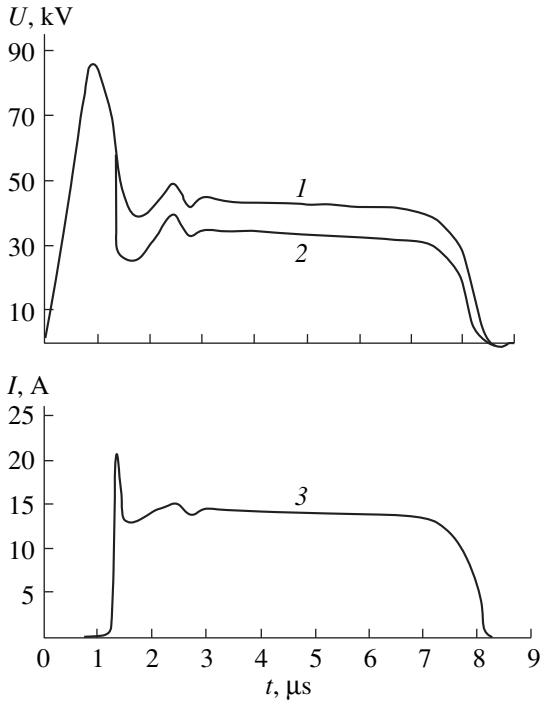

Fig. 2. Oscillograms of the cathode voltage pulse: (*1*) modulator idling; (*2*) beam generation mode; and (*3*) Faraday cup beam current mode.

surge varied from 10 to 100 kV; the droop duration was about 1 μs (Fig. 2).

EXPERIMENTAL RESULTS AND DISCUSSION

1. Electron beam formation. The processes of electron beam formation were studied using magnetron diodes with different values of the ratio r_a/r_c (1.2–15). Typical oscillograms of the cathode voltage pulse and Faraday cup beam current obtained using a magnetron gun (r_a = 1.3 cm; r_c = 0.25 cm) are shown in Fig. 2. Secondary emission was induced by the cathode voltage pulse surge droop. As seen from the oscillograms, the electron layer is formed and the beam is generated upon the surge droop. The process of beam current generation during the voltage droop occurs in two stages: a relatively slow voltage droop (hundreds of nanoseconds) and a fast voltage droop (nanoseconds or tens of nanoseconds). During the first stage (curve *1*), autoelectronic emission and accumulation of autoelectrons in the anode–cathode gap occur (idle run of modulator). During the second stage (curve *2*), the electron layer is formed and the beam is generated. It should be noted that duration of the slow droop stage depends on the electric intensity *E*. A decrease in the electric intensity from 120 to 40 kV/cm caused an increase in the slow droop stage duration from 250 to 600 ns.

Autoelectrons are emitted during the slow droop stage. The energy of the electrons moving in a decreasing electric field and a static magnetic field along cycloidal trajectrories increases. This causes electron bombardment of the cathode. As the electric field near the cathode decreases, at a certain moment the electron energy becomes sufficient for removing secondary electrons with a secondary-emission coefficient σ greater than one. Secondary electrons can be removed both by electrons with the energy of 0.3–1 keV incident at a right angle and by electrons with lower energies incident at smaller angles. Once the electrons have been emitted, the stage of fast voltage droop begins. This stage is characterized by an avalanche-like increase in the number of bombarding and knocked-out electrons. A sharp increase in the number of electrons causes a decrease in the voltage level, thereby enhancing the voltage droop rate. As a consequence, the energy of electrons bombarding the cathode increases and reaches a point where $\sigma > 1$. This causes a further increase in the number of electrons near the cathode until a dynamic equilibrium is attained, and the electron layer near the cathode is formed. On attaining equilibrium, the stage of steady-state secondary-electron multiplication begins and the beam is generated. It was revealed experimentally that the beam was generated when the electron-drift velocity $v = E/H$ was equal to 0.1–0.2 s (depending on experimental conditions).

It was found experimentally that the fast droop stage duration and, therefore, the beam current buildup time depended on the anode and cathode diameters and the voltage droop rate and duration. For example, if the cathode diameter exceeded 0.2 cm, the anode diameter exceeded 2.2 cm, the droop duration was 0.1–0.5 μs, and the droop rate exceeded 20 kV/μs, the beam current pulse buildup time exceeded 10 ns. Experiments with magnetron diodes with cathode and anode diameters of 0.2 and 1.0 cm, respectively, showed that the beam current was generated within a shorter time interval (1–10 ns). The duration of the time interval was equal to the droop duration. The droop rate was 1200–300 kV/μs. At such a high droop rate, the number of primary autoelectrons is rather small. However, because of a high pulse droop rate, the energy of these autoelectrons acquired in 10–20 gyroperiods reaches a value at which $\sigma > 1$.

Collective motion of electrons during their multiplication can cause space charge oscillations. This becomes possible at the steady-state stage, when the space charge reaches a certain minimum density and the electron layer begins to shield the cathode, which results in potential sagging in the anode–cathode gap and a decrease in the electric field strength near the cathode. The energy of electrons is changed as a result of their interaction with the fields induced by space-charge oscillations. The noise accompanying this interaction can be observed in the beam current pulse. It was found experimentally that the noise was stronger when a large-diameter cathode (80 mm) was used; i.e., the noise increased with decreasing electric intensity. The noise amplitude reached 20% of the beam current

amplitude. The noise wave period was equal to several nanoseconds.

2. The dependence of the beam generation current on the diode dimensions. The dependence of electron beam parameters on the diode dimensions was studied. The cathode diameter d ranged from 2 to 80 mm; and the anode diameter D, from 10 to 140 mm. The values of the Faraday cup beam current I, cathode voltage U, and the magnetic field intensity H for various anode and cathode diameters are given in the table for eight modifications of the magnetron diode. The electric field strength near the cathode varied from 20 to 125 kV/cm (at a constant voltage pulse amplitude). It was found that the beam current obeyed the Langmuir–Child law. At a given voltage, the beam current amplitude could be maximized by tuning the magnetic field strength. Electron beams with a current ranging from 1 to 50 A and a particle energy ranging from 5 to 100 keV were obtained.

At the initial stages of beam generation, short (~1 μs) surges were observed on the beam current pulse plateau. These surges were caused by gas desorption from the cathode surface and gas ionization [4]. After the diode was aged, the beam current pulse amplitude and shape remained invariant.

The dependence of the beam current on the cathode diameter is shown in Fig. 3. These curves were obtained at a voltage of 24 kV and anode–cathode distance of 5 (curve *1*) and 20 mm (curve *2*). It can be concluded from these curves that the beam current is inversely proportional to the logarithm of the ratio between the anode and cathode diameters. This result is in good agreement with the dependence obtained for classical magnetrons [6]. As seen from Fig. 3, the beam current decreases with increasing anode–cathode distance. However, one of the advantages of the diode is that it operates at rather low magnetic intensities (≤1000 Oe). In this case, the magnetic coil heating is insignificant, so that it is not necessary to adjust the magnetic field. In addition, this allows a magnetron diode with specified parameters (current, magnetic field intensity, and dimensions) to be designed.

3. Effect of electric and magnetic fields on electron beam formation. The effect of the magnetic field strength on beam generation at a constant cathode voltage amplitude was studied. There was a sharp rise, a plateau, and a sharp decrease in the Faraday cup beam current amplitude with increasing magnetic field. Such behavior was caused by changes in the electron trajectories and processes of energy accumulation by electrons in the anode–cathode gap in an increasing magnetic field. If the anode–cathode spacing was significantly larger, the dependence became smoother. This allowed beam current tuning over a wide range. The experimental data for the magnetron diode (anode diameter, 50 mm; voltage, 60 kV) are given in the table. It follows that the current can be varied from 0.5 to 10 A by changing the magnetic field strength from 1100 to 2000 Oe.

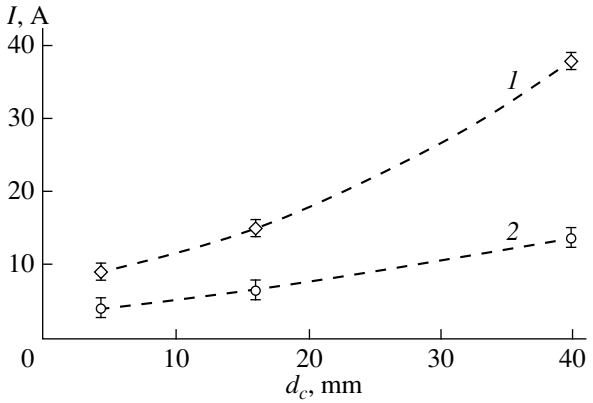

Fig. 3. Dependence of the beam current on the cathode diameter.

The diode working range dependence on the cathode voltage amplitude at a given magnetic intensity was studied. It was found that an approximately 20% variation in the voltage amplitude had virtually no effect on the process of electron beam formation in a static magnetic field. As the cathode voltage approached the ΔU limit from above or below, the conditions for beam generation were violated, and the voltage pulse instability caused secondary emission breakdown (Fig. 4).

The use of this effect in experiments with a magnetron diode with anode and cathode diameters of 78 and 40 mm, respectively, allowed the beam current to be completely modulated in amplitude at a carrier frequency of 1 MHz (Fig. 5). For this purpose, sinusoidal modulation of the voltage pulse peak was performed, and electron bunches were obtained at the diode output on the dips of the sinusoid. Complete modulation of current at a carrier frequency of ≥1 MHz was also attained in experiments with a magnetron diode of the same dimensions by tuning the magnetic field distribution over the diode axis. Multispiking generation of the electron beam could be attained by varying the amplitude and longitudinal distribution of the magnetic field.

Table

d, mm	D, mm	U, kV	I, A	H, Oe
2	10	7	1.6	2100
5	26	32	14	1900
5	50	60	1	1400
5	50	60	10	2000
16	50	17	5	600
40	50	30	50	2200
40	78	100	50	1800
80	100	19	8	1100

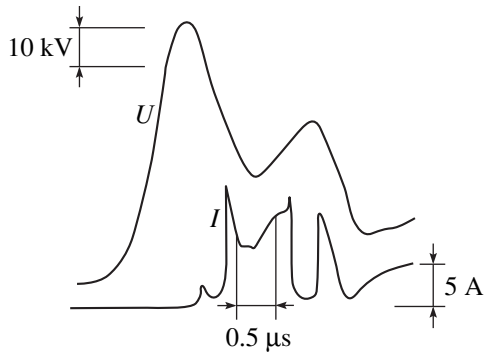

Fig. 4. Oscillograms of the cathode voltage *U* and beam current *I*; the cathode diameter is 5 mm, the anode diameter is 26 mm, and *H* = 1800 Oe.

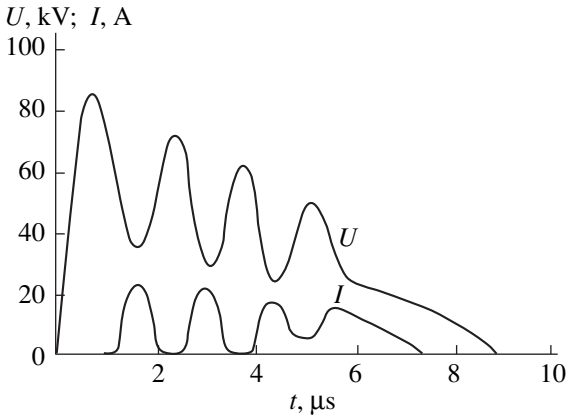

Fig. 5. Oscillograms of the beam current *I* and cathode voltage *U*; *H* = 700 Oe.

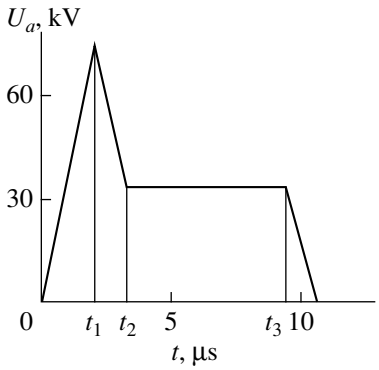

Fig. 6. Shape of the anode voltage pulse.

For example, an electron beam pulse with an amplitude of ~15 A and duration of 8 μs was obtained at a cathode voltage of 55 kV and a magnetic intensity of ~1150 Oe (cathode diameter, 40 mm; anode diameter, 78 mm). As the magnetic intensity decreased to 700 Oe, the beam current pulse assumed a spiky shape with a generation period of ~1 μs and a current amplitude of 30 A (duration of each spike was 10–30 ns).

The beam spot size on the collector was measured. It was determined that the beam cross section consisted of concentric rings with a uniform azimuthal intensity distribution. The beam diameter was approximately equal to the cathode diameter, and the beam wall thickness was 1–2 mm. Therefore, the energy of electrons in the electron layer was about 0.5–1 keV, which approximately corresponded to the maximum secondary emission coefficient for copper. Experiments with a magnetron diode with a cathode 80 mm in diameter showed that, if the cathode was exposed to a nonuniform magnetic field with a transverse component (about 5%), the beam spot image was partly sharp and partly fuzzy. The thicknesses of the sharp and fuzzy parts were about 2 and 3–4 mm, respectively. This fuzziness was caused by the nonuniformity of the magnetic field.

THEORETICAL STUDY OF THE ELECTRON FLOW MULTIPLICATION

Theoretical simulation of the secondary emission development and the steady-state stage of the secondary-emission cathode operation was performed to study specific features of the processes occurring in a magnetron diode. The steady-state stage was analyzed using a three-dimensional mathematical model of a magnetron diode. The model was based on a self-consistent set of simultaneous differential equations of motion (for electron flow) and the Poisson equation (for calculating the space-charge forces).

The magnetron diode simulated by this model was assumed to have the following parameters: cathode diameter, 5 mm; anode diameter, 26 mm; peak anode voltage ($0 < t < t_1$), 75 kV; anode voltage plateau ($t_2 < t < t_3$), 35 kV; voltage droop duration ($t_1 < t < t_2$), 1.25 μs; voltage plateau duration, 60 μs; magnetic intensity, 2000 Oe; and cathode length, 90 mm. A linear approximation of the shape of the voltage pulse applied to the anode is shown in Fig. 6. The value of the anode voltage within the interval $t_2 < t < t_3$ was taken as the working value. The large-particle method described in [7] was used to study the dynamics of the electron flow formation processes.

Consider specific features of electron processes in the magnetron diode. These processes are nonsteady (secondary-emission multiplication occurs at the anode voltage pulse surge droop ($t_1 \le t \le t_2$)). On the other hand, the spatial distribution of the electron flow should be studied using a three-dimensional cylindrical coordinate system (r, φ, z). The mathematical model of the magnetron diode is based on the self-consistent set of simultaneous equations of motion for the electron flow

$$\frac{\partial v_r}{\partial t} = \eta E_0^r(t) + r v_\varphi^2 - \omega_c r v_\varphi,$$

$$\frac{\partial(r\Theta)}{\partial t} = \eta\frac{1}{r^2}E_0^{\varphi}(t) - 2\frac{v_r v_\varphi}{r} + \omega_c\frac{v_r}{r},$$
$$\frac{\partial v_z}{\partial t} = \eta E_0^z(t), \quad \frac{\partial r}{\partial t} = v_r, \tag{1}$$
$$\frac{\partial\varphi}{\partial t} = \Theta, \quad \frac{\partial z}{\partial t} = v_z,$$

where $v_{r,\varphi,z}$ are the electron velocity components, $\omega_c = \eta B_0$ is the cyclotron frequency, and $\eta = e/m$ is the reduced electron charge; and on the Poisson equation (for calculating the space-charge field)

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial U}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 U}{\partial\varphi^2} + \frac{\partial^2 U}{\partial z^2} = -\frac{\rho}{\varepsilon_0}, \tag{2}$$

where $\rho = \rho(r, \varphi, z)$ is the space charge (SC) density.

The set of simultaneous equations (1) and (2) was solved in the quasi-stationary approximation for a constant anode voltage within the limits of the motion equation integration step (beginning at $t = t_1$). The step was taken to be $\Delta T = (1/10)T_c$, where $T_c = 2\pi/\omega_c$ is the cyclotron oscillation period. The equation of motion was solved numerically using the Runge–Kutta method of the fourth order. The quiet start model [8] was used as the model for determining the initial coordinates and velocities of the particles in the magnetron diode interaction space.

The Poisson equation (2) was solved by the Hockney method of finite differences using fast Fourier transforms [9, 10] at given initial and boundary conditions. Numerical differentiation of the SC potential at the nodes of the finite-difference mesh was used for determining the electrostatic field intensity $\mathbf{E}_0 = -\text{grad}U$. Discrete values of the SC potential (least-squares method) were locally smoothed to reduce fluctuations of the calculated SC field.

The results of the simulation are presented in Figs. 7 and 8. The energy distribution of the primary electrons (large particles (macroparticles) with a charge of $q = 0.8 \times 10^{-13}$ C) and the theoretical approximation of the experimental dependence of the secondary-emission coefficient for copper on the primary electron energy [11] are shown in Fig. 7. As seen from Fig. 7, the number of low-energy macroparticles (i.e., particles with energy corresponding to $\sigma \le 1$) is more than 60% of the total number of macroparticles bombarding the cathode. The presence of low-energy particles affects the SC field distribution near the cathode surface (within a distance equal to the selected mesh size $\Delta r = (r_a - r_c)/32$). Changes in the number of low-energy particles affect the cathode current (limitation of emission by the SC field). The number of macroparticles with energies higher than the first critical potential (i.e., the energy for which $\sigma > 1$) depends on the voltage applied. The macroparticle energy increases with increasing voltage. It should be noted that the critical potential value used

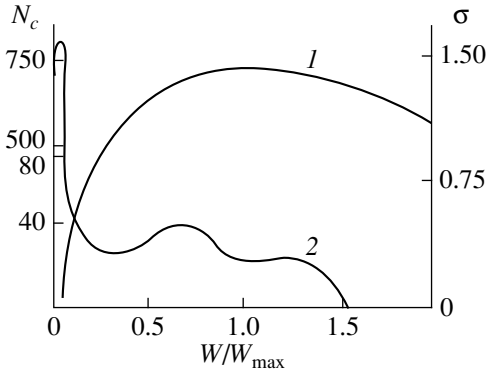

Fig. 7. (*1*) Dependence of the secondary-emission coefficient σ on the particle energy and (*2*) dependence of the number of primary electrons N_c on the particle energy ($U_a/U_{cr} = 0.7$).

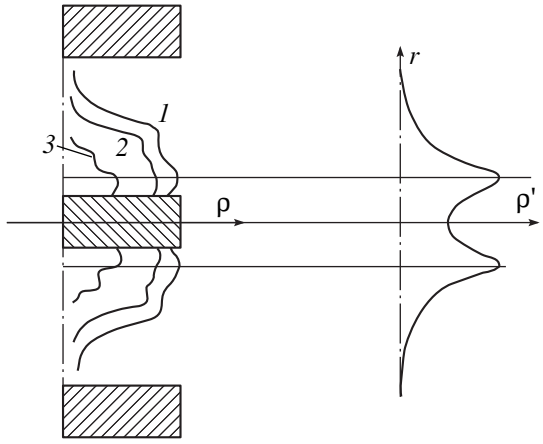

Fig. 8. U_a/U_{cr} = (*1*) 0.7; (*2*) 0.5; and (*3*) 0.3.

for voltage normalization was determined from the Hull cutoff field. For the magnetron diode under study, the critical potential was found to be 139 kV.

The theoretical radial distribution of the SC density in the diode interaction space and the experimental radial distribution of the beam intensity at the collector surface are shown in Fig. 8. The maximum SC densities in the three cases under consideration were calculated: in the first case, $\rho/\rho_{br} = 0.19$; in the second case, $\rho/\rho_{br} = 0.16$; in the third case, $\rho/\rho_{br} = 0.1$, where ρ_{br} is the Brillouin density of the space charge. In the third case ($U_a/U_{cr} \approx 0.3$), the theoretical and experimental distributions of SC density are in satisfactory agreement. The presence of electrons in the paraxial region is caused by cycloidal motion of electrons.

Numerical simulation of secondary-electron emission from the cathode surface at the steady-state voltage stage ($t_2 \le t \le t_3$) was performed. The secondary-electron emission was caused by bombardment of the cathode with primary electrons with energies of 300, 560, and 700 eV. The plateau voltage was equal to $U =$

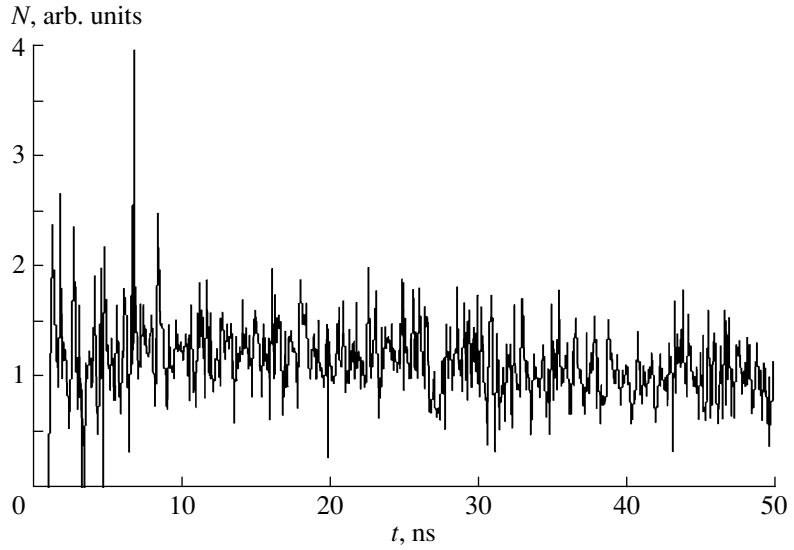

Fig. 9. Time dependence of the number of electrons in the electron layer (E = 300 eV).

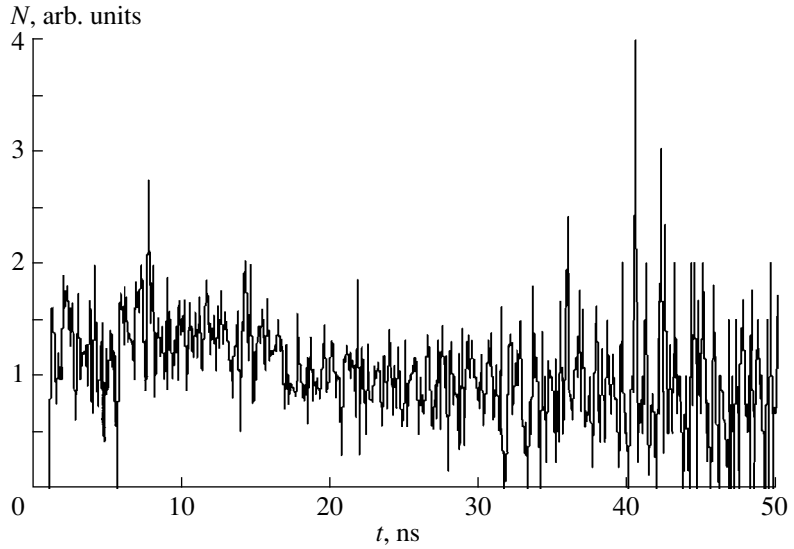

Fig. 10. Time dependence of the number of electrons in the electron layer (E = 700 eV).

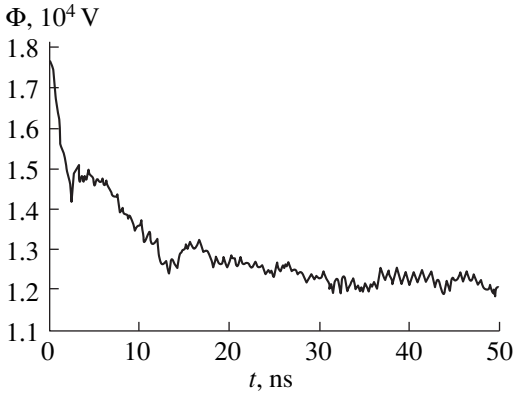

Fig. 11. Time dependence of the potential in the middle of the anode–cathode gap.

40 kV; the magnetic field intensity, H = 2000 Oe. A hot cathode was the source of primary electrons. When the hot cathode had worked for several nanoseconds, it was switched off. The initiated process of electron multiplication was studied. When the energy of the primary electrons reaches 700 eV, irregular disturbances of the electron density of the space charge occur with a delay of >30 ns (Fig. 10). If the energy of the primary electrons is about 300 eV, these disturbances are insignificant (Fig. 9). These results are in qualitative agreement with the results presented in [5]. As seen from the obtained results, the secondary-emission coefficient of the cathode material should be maximal at an incident electron energy of ~500 eV. The time dependence of the potential in the middle of the anode–cathode gap is

given in Fig. 11. As seen in Fig. 11, there is a decrease in the potential caused by thermoelectrons in the beginning of the curve. When thermionic emission ceases, the accumulated electrons disperse, causing a slight increase in the potential. Then, there is a decrease in the potential (the duration of the decrease is 10–15 ns) caused by the completion of transient processes and formation of the space charge maintained by the secondary emission.

The results of calculations are in satisfactory agreement with the results of experiments on electron beam generation using magnetron diodes.

CONCLUSION

It was shown that magnetron diodes with metal secondary-emission cathodes can be used to generate straight electron beams with high current densities. Tubular electron beams with a current density of up to 50–70 A/cm^2, the outer diameter of 3.5–84 mm, a wall thickness of 1.5–2 mm, and a particle energy of 5–60 keV were obtained (pulse duration was 10 μs). It was revealed experimentally that a beam current pulse train could be generated using a single voltage pulse. Magnetic field variation was shown to allow 10- to 20-fold tuning of the current. The processes of electron multiplication and the steady-state stage of secondary emission were studied theoretically. It was shown that magnetron diodes could be used as electron sources for high-power microwave devices and charged-particle accelerators. They also can be used as fast high-voltage switches with characteristic commutation times of several nanoseconds.

REFERENCES

1. V. M. Lomakin and L. V. Panchenko, Élektron. Tekh., Ser. 1, No. 2, 33 (1970).
2. J. F. Skowron, Proc. IEEE **61** (3), 69 (1973).
3. S. A. Cherenshchikov, Élektron. Tekh., Ser. 1, No. 6, 20 (1973).
4. A. N. Dovbnya, V. V. Zakutin, *et al.*, in *Proceedings of the 5th European Particle Accelerators Conference, Sitges, 1996*, Ed. by S. Myers, A. Pacheco, R. Pascual, *et al.* (Inst. of Physics Publ., Bristol, 1996), Vol. 2, p. 1508.
5. A. V. Agafonov, V. P. Tarakanov, and V. M. Fedorov, Vopr. At. Nauki Tekh., Ser. Yad.-Fiz. Issled., Nos. 2–3, 137 (1997).
6. G. G. Sominskiĭ, D. K. Terekhin, and S. A. Fridrikhov, Zh. Tekh. Fiz. **34** (9), 1666 (1964) [Sov. Phys. Tech. Phys. **9**, 1286 (1965)].
7. R. Hockney and J. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1984; Mir, Moscow, 1987).
8. A. S. Roshal', *Modeling of Charged Beams* (Atomizdat, Moscow, 1979).
9. R. W. Hockney, J. Assoc. Comput. Mach. **12** (1), 95 (1965).
10. J. W. Cooly and J. W. Takey, Math. Comput., No. 19, 161 (1965).
11. I. M. Bronshteĭn and B. S. Fraĭman, *Secondary Electron Emission* (Nauka, Moscow, 1969).

Translated by K. Chamorovskiĭ

Technical Physics, Vol. 46, No. 2, 2001, pp. 234–239. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 105–110.

Processes in Open Systems on Crystal Surfaces with Low Miller Indices

V. A. Voĭtenko

St. Petersburg State Technical University, ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia

Received March 31, 2000

Abstract—Exposure characteristics that were obtained when growing various films on natural low Miller index surfaces of several crystals were collected and analyzed. An evolution theory that explains their special form was constructed. The type of dose characteristics obtained suggests that the surface underwent a reconstruction, i.e., a nonequilibrium phase transition that occurs on the surface. A quantitative analysis of the experiments available has been performed. In particular, a quantitative estimation was obtained of to what extent coverage with lead hinders the oxidation of the surface of a nickel crystal. Upon intense light or electron irradiation of silicon, divacancies are the predominant centers of the formation of point and extended radiation defects, as well as of local regions of melting. For some two-dimensional systems (divacancies, sulfur atoms on the surfaces of passivated semiconductors, oxide films), delay times and evolution times for the self-organized structure formed were determined. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The processes of film nucleation in an open growth system on a crystal surface are characterized by periodic fluctuations of the adatom concentration. Various ways of the development of two-dimensional nuclei of a new phase, e.g., by the mechanism of Volmer–Weber or Stranski–Krastanov, are possible, but the predominant factor always is the appearance of large-scale periodic fluctuations [1].

In recent works [2, 3], magnetic films were applied (using a shrouded sublimation source cooled by liquid nitrogen) on the surface of semiconductors such as Ge and GaAs subjected to sulfide passivation. Analogous evaporation of oxygen onto the surface of pure and lead-covered nickel was performed in [4]. Passivation of the silver surface with chlorine upon the evaporation in an ultrahigh vacuum was studied in [5]. At present, automated methods of controlling growth surfaces are widely developed. One of such methods is based on the measurement of the temporal dependence of the intensity of electron-diffraction patterns [6] (which is studied in our work as well). These investigations confirmed the result known from biology [7] and economics [8] according to which the properties of open systems are connected with the character of their microscopic interactions more closely than in the equilibrium case. Nevertheless, one can usually reveal the general empirical regularities. This work is devoted to their explanation for the case of thin films. In particular, the role of evaporation temperature upon film preparation is discussed. On the one hand, with increasing evaporation temperature (e.g., upon deposition of magnetic films or upon passivation or oxidation), a significant interdiffusion of the main contacting components occurs. In this case, the magnetic moment of the iron film vanishes and the superlattice becomes destroyed. On the other hand, with decreasing evaporation temperature, the film grown (e.g., a photocathode) proves to be very coarse and highly defective because of the decreased diffusion along the surface of contact. The magnetic structure of a film or the crystal structure of individual layers of a photocathode without deteriorating the sharpness and flatness of the layer boundaries can be preserved by using buffer layers.

The maintenance of stable growth of the energetically unfavorable free surface of iron upon gas-phase deposition of fcc iron films on a sulfided (100) surface of semiconductors [2, 3] is ascribed to the interaction of sulfur atoms with the growing surface. The decrease in the growth entropy appears to be due to the floating out of a sulfur layer on top of the surface. The processes that occur on the surface are controlled by Auger electron spectroscopy (AES) [2–4] and low-energy electron diffraction (LEED) [2, 3]. In spite of the great importance of the problem from the viewpoint of practical application and the large number of theoretical models [9, 10], the evolution processes that determine the adhesion of thin films with low-index surfaces (surfaces with low Miller indices) have yet been clarified at present insufficiently.

On a microscopic level, the guarantee of the formation of an ideal film on a semiconductor surface appears to consist in that the applied atoms have no electron orbitals that could produce energy levels in the forbidden zone of the main crystal. In this case, no charge transfer occurs onto the bonding orbitals, and the adatoms can freely move along the surface or a boundary. It is exactly upon such movement that the above-mentioned coarse-scale fluctuations arise that serve as a basis for the formation of clusters of a new phase, just

1063-7842/01/4602-0234$21.00

as upon the diffusional decomposition of solid solutions [1]. On the whole, the main factor that determines the structure, morphology, and sharpness of the growing film is the diffusion of the lighter atoms (sulfur, oxygen) along the boundary between the heavier substances (semiconductor and iron, nickel, and lead). For example, the probability of the formation of sulfide bonds on a semiconductor surface upon sulfiding is determined by the diffusion of oxygen vacancies in the oxide film [11]. The three-dimensional diffusion theory of growth that was applied in [4] to the case of oxidation of pure nickel failed in explaining the first plateau in the exposure dependence of the intensity of the oxygen Auger peak. The exposure curve of oxidation of nickel with a lead coating [4] cannot be interpolated at all; no model can be associated with the experimental points. The authors of [2, 3] also give no approximation for the exposure dependence that they obtained for the intensity of the LEED pattern. All this was done in the present paper. We show that a satisfactory agreement between the theory and experiment can be obtained assuming that the evolution processes on the surface begin from the coalescence of a pair of diffusing adatoms. The grid of nanoclusters that arises in the process of deposition on the surface is determined by the number of initial protopairs. Formally, the situation is as if, during the entire process of condensation of adatoms, the critical nuclei contain only one pair of adatoms.

STRANSKI–KRASTANOV THEORY OF EVOLUTION PROCESSES

The adhesion of iron films to sulfided (100) surfaces of semiconductors is stabilized owing to the continuous segregation of sulfur through an ordered $c(2 \times 2)$ iron layer. The stabilization of the process, as follows from the similarity of dose characteristics, occurs also upon oxidation of the (100) nickel surface; here, the movement normal to the surface occurs inward. In both cases, the key event is the formation of molecules such as FeS_2 and NiO_2, e.g.,

$$Fe + 2S = FeS_2. \quad (1)$$

Suppose that only the lightest adatoms (sulfur, oxygen) participate not only in the transverse motion but also in the movement along the interface, whereas the migration of heavy atoms (iron, lead) is absent. Upon the deposition of an iron film, diffusional motion in the preliminarily adsorbed layer of sulfur atoms is initiated and is accompanied by the floating of sulfur on top of the free surface in the perpendicular direction. The diffusion occurs according to the equation

$$\frac{\partial n}{\partial t} + \mathrm{div}(nb\mathbf{F} - D\nabla n) = G(n) - \frac{n}{\tau}. \quad (2)$$

Here, n is the concentration of adatoms per unit surface area; D is the coefficient of atomic diffusion parallel to the surface; τ is the lifetime of atoms on the surface with respect to desorption; $G(n)$ is the rate of generation of adatoms on the surface; $\mathbf{F}$ is the elastic force [12] due to the lattice-parameter dependence on the film composition; and b is the mobility. The quantities $G(n)$ and $\mathbf{F}$ in Eq. (2) are considered to be functions of the concentration n. According to reaction (1), the formation of one sulfide bond that favors the adhesion of the iron film (coating) to the sulfided semiconductor substrate requires the simultaneous localization of two sulfur atoms at one immobile iron atom. The nickel oxide has a variable composition NiO_x with $x = 0.98$–1.7 and therefore [13] can be regarded as a solid solution of two compounds, namely, NiO and NiO_2. Thus, the probability of the independent localization of a pair of oxygen atoms near one nickel atom is important for the formation of the oxide film as well.

The homogeneous lattice chemical reactions, such as the formation of complex secondary radiation defects [14, 15], or oxidation or synthesis of molecules (e.g., by Eq. (1)), satisfy the balance equation

$$G(n) = \frac{n}{\tau}. \quad (3)$$

In the two-dimensional case [11, 15–17], Eq. (3) can lead to a bistability of the atomic coating, i.e., to its tendency to be in one of two states with different steady-state concentrations n. For this to occur, it is necessary that the rate of generation $G(n)$ of primary quasiparticles be a nonlinear function of the concentration n. Among the variety of open systems, objects of this type occur quite frequently. Thus, the transition to bistable solutions of the homogeneous balance Eq. (3) occurs, e.g., in the case of heat-conducting [16] or optical [17] systems, as well as upon radiation defect formation [14, 15]. Such a structural phase transition can be due to the fact that the formation of chemical bonds at the initial and final stages of a reaction occurs at different rates. For example, the formation of iron sulfide FeS molecules is a process with a higher activation energy, characterized by a smaller probability G_1 than the probability of attachment (G_2) of a sulfur atom to a layer that already contains such molecules. This means that $G_2 > G_1$ and that the curve of the $G(n)$ function has the form shown in Fig. 1 by the solid line. The existence of several stages of oxidation was reliably established, e.g., for the oxidation of nickel with oxygen [4] or silver with chlorine [5]; they were shown to include (1) chemisorption of an oxidizer on the free surface, (2) complete oxidation of the surface up to the formation of one or several monolayers of an oxide, and (3) a slow thickening of the oxide film.

The multistage character of the reaction causes a nonmonotonic reaction rate as a function of n, which corresponds to a nonlinear $G(n)$ dependence. As was indicated above, Eq. (3) can have three solutions if $G(n)$ is nonlinear. Among the solutions shown in Fig. 1, corresponding to "sparse" (n_l), intermediate (n_c), and "dense" (n_h) states of the system of adatoms or other quasiparticles, two extreme solutions represent stable

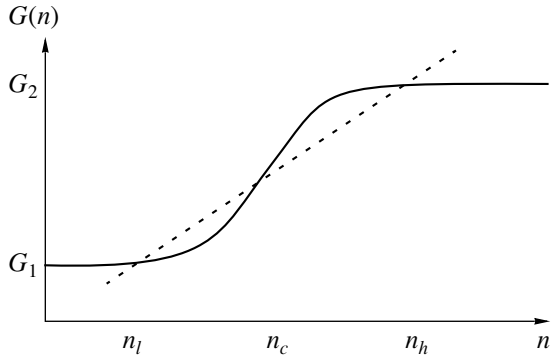

Fig. 1. Schematic dependences of the incoming ($G(n)$, solid line) and outgoing (n/τ, dashed line) terms of the kinetic Eq. (2) on the concentration of particles n. Along the horizontal axis, stable (n_l and n_h) and unstable (n_c) solutions of the stationary homogeneous Eq. (3) are indicated; along the vertical axis, the limiting values of G_1 and G_2 for the function $G(n)$ are given.

states. In the intermediate range of concentrations $n_l < n < n_h$, the balance Eq. (3) leads to bistability of the homogeneous state of the system. The arising instability of solutions of the diffusion Eq. (2) can be considered by analogy with bifurcations known from hydrodynamics [18].

Two regimes of the loss of stability of solutions of differential equations are known from hydrodynamics, which differ in their symmetry [18]. Their experimental effects are quite various; for surfaces, they consist in the formation of geometrical patterns of various shapes. They are observed, e.g., on metallic electrodes upon electropolishing [19] or in the form of local melting regions on illuminated silicon [20]. The regime of isotropic turbulence of the Qouette flow (e.g., in the space between coaxial cylinders) corresponds in the case of Eq. (2) to zero elastic force $\mathbf{F} = 0$. This regime leads to a transformation of closed isolines. Another regime that admits a linear analysis of stability using the Lyapunov criterion (see, e.g., [21]) is characterized by plane symmetry and corresponds to Poiseuille flow [18]. The Lyapunov exponents transform a laminar flow into a turbulent regime; a generalization of this law onto the phase space determines a new characteristic time of the system [22]. It is exactly the plane regime that is realized upon growing iron films on the surface of semiconductors such as Ge and GaAs. Upon the floating out of the layer of sulfur atoms that passivates the boundary of the semiconductor through the layer of iron atoms, the concentration n is maximum at the boundary and falls off to vanish at the free surface because of evaporation. This kinetic state is described approximately by the settling of a linear gradient $\partial n/\partial x = \text{const}$. As the x axis, the perpendicular to the surface plane was taken. By performing a Fourier transform with respect to the longitudinal coordinate y, which is oriented parallel to $\mathbf{F}$ [15, 18], we obtain the following expression for the linear addition to the steady-state concentration:

$$\delta n = A(t)\exp\{i(ky - \omega t)\} f(x), \qquad (4)$$

where k and ω are the wave vector and the frequency of the wavelike disturbance.

In the same approximation, we represent the functions $G(n)$ and $F(n)$ in the form of expansions in a small addition δn:

$$F(n) = C\delta n, \quad G(n) = \left(\frac{1}{\tau} + \nu\right)\delta n. \qquad (5)$$

Here, C is the coefficient of the linear expansion that is determined by the elastic force [12] that is due to the effect of the irradiating beam of particles, and the frequency ν is defined as

$$\nu = \frac{\partial G}{\partial n} - \frac{1}{\tau}. \qquad (6)$$

Substituting Eqs. (4)–(6) into Eq. (2), we obtain (for details, see [15]) that a longitudinal wave is established along the surface; for a constant gradient $\partial n/\partial x$, it is characterized by the frequency $\omega = knCb$ and damping $\gamma = k^2D - \nu$. Therefore, the condition sufficient for the solutions to Eq. (2) to be stable in the Lyapunov criterion sense is a significant sink of particles, i.e.,

$$\frac{1}{\tau} \geq \frac{\partial G}{\partial n}. \qquad (7)$$

Condition (7) is fulfilled in the region of the extreme roots n_l and n_h, where the chemical lattice reaction occurs steadily and uniformly. If the sink of particles is insufficient, so that $\nu > 0$, then a negative damping is possible, $\gamma = k^2D - \nu < 0$. This occurs in the central portion of Fig. 1, where, in a certain region near the intermediate root of Eq. (3), the straight line n/τ has a slope that is smaller than that of the $G(n)$ curve. If the stability criterion (7) is violated, the amplitude $A(t)$ increases in accordance with the Lyapunov exponent [22] because of the negative damping γ. However, this growth is limited by the nonlinear nature of Eq. (2). The equation for the squared modulus of the amplitude $A(t)$ averaged over the wave period, which describes its saturation, can be obtained by the Landau method [18]. Certainly, there are restrictions inherent in this method, but the comparison with the experiment [2–4] (see below) indicates that apparently all of these restrictions are met. It is the real boundary of stability found from the Lyapunov criterion [21, 22] that should be determined by that unique type of disturbances and by the same frequency $\omega(k)$ that yield zero damping ($\gamma = 0$). In this case, we can expand the averaged temporal derivative of $|A|^2$, i.e., $\langle\partial|A|^2/\partial t\rangle$, in A, in which, for the reasons that were indicated in [18], it is sufficient to retain only a few lowest terms that do not vanish upon averaging. The terms of odd orders in amplitude A necessarily contain a periodic factor and vanish upon averaging.

Among the terms of even, e.g., fourth, orders in A, there are terms such as $A^2A^{*2} = |A|^4$, which do not vanish upon averaging.

Note here, returning to the restrictions of the method, that the above temporal derivatives of the type $\langle\partial|A|^2/\partial t\rangle$ are direct analogs of the "probabilities of transition per unit time" that are calculated in perturbation theory on the basis of the nonstationary Schrödinger equation of quantum mechanics [23]. In this sense, the Landau method is applicable to nondegenerate systems. At a temperature close to absolute zero, any system is in a nondegenerate state. For open systems, the analog of temperature appears to be the intensity of an external action, so that the theory that is developed here is applicable for the case of negligibly small external flows. In economics [8], the conclusions that will be made below may be of interest in the area of insufficient financing. Our interest in the squared modulus $|A|^2$ in this case is related to the fact that the probability of coalescence of a pair of atoms per unit area or of the formation of other similar microstructures, e.g., divacancies, is proportional to the statistical weight, i.e., the number of combinations $C_n^2 = n(n-1)/2$. By averaging the expansion of this number into a series in the amplitude A, we obtain precisely $|A|^2$, and the rate of generation of sulfide bonds, molecules of FeS_2 or NiO_2, or divacancies in silicon wafers is equal to

$$\frac{\partial N_W}{\partial t} = \frac{1}{2}\left\langle\frac{\partial|A|^2}{\partial t}\right\rangle. \quad (8)$$

This rate is given by the average temporal derivative from the quantity of this type. By writing the Landau constant α, which was introduced in [18] in a similar expansion, through the limiting concentration of double bonds N_{0W}, we obtain

$$\frac{\partial N_W}{\partial t} = \frac{N_W}{\tau_l}\left(1 - \frac{N_W}{N_{0W}}\right), \quad (9)$$

where τ_l is the characteristic time determined by the Lyapunov exponent [22].

It was shown previously [11, 15] that coarse-scale fluctuations δn are settled in the system after this time, which lead to the evolution of the structure of semiconductor wafers and, in some cases, can be observed in the form of characteristic patterns on the surface. The solution to Eq. (9) is the exposure dose characteristic of the evolution type

$$N_W = \frac{N_{0W}}{1 + \exp[(t_0 - t)/\tau_l]}. \quad (10)$$

Scaling considerations permit us to unambiguously correlate the parameters of deposition of the films in [2–4] or burning out of molten regions in [20] with the theoretical constants that describe them; actually, this has already been done in Eq. (10). Since Eq. (9) is a first-order differential equation, its integration results in a single constant, which is designated in Eq. (10) as $\exp(t_0/\tau_l)$. There are grounds to believe that the constant t_0 that occurs in this equation is precisely that time that is required for the kinetic energy that is necessary to "swing" a corresponding fluctuation to be accumulated. For example, a delay t_0 can take place between the beginning of the bombardment of the surface with iron atoms and the floating out of sulfur atoms from the surface because of the inertia of the latter ones. The delay time t_0 may be considered independent of the intensity of the external action I, in particular, of the density of the flux of atoms from the outside. Note the dependence of the evolution time τ_I on the intensity of action I. Taking into account what we said above about the smallness of the external action I, we perform an expansion of the reciprocal time into a series in I: $\tau_I^{-1} = \xi + \eta I$, where ξ and η are positive phenomenological constants. To complete the consideration of the Stranski–Krastanov theory of evolution processes, we note that it is exactly that mechanism that is realized upon growth of lateral structures such as quantum filaments and quantum dots. They are formed via the diffusion mechanism of the development of a single monolayer, which is accompanied by decomposition through the Stranski–Krastanov mechanism. On the contrary, for growing superlattices, a vicinal surface should be used [24], and, in this case, thermal diffusion leads to the destruction of the growth steps.

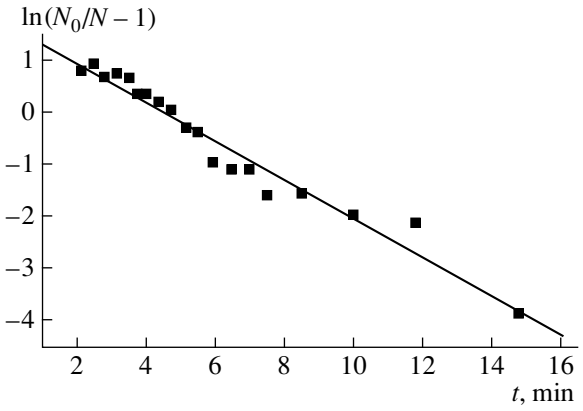

Fig. 2. Time variation of the intensity of the [1/2, 1/2] S/Fe(100) diffraction reflection in the process of iron deposition at 150°C. The experimental points are given according to [2]; the curve was obtained by the least-squares method using Eq. (10). The adjustable parameters are $a = t_0/\tau_l = 1.66$ and $b = 1/\tau_l = -0.37$ min^{-1}.

ANALYSIS OF EXPERIMENTAL RESULTS

The structure of iron films deposited on passivated surfaces of semiconductors was controlled in [2, 3] by the method of low-energy electron diffraction (LEED). The above theory permits one to describe the characteristic evolution type of the exposure characteristic con-

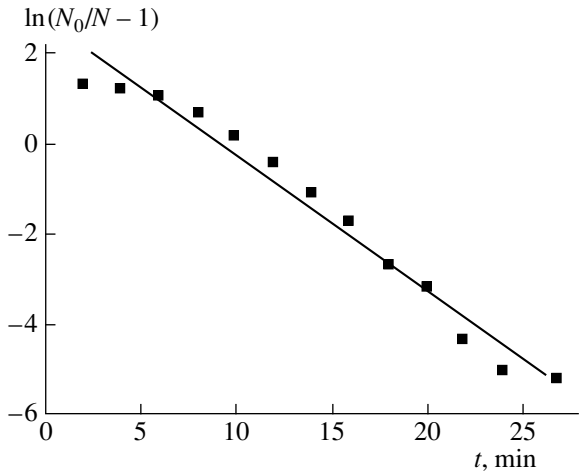

Fig. 3. The height of the oxygen Auger peak as a function of the time of oxygen deposition onto a pure nickel surface (100). The experimental points are given according to [4]; the curve was obtained by the least-squares method using Eq. (10). The adjustable parameters are $a = t_0/\tau_l = 2.77$ and $b = 1/\tau_l = -0.33$ min^{-1}.

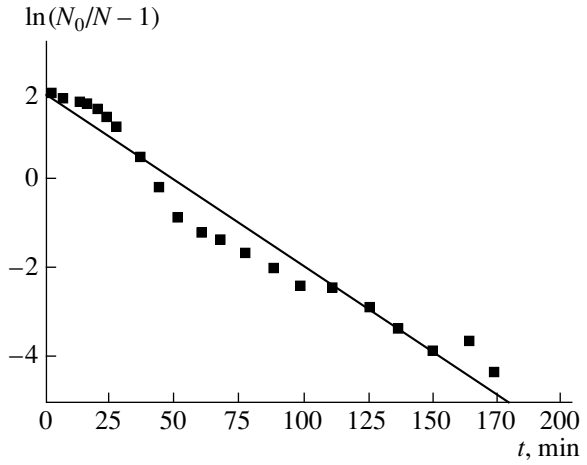

Fig. 4. Same as in Fig. 3 for the case of oxygen deposition onto a (100) surface coated with lead. The adjustable parameters are $a = t_0/\tau_l = 1.84$ and $b = 1/\tau_l = -0.038$ min^{-1}.

structed using LEED investigations. Upon oxidation of pure and lead-coated surfaces of nickel [4], the intensities of the oxygen Auger signal were measured, which were proved to depend on the exposure time too. In the latter case, the lead film also was subjected to a partial oxidation; the process of oxidation consisted in the diffusion of oxygen into the space between the nickel and the lead films. On the whole, the situation was the same as upon the deposition of iron films on the sulfided surface. A satisfactory agreement between the theory and the experiment indicates the nonuniform reconstruction of semiconductor and metallic surfaces in the process of deposition.

Figure 2 displays the intensity of the signal of diffraction reflections from a square lattice of sulfur atoms on the surface of iron (S/Fe(100) [1/2, 1/2] LEED spots) in the process of iron deposition at a temperature of 150°C according to [2]. In the process of deposition of iron and floating of sulfur atoms, a chain is formed around pairs of sulfur atoms, and, in this way, a cluster of the square lattice of sulfur atoms is formed. We assume that the brightness of the S/Fe(100) [1/2, 1/2] LEED spots is proportional to the number of nanoclusters that are formed in this way, i.e., to the number of the initial pairs (FeS_2 molecules). The straight line in Fig. 2 represents the fitting of the dose dependence of the above diffraction reflections using Eq. (10). In the caption to Fig. 2, we give the numerical values of the coefficients that describe the straight line $y = a + bt$ (fitting parameters); the corresponding evolution time $\tau_l = -1/b$ for Fig. 2 turned out to be $\tau_l = 2.70$ min. It was the fulfillment of the conditions of the theory of partially inhomogeneous reconstruction of the surface (see Eq. (10)) that, in our opinion, predetermined the retention of magnetic properties of the films [2]. It is of interest that this high quality was obtained at a moderate temperature of deposition $T = 150$°C, which turned out to be possible due to the advantages of the technique of sulfiding [11]. The adjustable delay time proved to be $t_{02} = 4.48$ min. The magnetic structure of the film is controlled by the formation of sulfide pairs inside the film. The division of the entire iron layer into domains with a low (n_l) and a high (n_h) concentration of sulfur explains the retention of the magnetic properties of the film upon its saturation with sulfur. In this case, the magnetic domains become closed through the regions with a low concentration of sulfur n_i.

Figures 3 and 4 display dose characteristics of the processes of oxidation of the pure and lead-coated surfaces of nickel. The straight lines represent the fitting of the experimental data points (squares) obtained by Auger spectroscopy using Eq. (10) by the linear least-squares method. The evolution time $\tau_l = -1/b$ was obtained to be 3.00 min for the pure surface (Fig. 3) and 26.3 min for the surface coated with lead (Fig. 4). The delay times t_0 also differ by about an order of magnitude, namely, $t_{01} = 8.3$ min for the pure surface and $t_{02} = 48.4$ min for the surface coated with lead. The good agreement between the experiment and the theory developed, just as in the case of iron films, indicates the high quality of the films; oxygen films were deposited at room temperature. The difference in the times for the two different regimes of oxidation characterizes the protecting effect of the lead coating and also confirms the adequacy of the theory.

REFERENCES

1. V. G. Karpov and M. Grimsditch, Phys. Rev. B **51** (13), 8152 (1995).
2. G. W. Anderson, P. Ma, and P. R. Norton, J. Appl. Phys. **79**, 5641 (1996).

3. G. W. Anderson, M. C. Hanf, and P. R. Norton, Phys. Rev. Lett. **74** (14), 2764 (1995).
4. C. Argile, Surf. Sci. **409**, 265 (1998).
5. B. V. Andryushechkin, K. N. El'tsov, and V. V. Martynov, Poverkhnost, No. 8, 72 (1999).
6. V. N. Petrov, V. N. Demidov, N. P. Korneeva, *et al.*, Zh. Tekh. Fiz. **70** (5), 97 (2000) [Tech. Phys. **45**, 618 (2000)].
7. E. G. Rapis, Zh. Tekh. Fiz. **70** (1), 122 (2000) [Tech. Phys. **45**, 121 (2000)].
8. A. Cavagna, J. P. Garrahan, I. Giardina, and D. Sherrington, Phys. Rev. Lett. **83** (21), 4429 (1999).
9. J. A. Venables, G. D. T. Spiller, and M. Hanbucken, Rep. Prog. Phys. **47** (4), 399 (1984).
10. S. A. Kukushkin and V. V. Slezov, *Dispersive Systems on Solid Surfaces: Mechanism of Formation of Thin Films* (Nauka, Leningrad, 1996).
11. V. A. Voĭtenko and S. E. Mal'khanov, Poverkhnost, No. 10, 72 (1999).
12. I. P. Ipatova, V. G. Malyshkin, V. A. Shchukin, *et al.*, Phys. Low-Dimens. Struct. **3–4**, 23 (1997).
13. A. Ya. Kipnis, *Popular Library of Chemical Elements* (Nauka, Moscow, 1977), Vol. 1.
14. V. A. Voĭtenko and S. E. Mal'khanov, Zh. Éksp. Teor. Fiz. **112** (2), 707 (1997) [JETP **85**, 386 (1997)].
15. V. A. Voĭtenko and S. E. Mal'khanov, Zh. Éksp. Teor. Fiz. **114** (3), 1067 (1998) [JETP **87**, 581 (1998)].
16. A. V. Subashiev and I. M. Fishman, Zh. Éksp. Teor. Fiz. **93** (6), 2264 (1987) [Sov. Phys. JETP **66**, 1293 (1987)].
17. N. N. Rozanov, *Optical Bistability and Hysteresis in Distributed Nonlinear Systems* (Nauka, Moscow, 1997).
18. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).
19. V. V. Yuzhakov, Hsueh-Chia Chang, and A. E. Miller, Phys. Rev. B **56**, 12608 (1997).
20. Ya. V. Fattakhov, I. B. Khaĭbullin, R. M. Bayazitov, *et al.*, Poverkhnost, No. 11, 61 (1989).
21. G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1961; Nauka, Moscow, 1973).
22. A. K. Pattanayak, Phys. Rev. Lett. **83** (22), 4526 (1999).
23. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1956; Pergamon, New York, 1977.
24. I. L. Aleĭner and R. A. Suris, Fiz. Tverd. Tela (St. Petersburg) **34** (5), 1522 (1992) [Sov. Phys. Solid State **34**, 809 (1992)].

Translated by S. Gorin

Technical Physics, Vol. 46, No. 2, 2001, pp. 240–244. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 111–116.

Interaction of Graphite with Atomic Hydrogen

E. A. Denisov and T. N. Kompaniets

St. Petersburg State University, Universitetskaya nab. 7/9, St. Petersburg, 198904 Russia

Received January 31, 2000; in final form, March 29, 2000

Abstract—The kinetics of sorption of atomic hydrogen by pyrolytic, quasi-single-crystal, and RGT commercial-grade graphite was studied. The processes of sorption and subsequent thermal outgassing are shown to proceed in a similar manner for all the three types of graphite. Thermal desorption spectra obtained during linear heating of hydrogen-saturated samples have two peaks. A mathematical model including features of the thermal desorption kinetics that are observed when heating is terminated is suggested. According to this model, two types of traps with binding energies of 2.4 and 4.1 eV are present in graphite. The physical justification of the model is given. © *2001 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

Over recent years, most of studies on hydrogen–graphite interaction have been concerned with applications, namely, with the use of graphite for protecting the first wall of fusion reactors. Therefore, interactions of several-keV hydrogen ion beams with commercial-grade graphites have largely been considered. A large body of both experimental and theoretical data is available in this field [1–12]. In these works, emphasis has been given to chemical erosion of the graphite surface [8, 13–18]. As to desorption, only that of hydrocarbons has been studied extensively in spite of the fact that the amount of hydrogen incorporated into hydrocarbons leaving the surface during thermal desorption is much smaller than the amount of hydrogen desorbed as H_2 molecules. In addition, in most of the studies it is assumed that atomic hydrogen–graphite interaction is a purely surface process and the possibility of hydrogen dissolution in graphite is usually ignored. We believe that, when graphite is exposed to an atomic hydrogen flux, the possibility of hydrogen dissolution may markedly increase in comparison with the situation when equilibrium molecular hydrogen is involved. At least some hydrogen atoms may occur beneath the adsorbent surface even at room temperature. Unfortunately, the simulation of atomic hydrogen sorption by graphite with volume processes taken into account has been reported only in two of the works cited above [17, 18].

As was noted, most of sorbed hydrogen molecules leave graphite as H_2 molecules as the temperature rises. Available literature data on the kinetics of this process are, however, scarce and refer mostly to the estimation of the number of hydrogen molecules from thermal desorption spectra (TDS). The only attempt to determine the energy of activation of desorption was based on the *a priori* assumption that this process is of a dissociative nature [8]. Therefore, is was the aim of this work to determine the kinetic and energy parameters of desorption processes that take place after the graphites have been irradiated by atomic hydrogen. Based on the kinetic information, one can develop models of processes occurring in this system and, in particular, separate surface and volume effects contributing to atomic hydrogen sorption.

MATERIALS

Commercial-grade RGT graphite (ρ = 2.20–2.26 g/cm^3) contains about 7.5 at. % of titanium and is produced by unidirectional hot pressing of carbon–titanium powder [19]. Most of the RGT grains are disk-shaped and lie parallel to the basal plane of the graphite lattice. The mean grain size is about 10 μm.

Pyrolytic graphite is of density ρ = 2.186 g/cm^3. On a microscale, it has a layered structure with a spacing of 0.5–1.0 μm in the growth (*c*-axis) direction because of the step growth kinetics. We studied two types of pyrolytic graphite that differed in temperature of final annealing: true pyrolytic graphite (PG) and quasi-single-crystal graphite (QSCG). The former looked like commercial-grade graphite with a rough surface. QSCG had a much smoother and shiny surface, which indicates its higher structural order.

EXPERIMENTAL

The sorption/desorption kinetics was studied with thermal desorption spectrometry. Graphite samples were made in the form of ribbons measuring $1 \times 40 \times 0.5$ mm. The surface of the ribbon was parallel to the basal graphite plane. The sample was attached to current leads and placed into a vacuum chamber. The residual (mainly hydrogen) pressure was kept at 10^{-8} torr. The temperature, which can be varied according to a specified law, was measured with a W/WRe thermocouple. The desorbed hydrogen was detected by a magnetic sectorial mass spectrometer. Prior to sorption experiments, the samples were annealed for a long

1063-7842/01/4602-0240$21.00

time at 1200°C. At the end of annealing, the temperature was momentarily raised to 1400°C.

Purified hydrogen was supplied to the chamber through a diffusion filter. Atomization was carried out when the gas passed near a 100-μm-diam. tungsten filament heated to 2500°C. The filament was arranged parallel to the sample at a distance of 5–8 mm so that the atomic hydrogen flux struck the surface parallel to the basal plane of the sample. The irradiation dose was calculated with regard for the inlet hydrogen pressure, atomization yield, and mutual arrangement of the sample and the atomizer.

The probability of hydrogen atomization on tungsten heated to 2100°C was taken equal to 0.3 [20]. During exposure, the hydrogen pressure was 10^{-2} torr. The flux of hydrogen atoms toward the front side of the sample was estimated at 5×10^{13} $H^0/(cm^2\ s)$ in view of the experiment geometry.

EXPERIMENTAL RESULTS

Unlike the RGT samples [21], the PG and QSCG samples did not sorb molecular hydrogen at pressures below 5 torr and temperatures below 600°C. During linear heating, noticeable desorption of H_2 was observed only after these graphites had been irradiated by atomic hydrogen. A typical postirradiation TDS for the PG is presented in Fig. 1. For the other types of graphites, the spectra are similar. No marked differences were also observed in the atomic hydrogen sorption kinetics. Dependences of the sorbed hydrogen quantity on the sorption temperature and irradiation dose for all the three graphites are much the same. The structure of the graphites and their purity seem to insignificantly affect the sorption/desorption kinetics with atomic hydrogen involved.

Two TDS peaks at 850 and 1250°C in Fig. 1 are related to hydrogen release from two states with different binding energies. The temperatures of the desorption peaks were independent of the initial hydrogen concentration, indicating the first order of the desorption kinetics. If heating is linear and the kinetics is of the first order, $\ln(\alpha/T_m^2)$ should vary linearly with E_d/kT_m (α is the rate of heating, T_m is the temperature of the desorption maxima, E_d is the energy of activation of desorption, and k is the Boltzmann constant) [22]. From the slope of this dependence, one easily obtains the energy of activation of desorption. For the first state, the energy of activation estimated for three rates of heating was found to be 2.4 eV. For the second state, the energy of activation can be estimated only roughly, since the temperature of heating was close to that of the second desorption peak (at repeated heating to higher temperatures, the sample rapidly broke down). The value of E_d for the second state was evaluated at about 4 eV. Note that the above energies of activation were obtained under the assumption that hydrogen is present only on the surface.

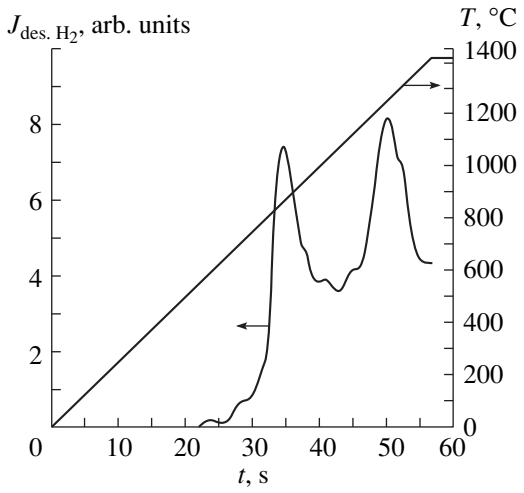

Fig. 1. Typical TDS of PG irradiated by hydrogen atoms. Linear heating with a rate of 25 K/s.

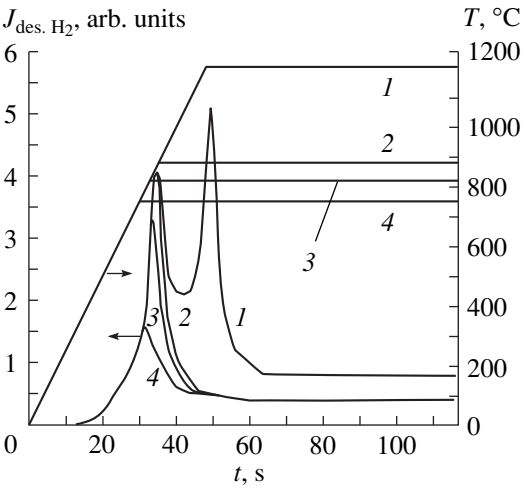

Fig. 2. TDS obtained after sorption of H^0 when linear heating was terminated at different temperatures. PG, heating rate 25 K/s.

In the next series of experiments, the sorption conditions were identical, but the final heating temperatures were different (Fig. 2). When heating was terminated (Fig. 2, curves *1–4*), the desorption rate began to fall sharply, the time of fall being almost independent of the final temperature of heating. Within 10–20 s, the desorption rate decreased to several percent of its maximum; i.e., hydrogen release was nearly completely stopped. The desorption rate sharply decreases in spite of the fact that the sample still contains a large amount of hydrogen, which is released at a subsequent rise in the temperature (Fig. 3, curves *1–3*). In the experiments the results of which are illustrated in Fig. 3, the sample was first heated to some intermediate temperature, which was kept for 45 s, and then heating was contin-

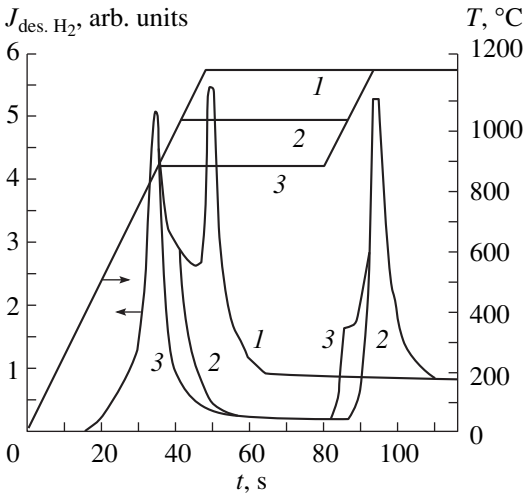

Fig. 3. TDS obtained after sorption of H^0 when linear heating was first terminated at different temperatures and then continued to 1200°C. PG, heating rate 25 K/s.

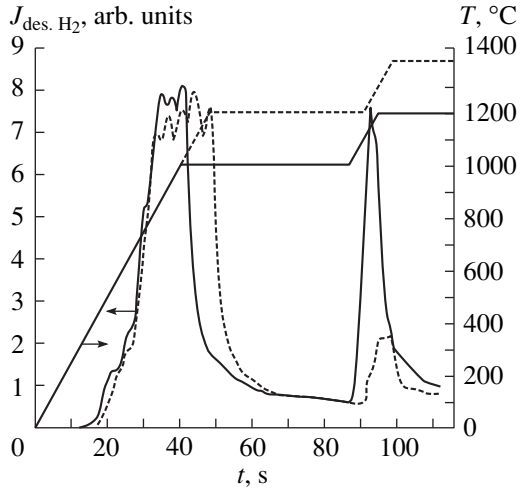

Fig. 4. TDS of PG irradiated by ~200-eV H^+ ions. Linear heating was terminated at 1000 and 1200°C, heating rate 25 K/s.

ued at the same rate to 1200°C. After the second heating, the remaining hydrogen was totally released, its amount being practically the same as in the case when it was desorbed at similar temperatures but under continuous heating. The kinetic characteristics of hydrogen release also remained unchanged. These features were typical of all three graphites.

Such an unusual desorption kinetics cannot be explained in terms of classical desorption from surface. For the first-order desorption kinetics, when heating is stopped, the desorption rate must decrease exponentially with a time constant proportional to $\exp(-E_d/kT)$; hence, at E_d = 2.4 eV, the rate of fall at 600°C and 800°C would differ by several orders of magnitude.

It should be noted that the unusual hydrogen desorption kinetics was also observed after the graphites had been irradiated by ~100-eV hydrogen ions. The hydrogen was desorbed in the same temperature interval; the desorption rate sharply dropped after heating had been terminated; and after the second rise in the temperature, the remaining hydrogen was entirely desorbed. The only difference was the absence of two distinct peaks (Fig. 4).

DISCUSSION

An adequate kinetic model is the first step toward understanding basic processes underlying gas–solid interaction. The model of choice must take into consideration the following. Upon irradiation, hydrogen is implanted into graphite part way down the surface (tens of nanometers for ion energies of several hundred electron volts [6]), and mass transfer conclusively plays a part in the outgassing process. Since the process kinetics are similar after irradiation by atoms and by ions, one can suggest that a considerable fraction of hydrogen atoms penetrates into the volume of the sample; hence, the outgassing kinetics also includes mass transfer in the bulk of the graphite. Note, however, that all the features of the hydrogen evolution kinetics (for example, the presence of two peaks) cannot be covered by the simple diffusion model.

Thus, graphite outgassing should be described in terms of at least three processes two of which are strongly temperature-dependent (i.e., have a large energy of activation) and show up as the TDS peaks. The third one, which is responsible for a decrease in the desorption rate when heating is stopped, must depend on temperature only slightly. This may be hydrogen diffusion between adjacent graphite layers [23]. This process requires a very small activation energy largely because these layers are loosely bonded and spaced at more than 3 Å apart. From our estimates [24], the energy of activation of diffusion in RGT graphite is no more than 0.5 eV. Therefore, the following model seems to be the most plausible. During sorption of the atoms or ions, a hydrogen-saturated layer of some thickness forms beneath the surface and two sorts of traps arise in the graphite. These traps, having large energies of activation, capture the hydrogen. Upon heating, the hydrogen escapes from the traps into the mobile phase, with its migration over the sample being accompanied with reverse trapping. It is also assumed that the rate of desorption from the surface far exceeds the rate of the other processes and is not the limiting stage of outgassing.

To mathematically represent this model (diffusion with reverse capture by traps of two sorts), we used a set of differential equations from our previous work [25].

Calculations which follow substantiate the determination of the activation energy of hydrogen detrapping

using the slope of the $\ln(\alpha/T_m^2)$ vs. $1/T_m$ curve. Within the model suggested, we evaluated TDS for linear heating with various rates. The energy of hydrogen evolution from traps of the first sort was assumed to be 2.4 eV, and the diffusion coefficient was varied between 10^{-7} and 10^{-4} cm^2/s. It turned out that the slope of the $\ln(\alpha/T_m^2)$ vs. $1/T_m$ curve obtained in the model was practically coincident with that of the curve constructed from experimental data (Fig. 5). Therefore, the value of 2.4 eV was taken as the energy of hydrogen evolution from traps of the first sort. Also noteworthy is that the energy of activation of detrapping can be determined without knowing the exact value of the diffusion coefficient. As the diffusion coefficient value, we took the one obtained in experiments on sorption of molecular hydrogen by RGT graphite: $D = 10^{-6}$ cm^2/s [24]. As follows from calculations, with such a diffusion coefficient and the saturated layer thickness of the order of the RGT grain size, the falls in the desorption curves when heating is switched off are adequately described if unactivated capture by traps of the first sort has a time constant of 25 s^{-1}.

Capture by traps of the second sort must be of an activation nature, since transitions from one state to the other were observed at none of the rates and final temperatures of heating used. The model adequately describes both thermal desorption under linear heating to 1200°C and thermal desorption falls when heating is terminated (Fig. 6). The parameter values used were as follows: preexponential in the expression for diffusion coefficient $D_0 = 10^{-6}$ cm^2/s, energy of activation of diffusion $E_D = 0$, preexponentials in trapping rate constants for traps of the first and second sort $r_1 = 25$ and $r_2 = 7.0 \times 10^7$ s^{-1}, energies of trapping activation for traps of the first and second sort $E_{r_1} = 0$ and $E_{r_2} = 2.0$ eV, preexponentials in detrapping rate constants for traps of the first and second sorts $b_1 = 2.0 \times 10^{11}$ and $b_2 = 5.0 \times 10^{13}$ s^{-1}, and energies of detrapping activation for traps of the first and second sort $E_{b_1} = 2.4$ and $E_{b_2} = 4.1$ eV.

Note that the vast majority of studies where hydrogen ion–graphite interaction was simulated points to the necessity of considering hydrogen capture by traps localized in the bulk of graphite. The binding energies of trapped hydrogen vary in wide limits: from 0.5 to 4.5 eV [1–3, 5, 7–11, 26–31].

The nature of the traps and specific sorption/desorption processes still remains unclear. Dangling bonds of carbon atoms in the $(10\bar{1}0)$ and $(11\bar{2}0)$ planes seem to be the most probable sites of hydrogen sorption. Linear defects like dislocations (dangling boundaries of graphite layers) may readily form in graphite because of creeping of one layer on another during pyrolysis. The C–H bond energy on the $(10\bar{1}0)$ graphite plane is

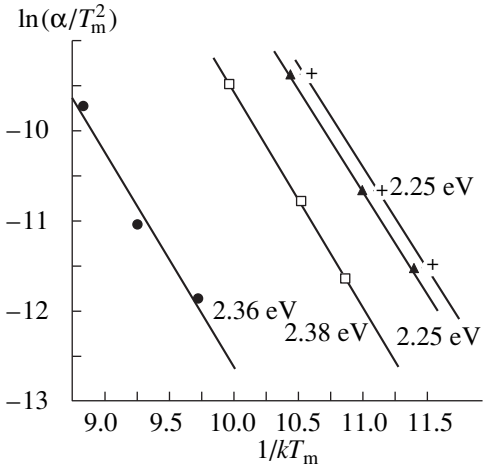

Fig. 5. $\ln(\alpha/T_m^2)$ vs. $1/T_m$ for the first desorption peak within the diffusion–recapture model. D = (●) 10^{-7}, (□) 10^{-6}, (▲) 10^{-5}, and (+) 10^{-4} cm^2/s.

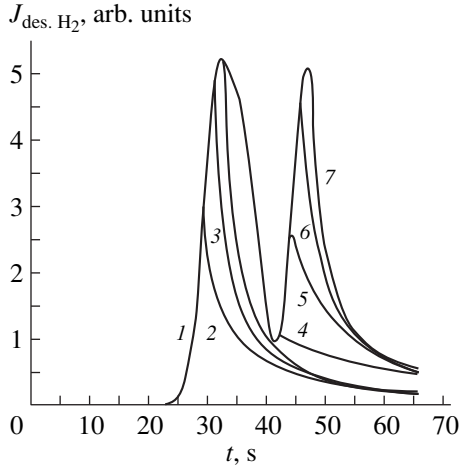

Fig. 6. Model TDS calculated for final linear heating temperatures (*1*) 730, (*2*) 770, (*3*) 820, (*4*) 1050, (*5*) 1100, (*6*) 1140, and (*7*) 1190°C. Heating rate 25 K/s.

higher than on the $(11\bar{2}0)$ plane by 0.8 eV [32]. Thus, the difference in bond energy per molecule for the different planes at the boundary of a graphite layer is 1.6 eV. This value is in good agreement with the above results.

We believe that hydrogen capture by the traps of the first or the second sort can be related to the production of C–H bonds with carbon atoms belonging to edge dislocations on the $(11\bar{2}0)$ or $(10\bar{1}0)$ planes, respectively, in the bulk of graphite. During sorption, transport of hydrogen in the form of atoms saturates dangling bonds of the edge dislocations to the maximum possible extent. During thermal desorption, the hydrogen is detrapped in the form of molecules and migrates along graphite layers until it reaches the crystallite boundary

and leaves the material through the pore network, grain boundaries, or other macrodefects.

It should be noted that this is only a possible physical interpretation of the mathematical model. Basically, thermal desorption spectrometry cannot elucidate the physical nature and location of the traps. In layered structures like graphite, sorption/desorption processes proceed in a rather unusual manner. For example, foreign atoms and molecules may be accumulated in large amounts between graphite layers [33]. Unfortunately, intercalation in the hydrogen–graphite system has yet received little attention. One can assume, however, that hydrogen atoms falling between graphite layers recombine to form molecules and become "blocked" in interplanar gaps. In this case, hydrogen release would require either molecule dissociation or severe deformation (up to local destruction) of the graphite plane; hence, energy must be spent to activate desorption. To confirm the validity of the above assumptions, atomic-hydrogen-induced modifications of the graphite surface must be examined.

CONCLUSION

Thus, when atomic hydrogen interacts with PG, QSCG, and commercial-grade RGT graphites, sorption/desorption processes are virtually independent of the material type, being governed by processes in the bulk. Hydrogen release during heating is well described by the mathematical model including diffusion with reverse capture by traps of two sorts. Physically, the traps are treated as dangling bonds of edge dislocations on the $(11\bar{2}0)$ and $(10\bar{1}0)$ planes. Diffusion of atoms goes both along and across graphite layers, while molecular diffusion takes place in interplanar gaps only.

REFERENCES

1. W. Moller, J. Nucl. Mater. **162–164**, 138 (1989).
2. K. L. Wilson and W. L. Hsu, J. Nucl. Mater. **145–147**, 121 (1987).
3. R. A. Causey, J. Nucl. Mater. **145–147**, 151 (1987).
4. P. G. Fischer, R. Hecker, H. D. Rohrig, and D. Stover, J. Nucl. Mater. **64**, 281 (1977).
5. K. Nakayama, S. Fukuda, T. Hino, and T. Yamashina, J. Nucl. Mater. **145–147**, 301 (1987).
6. W. R. Wampler and C. W. Magee, J. Nucl. Mater. **103**, 509 (1982).
7. G. Hansali, J. P. Biberian, and M. Bienfait, J. Nucl. Mater. **171**, 395 (1990).
8. V. Phyllipps, E. Vietzke, M. Erdweg, and K. Flaskamp, J. Nucl. Mater. **145–147**, 292 (1987).
9. S. Fukuda, T. Hino, and T. Yamashina, J. Nucl. Mater. **162–164**, 997 (1989).
10. D. K. Brice, Nucl. Instrum. Methods Phys. Res. B **44**, 302 (1990).
11. R. A. Causey, M. I. Blaskes, and K. L. Wilson, J. Vac. Sci. Technol. A **4**, 1189 (1986).
12. D. Federici and C. H. Wu, J. Nucl. Mater. **186**, 131 (1992).
13. P. C. Stangeby, O. Ausiello, A. A. Haasz, and B. L. Doyle, J. Nucl. Mater. **122–123**, 1592 (1984).
14. P. Hucks, K. Flaskamp, and E. Vietzke, J. Nucl. Mater. **93–94**, 558 (1980).
15. J. W. Davis, A. A. Haasz, and P. C. Stangeby, J. Nucl. Mater. **155–157**, 234 (1988).
16. I. S. Youle and A. A. Haasz, J. Nucl. Mater. **182**, 107 (1991).
17. T. Tanabe and Y. Watanabe, J. Nucl. Mater. **179–181**, 231 (1991).
18. M. Balooch and D. R. Olander, J. Chem. Phys. **63**, 4772 (1975).
19. A. P. Zakharov, Report on the Contract N 7/4 between Sintez NTO, St. Petersburg, Russia and Fusion Centre (Moscow, 1995).
20. J. Smith and W. Fait, in *Interaction of Gases with Surfaces* (Mir, Moscow, 1965), p. 362.
21. E. Denisov, T. Kompaniets, A. Kurdyumov, and S. Mazaev, J. Nucl. Mater. **233–237**, 1218 (1996).
22. D. Woodruff and T. Delihar, *Modern Techniques of Space Science* (Cambridge Univ. Press, Cambridge, 1986; Mir, Moscow, 1989).
23. B. M. U. Scherzer, J. Wang, and W. Moller, Nucl. Instrum. Methods Phys. Res. B **45**, 54 (1990).
24. E. A. Denisov, T. N. Kompaniets, A. A. Kurdyumov, and S. N. Mazaev, J. Nucl. Mater. **212–215**, 1448 (1994).
25. E. Denisov, T. Kompaniets, A. Kurdyumov, and S. Mazaev, Plasma Devices Op. **6**, 265 (1998).
26. E. Hoinkis, J. Nucl. Mater. **182**, 93 (1991).
27. W. R. Wampler, B. L. Doyle, R. A. Causey, and K. Wilson, J. Nucl. Mater. **176–177**, 983 (1990).
28. J. Redmond and P. L. Walker, J. Chem. Phys. **64**, 1093 (1960).
29. A. P. Zakharov, *Processes of Accumulation and Reemission of Hydrogen Isotopes in Carbonic Materials on Interaction with Ion and Plasma Streams (Report)* (Inst. Fiz. Khim. Ross. Akad. Nauk, Moscow, 1991).
30. S. L. Kanashenko, A. E. Gorodetsky, V. N. Chernikov, *et al.*, J. Nucl. Mater. **233–237**, 1207 (1996).
31. V. N. Chernikov, A. E. Gorodetsky, S. N. Kanashenko, *et al.*, J. Nucl. Mater. **217**, 250 (1994).
32. J. P. Chen and R. T. Yang, Surf. Sci. **216**, 481 (1989).
33. S. A. Solin and H. Zabel, Adv. Phys. **37**, 87 (1988).

Translated by V. Isaakyan

Technical Physics, Vol. 46, No. 2, 2001, pp. 245–248. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 117–119.

On the Possibility of Determining Spalling Strength of Metals in a Submillisecond Range from Experiments with a Penetrating Radiation Effect on a Target

T. A. Andreeva and S. N. Kolgatin

St. Petersburg State Technical University, ul. Politekhnicheskaya 29, St. Petersburg, 195251 Russia

e-mail: Kolgatin@spes.stu.neva.ru

Received December 1, 1999

Abstract—The strength of a metal under the action of short pulses is of great interest. Meanwhile, the problem concerning the magnitude of a tensile stress that leads to cleavage has not yet been solved theoretically. In this paper, an attempt is made to clarify, using the mathematical simulation method, the parameters of pulse radiation that can yield the clearest answer to the question of which of the existing theories of strength works best in metallic materials subjected to short pulses of micro- or nanosecond range. Recommendation are given on the optimal experimental design. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The strength of a metal under the action of short pulses of micro- or nanosecond range is of great practical interest. Meanwhile, the question of the magnitude of a tensile stress that leads to cleavage has not yet been solved theoretically. There are facts [1] that favor the kinetic theory of strength [2]. Some authors [3] believe that the strength attainable in the submillimeter range is independent of the pulse duration and is a result of motion of elementary carriers of plastic deformation. The final choice of this or that theory must be made on the basis of an analysis of additional experimental data. It would be convenient to extract such data from experiments on the effects of pulse penetrating radiation on a flat plate using the interferometric method of measuring the velocity of motion of the back surface of a target. In this case, the understanding of gas-dynamics processes that occur upon the absorption of energy in the target would favor the correct interpretation of the results obtained. In this work, we made at attempt, using the mathematical simulation method, to clarify the variation of which parameters of pulse radiation can yield the clearest (from the viewpoint of the problem stated) results. Recommendations are given on the optimal experiment design.

RESULTS AND DISCUSSION

There exist numerous types of devices based on pulsed sources of penetrating radiation that can be employed for generating compression waves in metallic targets (relativistic electron beams [4], lasers [5], X-ray sources based on *Z* pinches [6], etc.). The radiations differ in penetrating power, which can arbitrarily be characterized by parameters such as the average mass coefficient of absorption χ, the pulse duration τ_{pulse}, and the power of the heat flux on the surface q_0. By varying one or two of these parameters, we can change (in certain limits) the criterion of momentariness [7] and, consequently, the amplitude of the shock pulse in metal. Of practical interest is the dependence of the time for which the metal exists in the state of tension on the pulse parameters, as well as the intervals of χ, τ_{pulse}, and q_0 in which one can distinguish the exact instant at which spalling occurs.

We describe the formation and propagation of a compression wave in the target by a set of equations of gas dynamics, which, in the Lagrangian coordinates $s = \int_0^x \rho(\xi)d\xi$ (where ρ is the density, x is the Euler coordinate, and ξ is the integration variable), has the following form:

$$\frac{\partial(1/\rho)}{\partial t} = \frac{\partial v}{\partial s}, \quad \frac{\partial x}{\partial t} = v, \quad \frac{\partial v}{\partial t} = -\frac{\partial p}{\partial s}, \qquad (1)–(3)$$

$$\frac{\partial \varepsilon}{\partial t} = -p\frac{\partial v}{\partial s} - \frac{\partial W}{\partial s} + Q(s, t), \quad W = -k\frac{\partial T}{\partial s}, \quad (4), (5)$$

$$p = f_p(\rho, T), \quad \varepsilon = f_\varepsilon(\rho, T), \quad k = \rho f_\lambda(\rho, T). \quad (6)–(8)$$

Here, t is the time; v is the velocity; p, the pressure, ε, the internal energy; W, the heat flow due to thermal conductivity; Q, the specific energy contribution; k and λ, the thermal diffusivity and thermal conductivity coefficients, respectively; f_p and f_ε, the thermal and caloric equations of state, respectively, taken from [8]; and f_λ, the dependence of the thermal conductivity on the density and temperature, which is defined as in [7], using the dependences from [9, 10]. The energy contribution Q was calculated by the Bouguer law

$$Q = q_0\chi\exp(-\chi s)f_t, \qquad (9)$$

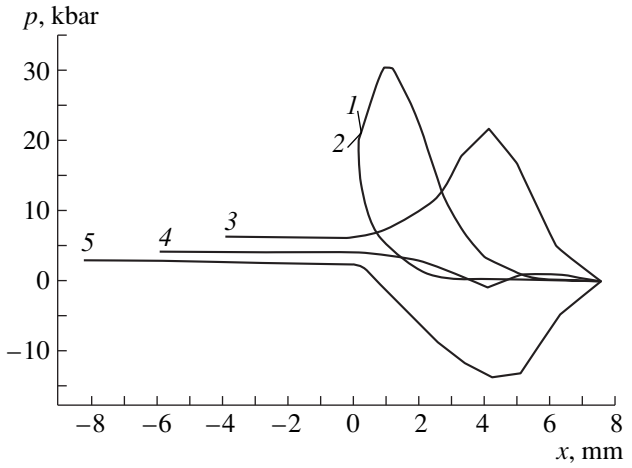

Fig. 1. Distribution of pressure over the thickness of the plate at various time moments.

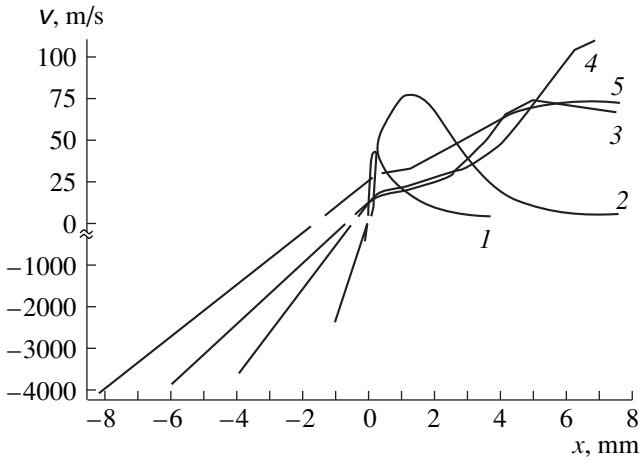

Fig. 2. Velocity fields inside the target at various time moments.

where q_0 is the density of the energy flux on the surface; χ is the mass-averaged absorption coefficient (a constant quantity); and f_t is a function of the temporal shape of the pulse, which approximately was assumed to be

$$f_t = \begin{cases} 1 - \left|1 - \dfrac{t}{t_{\text{pulse}}}\right|, & t \le 2t_{\text{pulse}} \\ 0, & t > 2t_{\text{pulse}}. \end{cases} \tag{10}$$

The initial conditions for the above set of equations had the form

$$\begin{aligned} v(0, s) &= 0; \quad p(0, s) = p_0; \\ \rho(0, s) &= \rho_0; \quad T(0, s) = T_0. \end{aligned} \tag{11}$$

On the surface ($s = 0$), the following boundary conditions were used:

$$p(t, 0) = p_0, \quad W(t, 0) = \sigma_{\text{St}}(T^4 - T_0^4), \tag{12}$$

where σ_{St} is the Stefan–Boltzmann constant and T_0 is the initial temperature.

On the second surface ($s = s_M$), we have

$$p(t, s_M) = p_0, \quad T(t, s_M) = T_0. \tag{13}$$

The set of gas-dynamics equations (1)–(13) was solved by the finite-difference method [11]. To provide the stability of calculations, a linear artificial viscosity was used. As a sample for the investigation, a copper plate 7 mm thick was taken. In order to answer the question posed at the beginning of the paper, the amplitude and the duration of the energy pulse, as well as the magnitude of the mass absorption coefficient, were varied.

Figure 1 displays the evolution of a compression wave in the plate under the effect of a radiation pulse with the parameters $t_{\text{pulse}} = 1$ μs, $\chi = 2.8$ m^2/kg, and $q_0 = 6 \times 10^{13}$ W/m^2. Curve *1* refers to the time moment $t = 0.5$ μs (radiation maximum); curve *2* refers to the instant $t_{\text{pulse}} = 1.0$ μs (the end of the pulse); curve *3*, to the instant $t_{\text{pulse}} = 2.0$ μs; curve *4*, to $t_{\text{pulse}} = 2.6$ μs; and curve *5*, to $t_{\text{pulse}} = 3.1$ μs. At the initial stage, there is formed a pressure peak at the wall (curve *1*), which then propagates to the target depth (curves *2*, *3*). After the compression pulse emerges onto the free back surface of the target, a rarefaction region is formed near it (curve *5*). The size of the rarefaction region is virtually independent of the parameters of the radiation pulse such as t_{pulse} and χ, whereas the amplitude of the tensile stress is mainly dependent on q_0. The maximum tensile stress and the rate of deformation are observed at a point that is located about 2.0 mm from the back surface (this corresponds to the distance passed by the sound in a period equal to half the pulse duration).

The pressure in the plate can be estimated from the velocity of motion of the back surface, which in turn can be measured by the interferometric methods [12].

The velocity fields at various time moments for the variant under consideration are shown in Fig. 2. At the initial stage, an evaporated gas layer is stripped off from the front surface of the target; the velocity of this layer in the coordinate system chosen is negative, whereas in the condensed medium a flow is formed that is directed toward the radiation flow (curves *1*, *2*). As the shock wave propagates into the metal, the velocity of the rear surface first increases gradually (curves *3*, *4*) and then decreases (curve *5*).

The allowance for the fracture of the target material was performed according to both above-mentioned theories. According to the first of them (the temporal theory of strength [2]), the material service life depends on the applied stress and temperature, which activate the process of breaking interatomic bonds. According to the second theory [3], the strength properties of the material in a submillimeter range are the result of motion of elementary carriers of plastic deformation

such as dislocations, disclinations, and grain boundaries.

The simulation of fracture in terms of the temporal theory of strength was performed by the following algorithm. After the temperature, velocity, density, Euler coordinate, internal energy, and pressure have been calculated, the elements of the array of deformation rates were calculated at a new time step at each kth point of the difference scheme:

$$b(k) = (-1)\frac{(p_k(t) - p_k(t - \delta t))}{\delta t}; \qquad (14)$$

then, for points where the pressure proved to be negative ($p_k(t) < 0$), the data obtained were compared with the limiting value of b_*. In this variant of the calculation, the scabbing occurred after b_* reached a limiting value equal to [1]

$$b_* \approx 100\frac{\text{kbar}}{\mu\text{s}}. \qquad (15)$$

In the second variant of the calculation, the fracture condition was assumed to occur when the negative pressure at one of the nodes of the difference mesh $\sigma(k) = (-1)p_k(t)$ reached the limiting value equal to [1]

$$\sigma \approx 50 \text{ kbar}. \qquad (16)$$

If the conditions of the target fracture were fulfilled at the kth node, then the following boundary conditions were set at this node, $p(t, s_k) = p_0$ and $T(t, s_k) = T_0$, and the subsequent calculation was performed only for the nodes from the $(k + 1)$th to the last node.

Figure 3 displays the variation of the velocity of the back surface V_w accessible for the experimental determination as a function of time. Curve *1* corresponds to calculations using Eq. (15); curve *2* was calculated by Eq. (16). It is seen that in the second case fracture occurs later. If the material does not fail, the $V_w(t)$ dependence falls off smoothly (curve *3* in Fig. 3). Note that the presence of scabbing simulated by any of the two models under consideration leads to a discernible difference in the calculated dependence for V_w.

Conditions necessary to check which of the theories is valid can be judged from Fig. 4, which gives the results of numerical simulation of scabbing on the rear surface in a wide range of the pulse parameters t_{pulse}, χ, and q_0. The lower surface corresponds to condition (15); the upper one, to condition (16). It is seen that, at the given values of the tensile strength, there is a significant gap between the surfaces in the whole interval of the pulse parameters. When the pulse parameters fall into this gap, the scabbing or the retention of the integrity of the sample can serve as arguments in favor of this or that physical model of fracture.

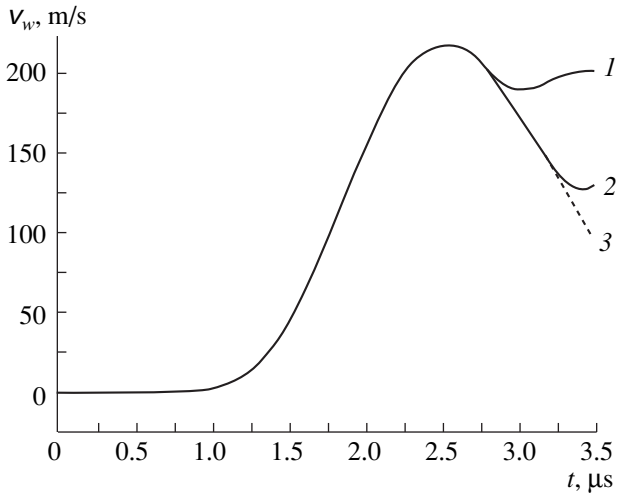

Fig. 3. Time dependence of the velocity of the rear surface of the plate.

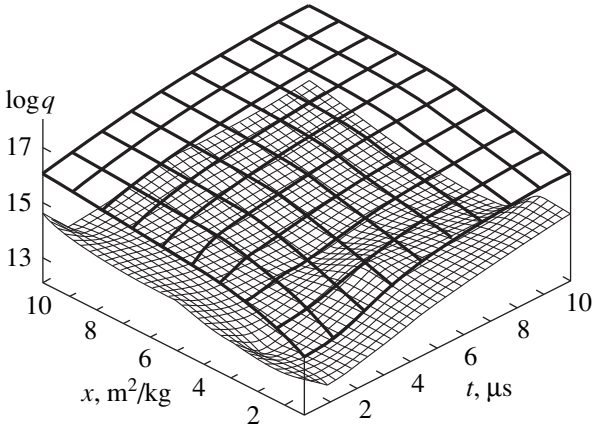

Fig. 4. Surfaces of the rear scabbing.

REFERENCES

1. N. A. Zlatin and B. S. Ioffe, Zh. Tekh. Fiz. **42** (8), 1740 (1972) [Sov. Phys. Tech. Phys. **17**, 1390 (1972)].
2. V. R. Regel', A. I. Slutsker, and É. E. Tomashevskiĭ, *Kinetic Nature of Solid Strength* (Nauka, Moscow, 1974).
3. Yu. I. Meshcheryakov, S. A. Atroshenko, T. V. Balicheva, *et al.*, Preprint No. 24, Leningradskiĭ Filial Inst. Mashinovedeniya im. A.A. Blagonravova, Akad. Nauk SSSR (Leningrad Branch of Blagonravov Institute of Engineering Science, Russian Academy of Sciences, Leningrad, 1989).
4. V. G. Volkov, K. G. Gaĭnulin, S. L. Nedoseev, *et al.*, Vopr. At. Nauki Tekh., Ser. Termoyad. Sintez **1** (7), 36 (1981).
5. N. E. Andreev, M. E. Veĭsman, V. V. Kostin, and V. E. Fortov, Teplofiz. Vys. Temp. **34** (3), 379 (1996).

6. K. S. Dyabilin, M. E. Ledebev, and V. P. Smirnov, Teplofiz. Vys. Temp. **34** (3), 479 (1996).

7. T. A. Andreeva, S. N. Kolgatin, and K. A. Khishchenko, Zh. Tekh. Fiz. **68** (5), 44 (1998) [Tech. Phys. **43**, 518 (1998)].

8. A. V. Bushman, G. I. Kanel', A. L. Ni, and V. E. Fortov, *Thermal Physics and Dynamics of Severe Pulse Actions* (Inst. Khim. Fiz. Akad. Nauk SSSR, Chernogolovka, 1988).

9. Ya. B. Zel'dovich and Yu. P. Raizer, in *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Nauka, Moscow, 1966; Academic, New York, 1966, 1967), Vols. 1, 2.

10. I. M. Bespalov and A. Ya. Polishchuk, Preprint No. 1-257, IVTAN (Institute of High Temperatures, USSR Academy of Sciences, 1988).

11. A. A. Samarskiĭ and Yu. P. Popov, *Difference Methods for Solution of Gas Dynamics Problems* (Nauka, Moscow, 1992).

12. N. A. Zlatin and G. S. Pugachev, Zh. Tekh. Fiz. **43** (9), 1961 (1973) [Sov. Phys. Tech. Phys. **18**, 1235 (1973)].

Translated by S. Gorin

Technical Physics, Vol. 46, No. 2, 2001, pp. 249–253. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 120–124.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Effect of Fullerene-Containing Additives on the Bearing Capacity of Fluoroplastics under Friction

B. M. Ginzburg and D. G. Tochil'nikov

Institute of Problems of Mechanical Engineering, Russian Academy of Sciences, St. Petersburg, 199178 Russia

E-mail: ginzburg@tribol.ipme.ru

Received June 15, 2000

Abstract—A modified procedure for determining the bearing capacity or maximum admissible pressure on a friction unit has been developed. The new feature of the procedure is a test of the wear capacity of a material as a function of pressure. The procedure has been tested on fluoroplastics doped with fullerene-containing soot lubricated with water. F-4 fluoroplastic with 1% fullerene-containing soot added to it shows a 30% higher bearing capacity. *© 2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The subject of this work, which is the bearing capacity of modified fluoroplastics under friction, happens to be at the junction of two branches of science—the physics of strength and polymer materials science. The bearing capacity of a material under friction is, in essence, a test of its strength under more complicated test conditions than simple stretching, compression, bending, and so on. It is, in fact, a shear deformation accompanied by contact interactions, often in the presence of a lubricant (producing a wedging action around microcracks or the Rebinder effect), which does not affect the main features of the phenomenon.

In the pioneering work of Khrushchov [1] a procedure has been developed for determining the maximum admissible pressure $P_{\max}$ on a material, which is run-in as one of the counterfaces in a friction unit. But the actual quantity $P_{\max}$ defined in this procedure is the strength characteristic of the material, which, by analogy with the bearing capacity of a friction unit (defined as the maximum admissible pressure on a friction unit, exceeding which may cause scouring), we call the bearing capacity of a material. Note that there is one more closely related performance characteristic: the maximum admissible operating pressure of a friction unit $P^w_{\max}$. This parameter is usually chosen to be lower than the bearing capacity of a friction unit by some arbitrary quantity so that the wear of a friction unit is small or close to zero at pressures of $P < P^w_{\max}$.

The purpose of this work was to modify the procedure for determining the bearing capacity of materials proposed in [1] to adapt it to new promising materials such as fullerene-containing fluoroplastics.

The procedure for determining $P_{\max}$ proposed in [1] is based on the stepwise loading of a planar specimen pressed against the periphery of a disk rotating at a constant speed (Fig. 1). In the specimen, a crater forms of length U that increases with time. The crater depth is always much less than the disk diameter. Initially, the specimen is loaded with a small load $F = q$. In the beginning, the counterfaces are nonconformal, their contact area is extremely small, and the unit pressure at the contact is very high. As the crater deepens, the pressure drops rapidly and the crater length asymptotically approaches a constant value (Fig. 1a). As soon as the measured crater length stops increasing (the so-called stable crater), a load $F = 2q$ is instantly applied to the friction unit. The crater length first increases rapidly, then the process slows down, and the crater length asymptotically approaches a new stable value. At the moment when the higher load is applied, the pressure P jumps up and then decreases (as the contact area grows due to attrition), finally settling at a higher level.

In a similar way, loads $F = 3q$, $4q$, and so on are applied to the friction unit, and the above process is repeated until, starting with a load $q_{\max} = nq$, the ultimate pressure on the stabilized crater no longer increases at this and still higher loads. Furthermore, the pressure seldom stays at the level reached; more often, it drops with every new load despite further increases in the rate of attrition and the crater length.

In Fig. 1b, a diagram is given of the crater lengths and pressures in U, P coordinates. To a first approximation, the crater area is equal to the area of its projection on the specimen plane; therefore, $F \approx PUb$ (where b is the crater width). Lines F = const in U, P coordinates are hyperbolas (point *1* lies on hyperbola $F = q$, points *1'*, *2* on hyperbola $F = 2q$, and so on). The line linking points *1*, *2*, *3*, etc., is the locus of points corresponding to the lengths of stable craters. Area *A* can be defined as corresponding to friction with low wear; area *B*, to friction with imperfect lubrication (boundary friction); and area *C*, to pressures unattainable for the given friction unit.

1063-7842/01/4602-0249$21.00

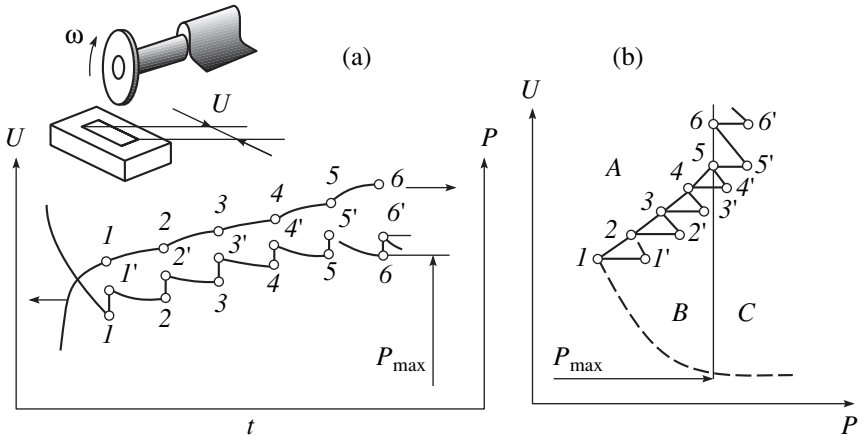

Fig. 1. Schemes of displaying the data used in [1]: (a) presentation of the dependence of the crater length U and pressure P on test duration (or the number of disk rotations); (b) same data in U–P coordinates; in the inset: schematic of the friction unit.

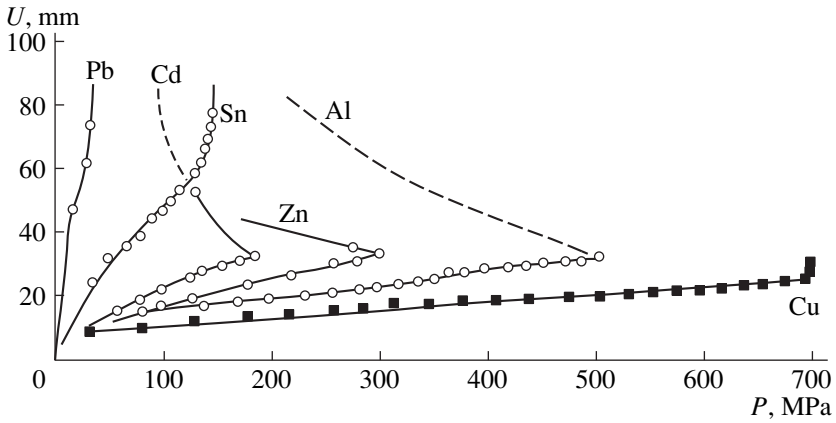

Fig. 2. Test results for six commercial purity metals in [1].

Figure 2 shows the diagrams in U, P coordinates for six commercial purity metals (lead, tin, cadmium, zinc, and copper) lubricated with kerosene [1]. The U–P diagram (magnitude of P_{max}, slope of the initial part of the $U(P)$ curve, the curve behavior at $F > q_{max}$, and so on) proved to be typical for a given friction unit, especially of a metal–lubricant combination. It has been found that the P_{max} values do not correlate with the Brinell hardness.

The existence of P_{max} has as yet no physical explanation. An assumption made in [1] is that the surface temperature of the specimen can rise up to the metal melting point (or the temperature which destroys some protective layer at the surface) for a short time after the application of a load in excess of q_{max}. The local temperature rise was discussed in a number of works and has been observed experimentally [2].

However, our opinion, as well as that of the author of [1], is that such a large temperature rise under abundant lubrication is unlikely to occur. In our studies, a hypothesis has been propounded that shear deformation can lower the melting point of crystal grains at the surface. According to this hypothesis [3], any structural changes in a polycrystalline specimen which hinder the migration of dislocations and the process of shear deformation of crystallites should cause an increase in P_{max}. This approach makes the lack of correlation between P_{max} and the Brinell hardness in metals comprehensible [1].

This experimental technique has a number of drawbacks. The wear particles sinking to the crater bottom can be abrasive; as the crater becomes deeper, the friction of the lateral surfaces against the crater walls becomes more significant, and building the U–P diagram is quite laborious. In this study, an attempt has been made to eliminate the above drawbacks of the technique and apply the modified technique to some new and promising materials. The basic approach is the same as in [1].

EXPERIMENTAL TECHNIQUE

A friction shaft-plane unit (Fig. 1) was used in the tests. This arrangement provides an easy escape for the

wear particles from the tribological contact zone; no lateral friction problems arise as the wear groove width a (which is an analog of the crater length in [1]) increases in size. The tests were carried out in a standard 2070 SMT-1 roller friction machine. The specimen tested was a rectangular plate 10 mm thick clamped in a specially designed holder. The plate was approached from below by a roller of diameter $D = 46$ mm and thickness 16 mm made of wear-resistant chromium-nickel-molybdenum steel of grade 18Kh2NChMA (GOST 4543-71) rotating at $\omega = 400$ min^{-1}, which corresponds to a linear sliding velocity of 1 m/s. The roller width was always larger than the specimen width b. The roller was submerged 2 mm into a 200 ml water bath. Lubrication with water prevented the surface temperature from rising; even at the heaviest loads, the water temperature in the bath did not exceed 50°C, which is too low to cause significant changes of the tribological characteristics of the specimens studied.

The tests were carried out for two experimental setups. In setup I (basic), the specimen and the roller were in contact along a line. The initial load F was 100 N. The friction torque M was continually monitored. Every 300 s, the test was interrupted, the groove width a was measured under an optical microscope, and then the specimen was loaded with a higher load. It was determined beforehand that 300 s is enough time for the groove width and the pressure at the contact to reach a quasi-stable value. The load was increased to a maximum of 1600 N in increments of 200 N. From the measured wear, the groove width nominal pressures in the beginning (P_s) and end (P_e) of the loading step were determined. As distinct from work [1], a $P_e(F)$ dependence was built using the obtained data. The load value beyond which the quantity P_e stopped to grow was assumed to be the maximum admissible load and the corresponding P_e the maximum admissible pressure.

In tests with setup II, relatively small area grooves of various widths were first abraded to determine the linear wear rate I_h at different initial loads. Note that I_h depends not only on pressure, but also on the initial load. Therefore, for the data comparisons to be correct, we carried out tests for the same load $F = 1600$ N, which is the highest of the loads in the range investigated. As will be seen below, in tests with setup II, an additional criterion, the closeness of I_h to zero, could be applied for estimating the maximum admissible pressures, which is especially important in cases where the quantity P_e could not be determined unambiguously from $P_e(F)$ data.

In every version of the tests (kind of setup, fluoroplastic specimen, load and pressure magnitudes), three to eight experiments were carried out. The resulting characteristics represent arithmetic means of the data for the particular version of the test. Relative values of the root-mean-square error for the arithmetic means of tribological characteristics are ~5% for nominal pressures P, ~6% for friction coefficients, and ~10% for the wear intensities. Details of the processing of measurement data and calculations of the tribological characteristics have been given in earlier works [3–5].

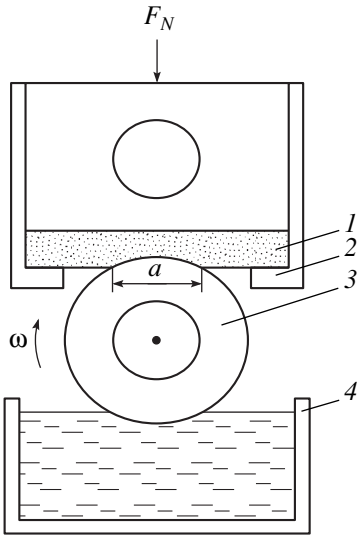

Fig. 3. Schematic of the friction unit test: (*1*) specimen; (*2*) specimen holder; (*3*) rotating steel disk; (*4*) water bath; a is the width of the wear groove.

MATERIALS AND PREPARATION OF SPECIMENS

Materials for the tribological tests were grade F-4 polytetrafluoroethylene (GOST 10007-80) and an F-4K20 composite (TU6-05-1412-76), which is a polytetrafluoroethylene containing 20 wt % of coal coke. All tested materials also contained 1 wt % fullerene soot introduced under the same conditions as the coal coke.

The fullerene soot was produced in the plasma of an electric arc [6]. About 8% of a fullerene mixture (mainly C_{60} and C_{70} in the ratio ~75/25) was extracted from the soot using toluene as a solvent. The soot introduced into the fluoroplastics was that which remained after extraction, and it still contained a considerable amount of higher fullerenes (~12%) [7]. About 1% of fullerene soot was introduced in the material under study under the same technical conditions as when introducing coal coke.

RESULTS AND DISCUSSION

In tests using setup I, it was found that following every increase in the load F, the rate of wear processes taking place in the tribological contact rapidly declines and stabilizes. As a result, the wear groove width a (analog of the crater length in [1]) and depth h initially increase drastically and afterwards, as the contact test goes on, change only slightly. The wear rate of the tribological contact is determined by the pressure, which

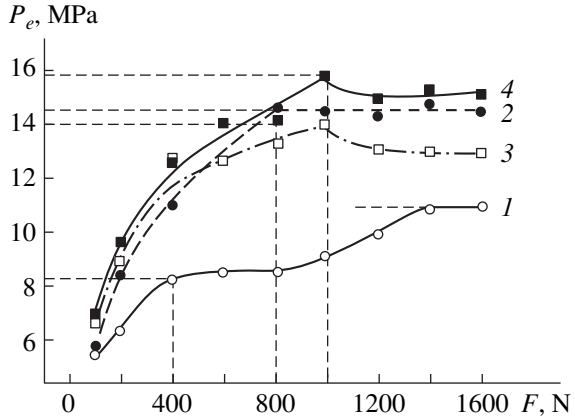

Fig. 4. Dependence of the pressure P_e at the end of test step on the load F applied to the specimen. (*1*) F-4 fluoroplastic; (*2*) F-4 + 1% fullerene soot (F-4SZh1); (*3*) fluoroplastic composite F-4K20; (*4*) F-4K20 + 1% fullerene soot (F-4K20SZh1).

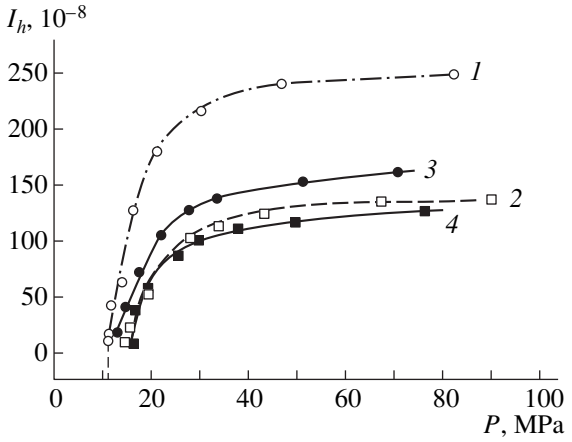

Fig. 5. Dependence of the linear wear rate I_h on pressure in the contact: (*1*) F-4; (*2*) F-4SZh1; (*3*) F-4K20; (*4*) F-4K20SZh1.

depends on the contact area. Alongside the increase and subsequent stabilization of the wear groove depth and the tribological contact area, a rapid decrease and leveling off of the pressure in the tribological contact occurs. Thus, the growth of the tribological contact area and lowering of pressure are interrelated; their rates and ultimate values depend on the load on the contact and the wear groove width, the latter being, in turn, the function of the wear resistance of the material.

In Fig. 4, $P_e(F)$ curves for the fluoroplastic specimens are shown. It is seen that P_e initially rises as F is increased and that, at F values of 800–1000 N, P_e levels off at different levels.

One exception is the starting F-4 fluoroplastic, for which two stable pressure levels have been observed: P_e = 8 MPa at F = 400 N and P_e = 11 MPa at F = 1400 N. But, as shown below, low wear is observed at pressures of $P_e < 11$ MPa, so this value should be taken as the maximum admissible value.

For the F-4K20 fluoroplastic containing 20% carbon coke, the admissible pressure is higher (~14 MPa), possibly due to significant transformation of the deformation mechanism of the polytetrafluoroethylene crystallites. The presence of numerous defects due to the coke impurity makes the dislocation mechanism of shear less effective.

It is remarkable that the introduction of just 1% soot produces an even slightly greater rise of the admissible pressure to P_e = 15 MPa. The cause, in our opinion, is the high chemical activity of fullerenes, which is particularly seen in the formation of a fullerene-polymer network on the friction surfaces [5].

At least two mechanisms of the growth of P_e involving fullerenes can be suggested. According to one, fullerenes penetrate only amorphous regions of the amorphous-crystalline network where its points are located. According to the other, the fullerenes penetrate the crystallites, which induces high concentrations of dislocations in them, thus impeding shear deformations of the crystallites and of the material as a whole. Which of the two mechanisms is applicable can be determined using different structural methods, and such an attempt will be undertaken in our next study. In a recent report [8], a simple technique for identifying the easiest slip planes in polymer crystallites was proposed. This technique can be used for estimating the effect of fullerenes on the susceptibility of polymer crystallites to shear deformation.

Introduction into F-4K20 of the fullerene soot strengthens this material still further: the admissible pressure rises to 16 MPa.

In Fig. 5 the results of the tests for setup II are shown. The introduction of 1% fullerene soot resulted in a significant reduction of the linear wear rate for the same pressures: by a factor of 2–3 in F-4 (specimen F-4SZh1) and a factor of 1.3–1.7 in F-4K20 (specimen F-4K20SZh1). In this figure, it is seen that at $P < P_{max}$ the quantity I_h approaches zero. Using the data in Fig. 5, an unambiguous choice of P_e value for the F-4 fluoroplastic could be made. Note that, in contrast to metals, pressures in excess of P_{max} are quite feasible, but the wear rates will be higher. Therefore, operating pressures for a material can be only those below P_{max}.

CONCLUSIONS

To conclude, a modified procedure for determining the bearing capacity of a material in a friction unit has been proposed. In addition to the dependence of pressure on the applied load, this procedure yields data on the linear wear rate as a function of pressure. The range of admissible pressures for a material or its bearing capacity is defined as a range in which the pressure is

not rising (with increasing load) and the wear is not high.

REFERENCES

1. M. M. Khrushchev, *Run-In of Bearing Alloys and Trunnions* (Akad. Nauk SSSR, Moscow, 1946).
2. *Contact Interaction of Solid States and Calculation of Friction and Deterioration Forces: Collection of Articles*, Ed. by A. Yu. Ishlinskiĭ and N. B. Demkin (Nauka, Moscow, 1971).
3. B. M. Ginzburg, Pis'ma Zh. Tekh. Fiz. **26** (8), 65 (2000) [Tech. Phys. Lett. **26**, 345 (2000)].
4. Yu. P. Kozyrev, B. M. Ginzburg, N. D. Priemskii, *et al.*, Wear **171**, 71 (1994).
5. B. M. Ginzburg, M. V. Baĭdakova, O. F. Kireenko, *et al.*, Zh. Tekh. Fiz. **70** (12), 87 (2000) [Tech. Phys. **45**, 1595 (2000)].
6. W. Krätschmer and D. R. Huffman, Philos. Trans. R. Soc. London, Ser. A **343** (1667), 33 (1993).
7. F. Beer, A. Gügel, K. Martin, *et al.*, J. Mater. Chem. **7** (8), 1327 (1997).
8. B. M. Ginzburg and N. Sultanov, Zh. Tekh. Fiz. **71** (2), 129 (2001) [Tech. Phys. **46**, 258 (2001)].

Translated by B. Kalinin

Technical Physics, Vol. 46, No. 2, 2001, pp. 254–257. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 125–128.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Parameter Optimization of a Silicon Pressure-Sensitive Element Based on Dual-Drain MIS Transistors

G. G. Babichev, S. I. Kozlovskiĭ, V. A. Romanov, and N. N. Sharan

Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, Kiev, 252650 Ukraine

Received April 18, 2000

Abstract—The object of investigation was an integrated pressure transducer with a differential pressure-sensitive element based on two parallel-connected *p*-channel dual-drain MIS transistors. The transistors were made on the planar side of a silicon membrane with the rigid central region near the edges of membrane's thin part (valley). The optimum design of the transducer and its basic characteristics were determined. © *2001 MAIK "Nauka/Interperiodica".*

The progressively growing demand for semiconductor pressure transducers [1] stimulates a search for new types of high-sensitivity elements that combine low power consumption and low intrinsic noise. In the previous work [2], we studied the sensitive elements based on silicon field-effect dual-drain pressure-sensitive transistors subjected to uniform uniaxial elastic mechanical strain.

In this work, we consider a more general case when the strain varies linearly with the MIS transistor (MIST) channel length. Such a distribution of elastic strain occurs in the valley of a specially shaped rigid-center membrane upon membrane loading. It has been shown [3, 4] that this membrane, combined with a differential sensitive element, decreases the effect of "parasitic" mechanical stresses due to transducer packaging. Moreover, it also provides a linear output characteristic of high-sensitivity thin-membrane transducers, in which the bending of the membrane upon loading is comparable to its thickness.

In what follows, we will optimize the design of the pressure transducer with a differential sensitive element built around two parallel-connected *p*-channel dual-drain MISTs. The schematics of the silicon membrane and the crystallographic orientation of the transistor channels are given in Fig. 1. The pressure-sensitive transistors are placed on the planar side of the membrane near the edges of its thin part (valley). When the membrane is loaded uniformly and the transistors are placed symmetrically about the plane passing through the center of the valley so that it is parallel to the edges of the valley and perpendicular to its plane, elastic mechanical stresses in the vicinity of the devices are equal in magnitude and opposite in sign [3, 4].

Figure 2 shows the layout and the connection diagram of the sensitive element. With appropriate bias voltages (U_0, U_G) applied and in the absence of pressure, the drain currents of transistors T_1 and T_2 are the same, since their layout and geometries are identical; hence, the output signal U_{out} (Fig. 2) is zero provided that load resistances are equal: $R_{D1} = R_{D2} = R_D$.

When pressure is applied, the membrane converts the load distributed over the surface to uniaxial elastic tensile (compressive) strain in the regions where the transistors are located. Because of this, the mobility μ_p

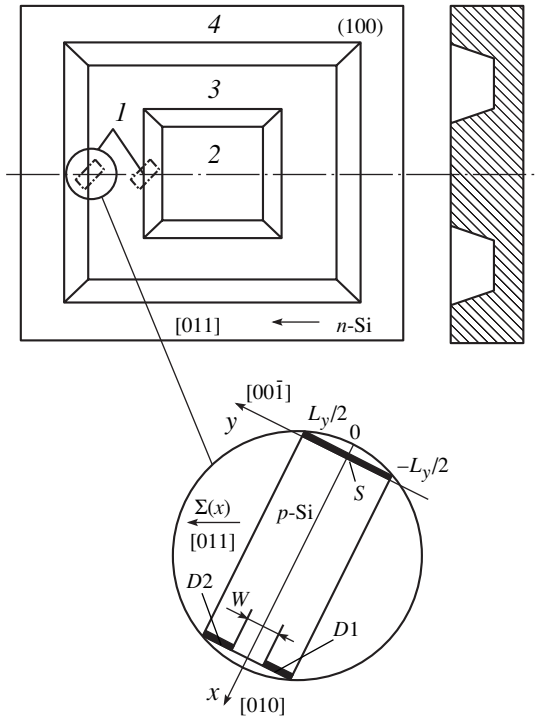

Fig. 1. Arrangement of the sensitive element on the membrane and the crystallographic orientation of the transistor channels: *1*, dual-drain MIS strain-sensitive transistors (*S*, source; *D*1 and *D*2, drains); *2*, rigid central island; *3*, valley; and *4*, membrane support.

1063-7842/01/4602-0254$21.00 © 2001 MAIK "Nauka/Interperiodica"

of majority carriers (holes) becomes anisotropic and additional strain-induced transverse and longitudinal electric fields arise in the FET channel [2]. They cause current disbalance in the drain circuits, as a result of which the output current appears.

Below, we calculate a strain-induced potential change in the MIST channel and the relative, S_R, and absolute, S_A, sensitivities of the transducer (S_R is also called the conversion efficiency).

Let the MIST channel be bounded by the regions

$$0 \le x \le L_x, \quad -L_y/2 \le y \le L_y/2, \quad 0 \le z \le L_z. \tag{1}$$

The potential will be calculated under the following assumptions and simplifications.

(i) The conductivity in the xy plane is anisotropic due to uniaxial elastic mechanical stress Σ aligned with the [110] direction (Fig. 1). The $\Sigma(x)$ dependence has the form [3, 4]

$$\Sigma(x) = \Sigma_0(k_0 + k_1 x/L_x), \tag{2}$$

where $|k_0|$, $|k_1| \le 1$ and $k_0\Sigma_0$ is the mechanical stress at the point $x = 0$.

(ii) Elastic mechanical stresses and the anisotropy parameter [2]

$$a(x) = \Pi_{44}\Sigma_0(k_0 + k_1 x/L_x)/2 = a_0(k_0 + k_1 x/L_x) \tag{3}$$

(Π_{44} is the piezoresistive coefficient of shear for p-Si) are sufficiently low. Hence, the inequality $\max|a(x)| \ll 1$ is valid (in calculations, a_0 will be set equal to 0.06) and the conductivity tensor components can be written as

$$\sigma_{xx} = \sigma_{yy} \cong \sigma(x), \quad \sigma_{xy} = \sigma_{yx} = \sigma(x)a_x. \tag{4}$$

(iii) The transistor channel is relatively thin (L_x, $L_y \gg L_z$), and its conductivity in the smooth approximation [5, 6] is given by

$$\sigma(x) = \frac{\mu_p C}{L_z}\sqrt{(U_C - U_G)^2 - 2\left(U_C - U_G - \frac{U_{DS}}{2}\right)\frac{U_{DS}x}{L_x}}, \tag{5}$$

where C is the capacity per unit area of the gate–semiconductor structure, U_{DS} is the potential difference between the source and the drain, U_C is the cutoff voltage (gate voltage at which the channel conductivity vanishes), and U_G is the gate voltage. We assume that $U_{DS} < U_C - U_G$ (U_{DS}, U_C, $U_G > 0$) and the transistor is in the on state.

In view of the above simplifications, the potential distribution $\varphi(x, y)$ in the channel can be considered as quasi-planar; then, in the first approximation with respect to the anisotropy parameter, the spatial distribution of the potential is expressed as

$$\frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial \ln(\sigma(x))}{\partial x}\frac{\partial\varphi}{\partial x} = 0 \tag{6}$$

with the boundary conditions

$$y = \pm\frac{L_y}{2}, \quad a(x)\frac{\partial\varphi}{\partial x} + \frac{\partial\varphi}{\partial y} = 0, \tag{7}$$

$$x = 0, L_x \quad \varphi(0, y) = 0, \quad \varphi(L_x, y) = U_{DS}. \tag{8}$$

Equations (6)–(8) are solved by the Fourier method:

$$\varphi(x, y) = U_C - U_G - \sqrt{(U_C - U_G)^2 - 2\left(U_C - U_G - \frac{U_{DS}}{2}\right)\frac{U_{DS}x}{L_x}} + \Phi(x, y), \tag{9}$$

where

$$\Phi(x, y) = \frac{2a_0}{\pi}\left(U_{DS} - \frac{U_{DS}^2}{2(U_C - U_G)}\right) \times \sum_{n=1}^{\infty}\frac{F_n(k, U_C, U_G, U_{DS})}{n\cosh(\pi n L_y/2L_x)}\sinh\left(\frac{\pi n y}{L_x}\right)\sin\left(\frac{\pi n x}{L_x}\right), \tag{10}$$

$$F_n(k, U_C, U_G, U_{DS}) = \int_0^1 \frac{(k_0 + k_1 z)\sin(\pi n z)}{\sqrt{1 - [2(U_C - U_G) - U_{DS}]U_{DS}z/(U_C - U_G)^2}}dz. \tag{11}$$

Figure 3 depicts the distribution of the transverse potential difference $\Delta\varphi(x)$ along the channel when the anisotropy parameter depends in various ways on the coordinate x:

$$\Delta\varphi(x) = \varphi(x, L_y/2) - \varphi(z, -L_y/2) = \Phi(x, L_y/2) - \Phi(x, -L_y/2). \tag{12}$$

The form of the anisotropy parameter is seen to considerably affect the longitudinal distribution of $\Delta\varphi(x)$. Note that a problem similar to Eqs. (6)–(8) was numerically solved for the case when conductivity anisotropy

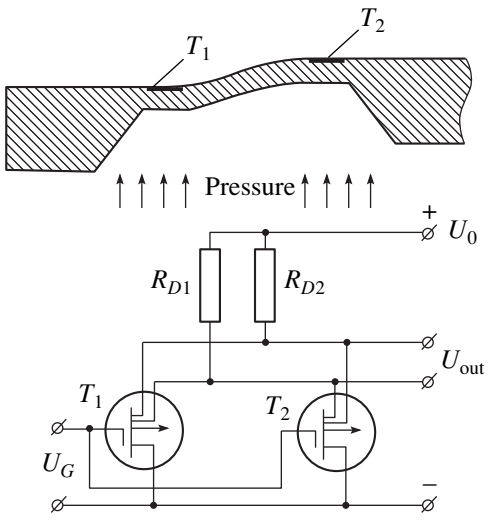

Fig. 2. Layout and connection diagram of the transistors.

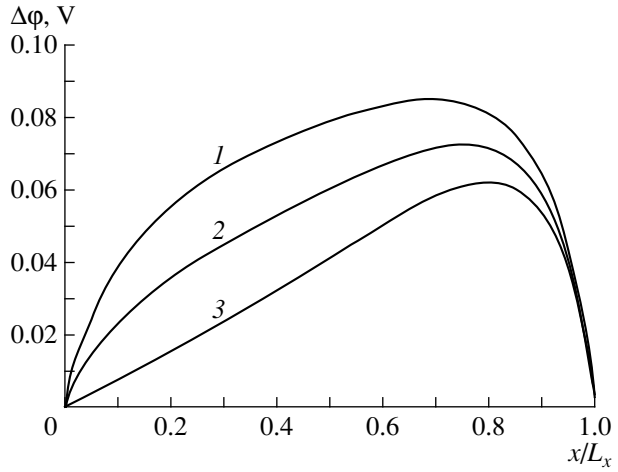

Fig. 3. Distribution of the transverse potential difference along the channel. U_{DS} = 3 V and $U_C - U_G$ = 5 V. Coefficients k_0 and k_1 in the expression for the anisotropy parameter $a(x) = a_0(k_0 + k_1 x/L_x)$ are (*1*) $k_0 = 1$, $k_1 = 0$; (*2*) $k_0 = 0$, $k_1 = 1$; and (*3*) $k_0 = k_1 = 0.5$.

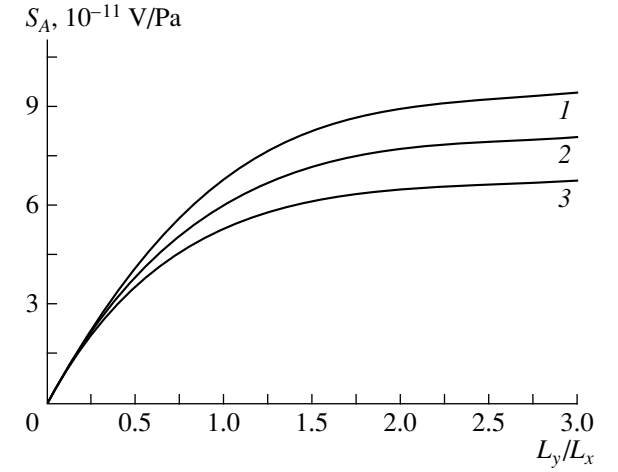

Fig. 4. S_A as a function of the channel geometry. (*1*)–(*3*) the same as in Fig. 3.

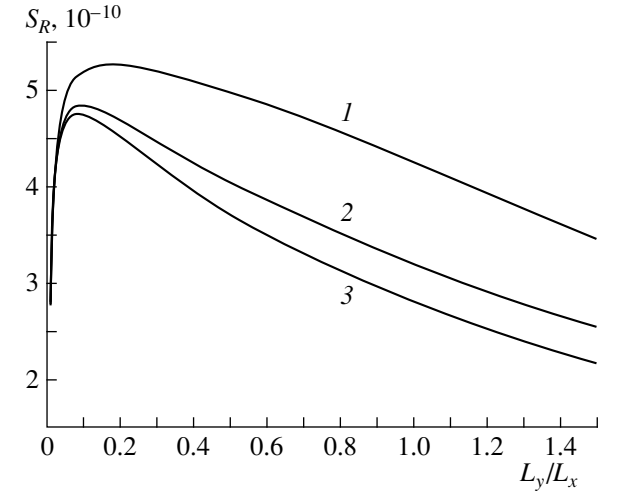

Fig. 5. S_R as a function of the channel geometry. (*1*)–(*3*) The same as in Fig. 3.

was induced by an external magnetic field. The $\Delta\varphi(x)$ curve for uniform strain ($k_0 \neq 0$, $k_1 = 0$) agrees well with the numerically calculated distribution of Hall voltage in the channel of an MIS magnetotransistor [7].

Now let us calculate the conversion efficiency S_R and the absolute sensitivity S_A of the pressure transducer. Let the drains $D1$ and $D2$ be separated by a gap W (Fig. 1) and have the same width $L_D = (L_y - W)/2$. The values of S_A and S_R are found by integration [2]:

$$S_A = 2C\mu_p R_D (U_C - U_G - U_{DS}) \int_{-L_y/2}^{-W/2} \frac{\partial \Phi}{\partial x}(L_x, x)\,dy, \quad (13)$$

$$S_R = C\mu_p \frac{(U_C - U_G - U_{DS})}{2I_D^0} \int_{-L_y/2}^{-W/2} \frac{\partial \Phi}{\partial x}(L_x, y)\,dy. \quad (14)$$

Here, $I_{D1}^0 = I_{D2}^0 = I_D^0$ are the drain currents of transistors T_1 and T_2 in the absence of strain:

$$I_D^0 = \mu_p C \frac{L_D}{L_x}\left(U_C - U_G - \frac{U_{DS}}{2}\right)U_{DS}. \quad (15)$$

Integrating Eq. (13) and (14) in view of Eq. (10) and (11) yields

$$S_R = \frac{\Pi_{44}}{2}\frac{(U_C - U_G - U_{DS})L_x}{(U_C - U_G)L_D}\Psi, \quad (16)$$

$$S_A = 2\Pi_{44} I_D^0 R_D \frac{(U_C - U_G - U_{DS})L_x}{(U_C - U_G)L_D}\Psi, \quad (17)$$

where

$$\Psi = \frac{2}{\pi}\sum_{n=1}^{\infty} \frac{(-1)^n F_n(k, U_C, U_G, U_{DS})}{n} \times \left[\frac{\cosh(\pi n W/2L_x)}{\cosh(\pi n L_y/2L_x)} - 1\right]. \quad (18)$$

The absolute and relative sensitivities of the strain-sensitive transistor as functions of the channel geometry are shown in Figs. 4 and 5. The run of the S_A vs. L_y/L_x curve is almost the same for the uniform and nonuniform distributions of mechanical strain (Fig. 4): S_A monotonically increases and saturates at $L_y/L_x \geq 2$. For $k_1 = 0$, this is consistent with the similar analytical dependence for an MIS magnetotransistor [7]. The dependence of S_R on L_y/L_x (Fig. 5) features a peak at $L_y/L_x = 0.1$–0.2 for the nonuniform case ($k_1 \neq 0$). From the curves represented in Figs. 4 and 5, one can find the optimal value of L_y/L_x at which S_R and S_A are maximal.

Figures 6 and 7 depict analytical dependences of S_A on the drain–source, U_{DS}, and gate–source, U_G, potential differences for $CR_D\mu_p = 10^{-2}$. As U_{DS} grows (Fig. 6), the absolute sensitivity linearly increases and peaks at

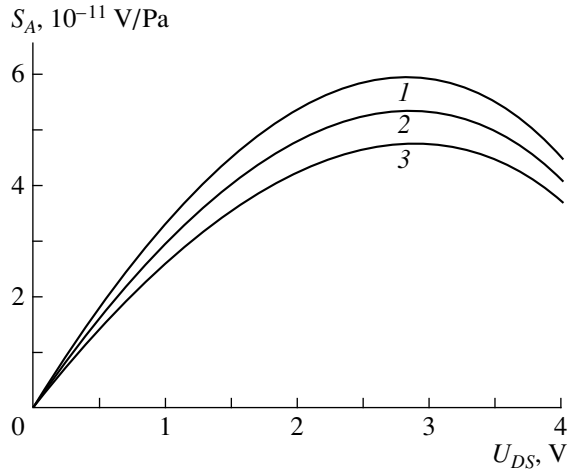

Fig. 6. S_A vs. U_{DS} for $U_C - U_G = 5$ V.

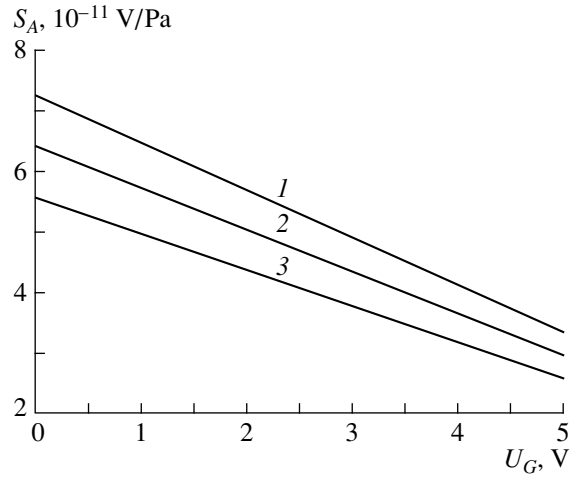

Fig. 7. S_A vs. U_G for $U_C = 10$ V and $U_{DS} = 3$ V.

$U_{DS} \approx (U_C - U_G)/2$. The run of curves *1–3* to a great extent is defined by the dependence $I_D^0(U_{DS})$ [expressions (15) and (17)]. With an increase in U_G, S_A drops practically linearly (Fig. 7, curves *1–3*). As follows from formula (16), S_R is virtually independent of U_{DS} and U_G.

To conclude, consider briefly one more basic property of a pressure transducer: linearity of the load characteristic. It is known [3, 4, 8] that possible reasons for the nonlinearity of piezoresistive membrane transducers are nonlinear conversion of pressure to elastic strain by that part of the membrane where the sensitive element is located and also nonlinearity of the piezoresistive effect. Experimental and theoretical studies [3, 4, 8] suggest that the nonlinearity of the load characteristic due to the above reasons is greatly reduced by the differential design of the sensitive element, where one of the transistors is under compression and the other, under stretching.

Thus, the design of the pressure transducer proposed in this work increases the absolute sensitivity twofold and provides a better linearity of the load characteristic compared to the single-transistor design [2].

REFERENCES

1. J. Bryzek, Sens. Actuators A **56**, 1 (1996).
2. G. G. Babichev, S. I. Kozlovskiĭ, and N. N. Sharan, Zh. Tekh. Fiz. **70** (10), 45 (2000) [Tech. Phys. **45**, 1276 (2000)].
3. A. Yasukawa, M. Shimazoe, and Y. Matsuoka, IEEE Trans. Electron Devices **ED-36** (7), 1295 (1989).
4. M. Nishihara, K. Yamada, and Y. Matsuoka, Hitachi Rev. **30** (6), 285 (1981).
5. I. M. Vikulin and V. I. Stafeev, *Physics of Semiconductor Devices* (Radio i Svyaz', Moscow, 1990).
6. R. S. Muller and T. Kamins, *Device Electronics for Integrated Circuits* (Wiley, New York, 1986; Mir, Moscow, 1989).
7. P. W. Fry and S. F. Hoey, IEEE Trans. Electron Devices **ED-16** (1), 35 (1969).
8. K. Suzuki, T. Ishihara, M. Hirata, and H. Tanigawa, IEEE Trans. Electron Devices **ED-34** (6), 1360 (1987).

Translated by V. Isaakyan

Technical Physics, Vol. 46, No. 2, 2001, pp. 258–261. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 129–132.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

X-ray Diffraction Determination of Easiest Slip Planes in Amorphous–Crystalline Polymers

B. M. Ginzburg* and N. Sultanov**

**Institute of Problems in Machine Science, Russian Academy of Sciences, Vasil'evskiĭ Ostrov, Bol'shoĭ pr. 61, St. Petersburg, 199178 Russia*

***Tajik State University, ul. Rudaki 17, Dushanbe, 734000 Tajikistan*

e-mail: ginzburg@tribol.ipme.ru

Received May 15, 2000

Abstract—A simple X-ray diffraction technique for determining the easiest slip planes in amorphous–crystalline polymer grains is suggested. The efficiency of the technique was demonstrated with polyethylene and polyamide-6. © *2001 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Determination of the easiest slip planes in solids is a complex, while well-established, procedure. Usually, bulk single crystals are deformed and then examined with electron microscopy, as well as electron and X-ray diffraction methods [1–4]. In the case of polymers, many small single crystals grown in dilute solutions are settled out on plastic substrates that are subsequently deformed [4].

It would be logical to assume that the easiest slip planes observed in polymer single crystals are also present in mosaic polycrystals; however, this statement is in doubt. Moreover, in polymer crystals, the easiest slip planes usually coincide with planes of chain folding, and the direction of the latter depends on growth conditions for both bulk single crystals and single-crystal grains. At least, folding planes in the crystals and grains may differ. Therefore, determination of slip planes (and the easiest slip planes, in particular) in polycrystalline polymer materials seems to be a topical problem, since this is necessary for a better understanding of deformation mechanisms in amorphous–crystalline polymers.

In this work, we present a simple technique for finding the easiest slip planes in amorphous–crystalline polymers.

EXPERIMENTAL

First, a sufficiently thick highly oriented amorphous–crystalline film of a crystallizable polymer is formed by any one of the preparation procedures. Its thickness, dictated by transmission X-ray diffraction analysis, usually ranges from 0.1 to 2 mm. Thick films are structurally more uniform; in addition, substrate effects are avoided in this case. Highly oriented crystallites are prepared, for example, by orientation drawing at temperatures sufficient for crystallization or recrystallization processes to occur [4]. Macromolecules in the crystallites are then oriented largely along the drawing direction. If the film is amorphized or does not crystallize during the orientation process (for example, if orientation drawing is carried out at temperatures below the glass-transition temperature), it can be annealed. The annealing conditions should be selected in such a way as to preserve the preferential orientation of the molecules along the drawing direction (e.g., annealing can be performed at a fixed length of the specimen). The need for initially thick films may show up at this stage as well, since the film geometry must prevent plane texturing of the crystallites. It is desirable that the film texture be axial relative to the texture axis, which coincides with the direction of macromolecules in the crystallites. When an X-ray beam is incident normal to the drawing direction, a large-angle diffraction pattern for such films is similar to a rotating-crystal pattern and exhibits a set of curved reflections on the layer lines (primarily on the zero layer line, or equator). The equatorial reflections are those from crystallographic planes parallel to the axes of macromolecules. It is just along these planes that share strain in polymer crystals takes place [4].

Once the highly oriented amorphous–crystalline film has been prepared, the next orientation drawing (reorientation) is carried out at a large angle (from 20° to 70°; 45° seems to be the best) to the direction of the first drawing. To make this procedure easier, a piece is cut out of the initial specimen in such a way that its long sides make the given angle with the first drawing direction (Fig. 1a).

During the second drawing, diffraction patterns are taken from strained specimens from time to time. The X-ray beam remains perpendicular to the specimen all the time. A typical change in the large-angle diffraction pattern is demonstrated in Figs. 1b and 1c. The diffraction pattern turns toward the direction of the second

1063-7842/01/4602-0258$21.00 © 2001 MAIK "Nauka/Interperiodica"

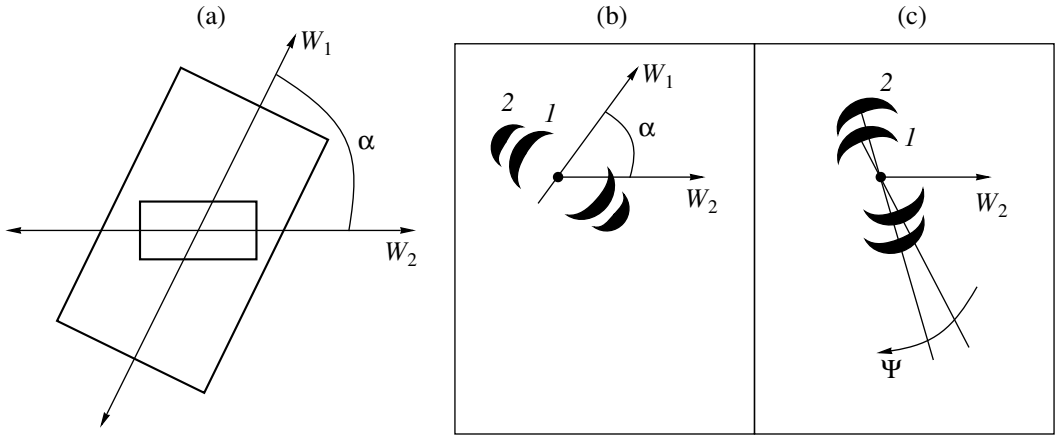

Fig. 1. (a) Specimen preparation: W_1, direction of the first orientation drawing of the polymer film at a temperature T_1; W_2, direction of the second orientation drawing at T_2; and α, angle between these directions. (b) Large-angle X-ray diffraction pattern for the initial specimen (after the first drawing); *1* and *2* are equatorial reflections. (c) Large-angle X-ray diffraction pattern for the specimen stretched in the direction W_2; the reflections are turned toward this direction, reflection *2* from the easiest slip planes being ahead of reflection *1*.

drawing, but the equatorial reflections from the easiest slip planes are ahead of those from other planes: the former are rotated through a larger angle toward the second drawing axis. Note that such a criterion for determination of the easiest slip planes is valid not only for reversible but also for low plastic strains.

In addition, slip, or shear strain, inside the crystallites may lead to radial broadening of reflections from the slip planes; this effect is, however, noticeable only for intense reflections. Moreover, it shows up only when reorientation occurs in a system of highly oriented crystallites where the macromolecule axis is aligned with the texture axis. If these axes do not coincide or the angle of reorientation is close to right angle, shear strain of the crystallites may cause the reverse effect: radial narrowing of the reflections [5].

RESULTS AND DISCUSSION

Polyethylene. First, we applied our technique to low-density polyethylene. Grains of commercial polyethylene of molecular weight $M = 25 \times 10^3$ were used to mold 1.4-mm-thick films. The operating temperature and pressure were 200°C and 5 MPa, respectively. The films were quenched in room-temperature water. Then, they underwent uniaxial stretching to an elongation by a factor of 5 (first orientation) at $T_1 = 85$°C and to an elongation by a factor of 9 at $T_1 = 20$°C. Next, specimens cut out of the films at various angles α to the first orientation direction (Fig. 1a) were subjected to step strain (second orientation) at temperatures $T_2 = 20$ and 85°C in a special chamber inside the X-ray chamber. After each step, the stressed specimens were held for 12 h and large-angle diffraction patterns were recorded on a flat X-ray film. The X-ray beam was directed normally to the X-ray film plane. Ni-filtered CuK_α radiation was used.

Figure 2 shows a typical large-angle diffraction pattern obtained under recoverable strain (~50%) in the experiments with $T_1/T_2 = 85/20$ for $\alpha = 75°$. The 200 reflection is an angle Ψ ahead of the 110 reflection when they rotate toward the vertical direction of a new equator. Like 110, other equatorial reflections, *hk*0 and 0*k*0, are also left behind. The same is true for the 100 reflection of the monoclinic modification of the polyethylene (the closest to the center of the pattern). The last-named reflection is likely to appear because of martensitic transformation. The azimuth angle of

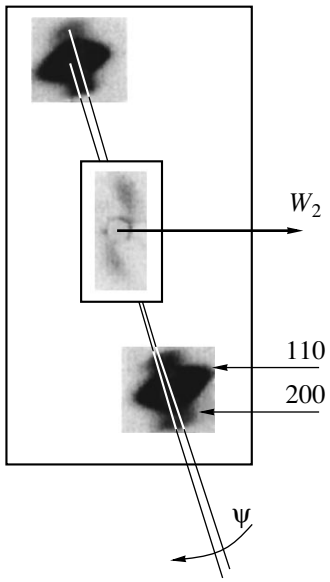

Fig. 2. Large-angle X-ray diffraction pattern for the polyethylene specimen after the first drawing at $T_1 = 85$°C and after uniaxial elastic stretching (~50%) at an angle $\alpha = 75°$ to the first drawing direction for $T_2 = 20$°C. At the center, the associated low-angle pattern is shown.

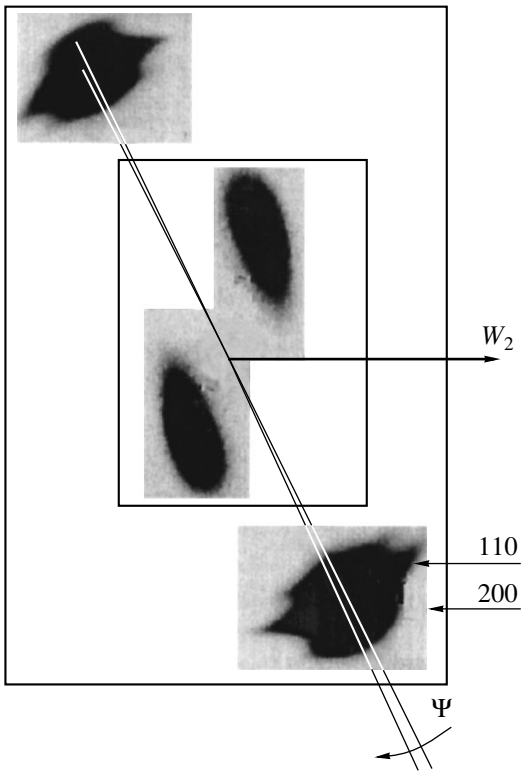

Fig. 3. The same as in Fig. 2 for the specimen drawn at T_1 = 20°C and stretched to 50% at T_2 = 20°C for α = 45°. The residual strain after load removal is 15%.

advance Ψ is small, no more than several degrees. This means that slipping along other paratropic planes proceeds with almost the same easiness; specifically, it obviously proceeds along the (110) planes.

The (110) reflection is the most intense; therefore, its radial broadening is seen immediately from the diffraction patterns without any measurements. Under elastic strain, this broadening is known to reversibly change [6]. It is associated with shear strain of the crystallites due to intracrystallite slipping. The sign of the shear strain is the distorted low-angle diffraction pattern at the center of Fig. 2.

Similar results were obtained for the other temperature conditions of the first and second drawings: T_1/T_2 = 20/20 and 85/85. In Fig. 3, a large-angle pattern for the 20/20 series is depicted; the strain of the specimen, 50%, is not fully recoverable: after load removal, the residual strain was ~15%. However, in this case, too, the 200 reflection leads and the angle of advance Ψ is small. The low-angle diffraction pattern at the center of Fig. 3 is an indication of the shear strain in the crystallites.

In polyethylene, the (110) plane is the most closely packed. Usually, it is also the easiest slip plane. In our case, however, the easiest slip plane is (200), second in packing density [7]. It appears that here the (200) plane is the folding plane, which in many ways explains the results observed.

Thus, in highly oriented mosaic amorphous–crystalline polyethylene, at least two slip systems, (200)[001] and (110)[001], are present, the former being the system of easiest slip. This is consistent with the literature data [7].

For right-angle reorientation (α = 90°), one might expect the (200)[020] slip system; however, this system was not revealed. The crystallites first were at short-term unstable equilibrium and then were separated into two sets with the (200)[001] and (110)[001] slip systems.

Polyamide-6. In polyamide-6, the macromolecule axes are aligned with the *b* axis. Intermolecular interaction in polyamide-6 is known to strongly depend on the crystallographic direction because of hydrogen bonds

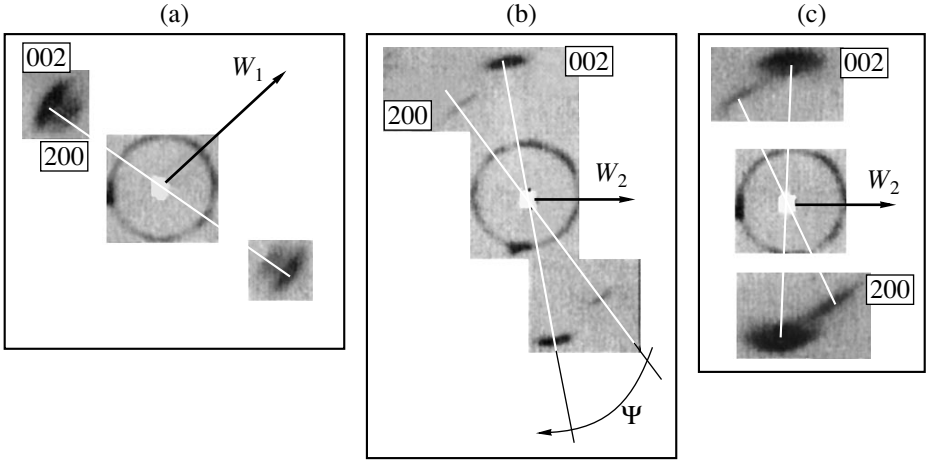

Fig. 4. Large-angle X-ray diffraction pattern for the oriented polyamide-6 film annealed at 200°C and then uniaxially stretched (second drawing) at an angle α = 45° to the initial orientation at T_2 = 20°C; the second drawing was performed in the horizontal direction. The strain ε = (a) 0, (b) 22% (fully recoverable strain), and (c) 300% (irreversible strain to prerupture with necking).

formed in the (002) planes. Slipping transverse to these planes is therefore greatly hindered in polyamide-6.

Let us see how this well-known situation is reflected in experiments using our approach. Polyamide-6 specimens were prepared by annealing (2 h, 200°C) oriented commercial PK-4 films of fixed length. Further procedures were the same as with the polyethylene.

Figure 4 demonstrates large-angle diffraction patterns obtained after reorientation under room conditions for $\alpha = 45°$. When the strain was recoverable (~22%), the angles of rotation for the equatorial reflections drastically differed: the 002 reflections turned through nearly 30°, while the positions of the 200 reflections remained almost unchanged. Hence, the easiest-slip plane is (002).

Two points seem to be noteworthy in our opinion. First, note that the 200 reflections turn insignificantly even at a subsequent plastic strain of 300%, when the 002 reflections are set on a new equator. Second, the 002 reflections markedly broaden in the radial direction. In the case of plastic strain, this may be associated with melting of initial crystallites and the nucleation of new, smaller, ones; for elastic strain, this indicates the presence of shear strain in the crystallites.

Thus, our approach was tested on two polymers with extremely different intermolecular interactions. It can be assumed that the vast majority of other crystallizable polymers are close to those studied or lie in between by intermolecular interaction force. It is therefore hoped that our technique for determination of easiest slip planes in mosaic amorphous–crystalline polymers will be applied to studying other crystallizable polymers.

REFERENCES

1. E. Schmid and W. Boas, *Kristallplastizät, mit besonderer berücksichtigung der Metalle* (J. Springer, Berlin, 1935; Gos. Ob"edinennoe Nauchno-Tekhnicheskoe Izd. NKTP SSSR, Moscow, 1938).
2. R. Berner and H. Kronmüller, *Plastische Verformiung von Einkristallen*, in *Moderne Probleme der Metalphysik*, Ed. by A. Seeger (Springer-Verlag, Berlin, 1965; Mir, Moscow, 1969).
3. B. I. Smirnov, *Dislocation Structure and Hardening of Crystals* (Nauka, Leningrad, 1981).
4. B. Wunderlich, *Macromolecular Physics*, Vol. 1: *Crystal Structure, Morphology, Defects* (Academic, New York, 1973; Mir, Moscow, 1976).
5. B. M. Ginzburg, V. N. Baranov, V. I. Gromov, and K. B. Kurbanov, Mekh. Polim., No. 1, 3 (1974).
6. B. M. Ginzburg, N. Sultanov, and D. Rashidov, J. Macromol. Sci., Phys. **9** (4), 609 (1974).
7. V. A. Marikhin and L. P. Myasnikova, *Supramolecular Structure of Polymers* (Khimiya, Leningrad, 1977).

Translated by V. Isaakyan

Technical Physics, Vol. 46, No. 2, 2001, pp. 262–265. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 133–136.

BRIEF COMMUNICATIONS

Luminescence from Mn-Doped PbI_2 Crystals

I. M. Kravchuk*, S. S. Novosad, and I. S. Novosad****

** Lviv Technologies State University, Lviv, 79013 Ukraine*

*** Franko National University, Lviv, 79000 Ukraine*

Received January 21, 2000

Abstract—Luminescent properties of PbI_2 and PbI_2 : 0.5 mol % $MnCl_2$ crystals under X-ray or N_2-laser excitation are studied experimentally. The measurements are performed at temperatures ranging from 85 to 295 K. For PbI_2 crystals under laser excitation, spectral bands with peaks near 495 and 512 nm, respectively, are observed at 85 K. With X-ray excitation at the same temperature, luminescence is observed in the 515- and 715-nm bands. The doping decreases the intensity in the 515-nm band, increases it for longer wavelengths, and shifts the highest peak to 700 nm. At 85 K, the doping has an insignificant effect on the excitation energy accumulated by trapped electrons. Certain PbI_2 crystals also exhibit a peak in a region of 580–595 nm. This peak becomes much higher if the crystal is treated with an N_2 laser at room temperature or if it is heated to 450–485 K. As the measurement temperature rises from 85 to 295 K, luminescence intensity decreases considerably. With X-ray excitation at room temperature, the yield of PbI_2 : Mn luminescence peaked at 660 nm for doped crystals is about three times larger than the yield peaked at 555 nm for nondoped crystals. The spectral curves and underlying radiative processes are discussed. © *2001 MAIK "Nauka/Interperiodica".*

The luminescence from PbI_2 crystals excited by X-ray pulses was addressed in [1]. With temperature raised from 77 to 293 K, it was revealed that the emission intensity decreases by a factor of ~25 and the spectral peak shifts from 520 to 550 nm. A remarkable feature of the crystals is that the temperature rise has an insignificant effect on the pulse luminescence decay time, which is 0.4 ns at most. The high density and short luminescence decay time of PbI_2 crystals suggest that they may be used as fast scintillators. Regarding the activated PbI_2 crystals, little attention has been given to their luminescent properties under pulsed or stationary X-ray excitation.

This experimental study examines the luminescence from PbI_2 : Mn crystals under stationary excitation by X-rays or laser radiation. We used an LGI-21 nitrogen laser operating at $\lambda = 337.1$ nm. Our aim was to explore the effect of Mn doping on the accumulation of excitation energy by trapped electrons, the specific light yield, and the luminescence spectrum for temperatures ranging from 85 to 295 K.

The crystals were grown by the Stockbarger–Bridgman technique [2]. During their growth, the melt was doped with ~0.5 mol % $MnCl_2$. The starting material (PbI_2) was of analytical-grade purity and was further purified by oriented crystallization [2]. Luminescent properties of the crystals were studied according to the technique described in [3].

First, let us consider the luminescence spectra of nondoped crystals. With laser excitation at 85 K, we detected an intense narrow band with a peak near 512 nm and a faint band centered at 495 nm. With X-ray excitation at the same temperature, the spectrum consists of bands at 515–520 and 715 nm. Some of the specimens also exhibited a peak in the region of 580–595 nm under both types of excitation (see curve *1* in Fig. 1). This peak becomes much higher after the crystal (in the cryostat) is heated to 450–485 K or treated with N_2 laser radiation at room temperature. These crystals retain their luminescent properties for about six months in storage at the laboratory.

If the measurement temperature is raised from 85 to 295 K, luminescence decreases considerably. With X-ray excitation at room temperature, the spectral curve has a low-intensity band centered at 555 nm and a wide plateau between 640 and 780 nm (see curve *2* in Fig. 1).

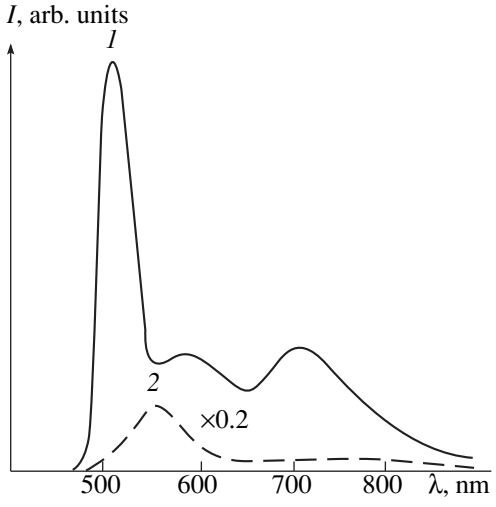

Fig. 1. Luminescence spectra of a PbI_2 crystal under X-ray excitation for (*1*) 85 and (*2*) 295 K.

1063-7842/01/4602-0262$21.00 © 2001 MAIK "Nauka/Interperiodica"

Similar shapes have spectra obtained with laser excitation near the higher temperature limit.

At 85 K, trapped electrons accumulate a small amount of energy during laser or X-ray excitation. When the temperature is raised to 295 K, nonelementary peaks were observed in the thermoluminescence curves at about 117 K.

Now, let us proceed to Mn-doped PbI_2 crystals. Consider the case of X-ray excitation. At 85 K, luminescence intensity was found to decrease by three times in the 515-nm band and to increase for longer wavelengths, with the highest peak shifted to 700 nm (curve *1* in Fig. 2). The plateau between 590 and 610 nm stems from a peak at 595 nm, which characterizes the thermally treated PbI_2 crystals.

A rise in temperature crucially reduces the peak at 700 nm. Specifically, for 150–295 K, its height is about 5% of that for 85 K. At 295 K, the spectrum is characterized by a low-intensity wide nonelementary band centered at 660 nm (curve *2* in Fig. 2). At this temperature, the specific light yield is about three times larger than that of nondoped crystals. The energy accumulated under X-ray excitation is small again.

Let us consider laser excitation. At 85 K, the spectrum has a low-intensity band centered at 700 nm as well as an intense narrow band at 515 nm (curve *1* in Fig. 3). If the temperature is raised to 295 K, the luminescence intensity is strongly quenched and the crystal emits at wavelengths longer than 550 nm (curve *2* in Fig. 3). The falling segment of the spectral curve indicates emission peaks at 660 and 700 nm.

Figure 4 shows the luminescence intensity in the bands at 515 and 700 nm against temperature for the case of laser excitation. It is seen that the two curves differ markedly. As temperature increases from 85 to 180 K, the intensity in the 515-nm band decreases rapidly and monotonically (curve *1*), whereas that in the 700-nm band increases slightly until 95 K is reached and then decreases monotonically (curve *2*). Eventually, intensity falls to a very low level in both cases.

Thus, the above results demonstrate that doping with Mn changes the PbI_2 luminescence spectrum. For laser excitation, a new band arises. For X-ray excitation, the intensity is redistributed in favor of longer wavelengths. At low temperatures, a similar effect was observed in crystals doped with CdI_2. Based on [4–7], we reason that interstitial Mn ions have a valency of 2. No luminescence from Mn^{2+} centers in the crystals was observed near the lower temperature limit. This may stem from the fact that the excited states of doping ions are in the conduction band or that the dopant forms no efficient recombination centers that could compete with those of the matrix [8].

Some authors attribute the edge emission at 512 nm to I_2^- centers in PbI_2 [9, 10]. The low-intensity peak located near 495–500 nm is associated with the radiative relaxation of free excitons. If PbI_2 is irradiated with

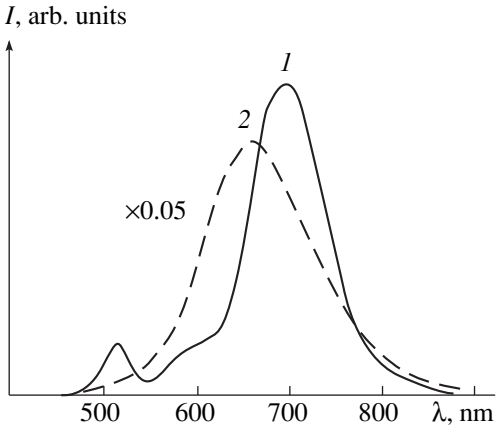

Fig. 2. Luminescence spectra of a PbI_2 : Mn crystal under X-ray excitation for (*1*) 85 and (*2*) 295 K.

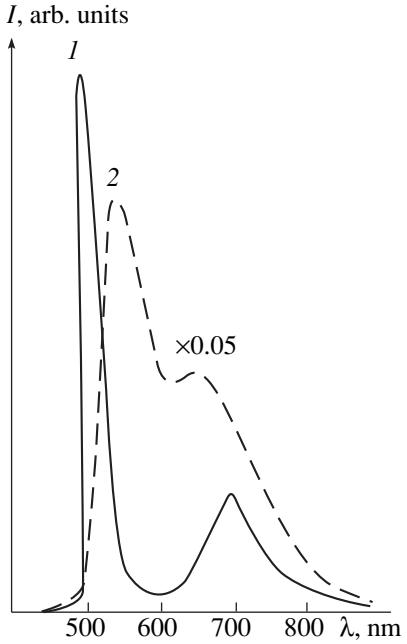

Fig. 3. Luminescence spectra of a PbI_2 : Mn crystal under laser excitation for (*1*) 85 and (*2*) 295 K.

photons corresponding to band-to-band absorption, an induced signal can be detected at 77 K. The signal represents electron paramagnetic resonance (EPR) and exhibits a line with a *g*-factor of 2.009. This EPR signal is ascribed to Pb^+ centers in PbI_2. They result from electron trapping by irregular Pb^2 ions located near extended defects or at sites where lead–iodine bonds are broken. Another EPR signal, which has a complicated nature and reflects hyperfine interaction, is associated with a hole trapped by two I^- ions with the formation of an I_2^- molecular ion [10, 11].

In single crystals and polycrystals of PbI_2, the 715-nm band has a high quantum yield mainly under

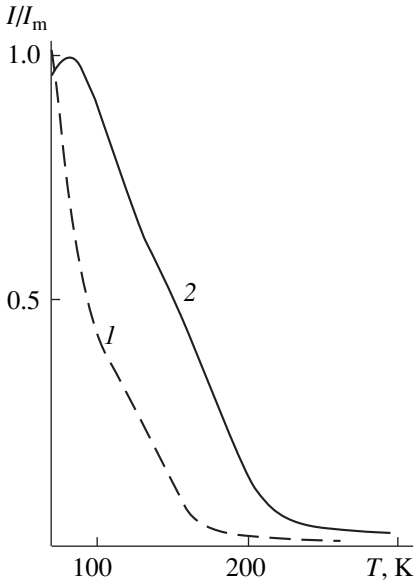

Fig. 4. Normalized luminescence intensity vs. temperature for (*1*) the 515- and (*2*) the 700-nm band in the case of a PbI_2 : Mn crystal under laser excitation.

optical excitation near the absorption band edge. It is conceivable that the EPR signal with $g = 2.009$ and the red band of photoluminescence have a common mechanism, since they both obey a boat-shaped temperature characteristic in a range of 80–150 K [10]. In all likelihood, the red band is caused by defects in the crystal bulk, and the photoluminescence probably stems from recombination. This explains why a nondoped perfect PbI_2 crystal does not luminesce in the 715-nm band when it is irradiated with an N_2 laser operating at a wavelength for which the crystal offers strong intrinsic absorption in a near-surface layer, but does luminesce when it is irradiated with X-rays, which have a higher penetrating power.

Based on [10], one could associate this optical emission with the $^3P_1 \longrightarrow {}^1S_0$ radiative transitions in Pb^{2+} ions induced by the recombination of holes with Pb^+ electron-trapping centers. Investigating the photovoltaic properties of PbI_2 crystals has revealed that this material is an *n*-type semiconductor if its temperature is below 180 K [12]. Otherwise, it is a *p*-type semiconductor. We therefore suggest that the 715-nm emission is produced by the recombination of electrons and the holes confined to halogen ions in the neighborhood of Pb^+ ions. The latter ions are associated with anionic vacancies [13] that are always present in PbI_2 as a result of the volatilization of iodine during crystal growth [14]. In short, we connect the 715-nm emission with α-centers (see also [15]). Since Mn^{2+} has a smaller ionic radius compared to Pb^{2+} [16], the presence of Mn^{2+} ions in a PbI_2 crystal causes lattice disordering and hence a rise in the concentration of the centers of long-wave luminescence both in the bulk and near the surface.

An abrupt decrease in the intensity–temperature characteristic measured at 515 nm may result from the temperature-stimulated delocalization of holes from V_k centers [8]. Some of the holes are then trapped by deeper centers; hence, a slight increase in the curve for 700 nm is observed. In this band, luminiscence quenching near the higher temperature limit may be induced by both the temperature-stimulated delocalization of holes from the centers, followed by their radiationless recombination at deep electron-trapping centers, and ionic and photochemical processes [2].

A peak in the region 580–595 nm, which is characteristic of a near-surface region in nondoped crystals, is probably related to oxygen [17]. As noted above, this peak can be made higher by heat or laser treatment of the crystal. The luminescence intensity markedly decreases as the temperature rises from 85 to 200 K.

We also studied the kinetic properties of luminescence from PbI_2 crystals under β-, γ-, or X-ray excitation according to the techniques described in [1, 17]. The results agree with [1]. This allows us to infer that, for the 515- to 520-nm and the 555-nm band, doped crystals under X-ray excitation also have a short luminescence decay time ($\tau_1 \approx 0.5$ ns). The longer luminescence decay time ($\tau_2 \approx 1.5 \times 10^{-5}$ s) of nondoped crystals measured between 500 and 800 nm at room temperature seems to stem from the radiative relaxation of excitons confined to defects.

REFERENCES

1. A. S. Voloshinovskiĭ, P. A. Rodnyĭ, and A. Kh. Kkhudro, Opt. Spektrosk. **76** (3), 428 (1994) [Opt. Spectrosc. **76**, 382 (1994)].
2. *Wide-Gap Layered Crystals and Their Physical Properties*, Ed. by A. B. Lyskovich (Vishcha Shkola, Lviv, 1982).
3. O. M. Bordun, I. M. Bordun, and S. S. Novosad, Zh. Prikl. Spektrosk. **62** (6), 91 (1995).
4. V. G. Abramishvili and A. V. Komarov, Fiz. Tverd. Tela (Leningrad) **31** (4), 68 (1989) [Sov. Phys. Solid State **31**, 583 (1989)].
5. M. S. Brodin, A. O. Gushcha, B. E. Derkach, *et al.*, Fiz. Tverd. Tela (Leningrad) **30** (11), 3481 (1988) [Sov. Phys. Solid State **30**, 1999 (1988)].
6. V. G. Abramishvili, A. V. Komarov, S. M. Ryabchenko, *et al.*, Fiz. Tverd. Tela (Leningrad) **29** (4), 1129 (1987) [Sov. Phys. Solid State **29**, 644 (1987)].
7. M. S. Brodin, I. V. Blonskiĭ, and V. N. Karataev, Ukr. Fiz. Zh. (Russ. Ed.) **34** (4), 526 (1989).
8. E. D. Aluker, D. Yu. Lusis, and S. A. Chernov, *Electron Excitations and Radioluminescence in Alkali Metals* (Zinatne, Riga, 1979).

9. J. Arends and J. Verwey, Phys. Status Solidi B **23** (1), 137 (1967).
10. I. Z. Irdutnyĭ, M. T. Kostyshin, O. P. Kasyarum, *et al.*, *Photostimulated Interactions in Metal–Semiconductor Structures* (Naukova Dumka, Kiev, 1992).
11. A. V. Patankor and E. E. Schneider, Phys. Chem. Solids **27** (3), 575 (1966).
12. Takeshi Hagihara, Shunji Nigata, and Nobuvoso Ayai, Jpn. J. Appl. Phys. **20** (5), 1003 (1981).
13. I. Baltag, S. Lefrant, L. Miyut, and R. P. Mondescu, J. Lumin. **63**, 309 (1995).
14. M. S. Brodin, V. A. Bobik, I. V. Blonskiĭ, and N. A. Davidova, Fiz. Tverd. Tela (Leningrad) **32** (2), 403 (1990) [Sov. Phys. Solid State **32**, 232 (1990)].
15. I. A. Parfianovich and É. E. Penzina, *Electron Color Centers in Ionic Crystals* (Vostochno-Sibirskoe Knizhnoe Izd., Irkutsk, 1977).
16. G. B. Bokiĭ, *Crystal Chemistry* (Nauka, Moscow, 1971).
17. V. V. Averkiev, I. M. Bolesta, I. M. Kravchuk, *et al.*, Ukr. Fiz. Zh. (Russ. Ed.) **25** (8), 1392 (1980).

Translated by A. Sharshakov

Technical Physics, Vol. 46, No. 2, 2001, pp. 266–267. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 137–138.

BRIEF COMMUNICATIONS

Glow Discharge with an Annular Hollow Cathode

E. I. Gyrylov

Section of Physical Problems at the Presidium of Buryat Science Center, Siberian Division, Russian Academy of Sciences, Ulan-Ude, 670047 Russia

e-mail: ofp@bsc.buryatia.ru

Received April 10, 2000; in final form, August 4, 2000

Abstract—A discharge system with peripheral discharge chambers (Penning cells) and common hollow cathode with only one cell supplied with power is studied. It is shown that a jumpwise transition from a dark discharge to a glow discharge is accompanied by the penetration of plasma into the hollow cathode. © *2001 MAIK "Nauka/Interperiodica".*

Due to a high plasma density in a glow discharge with a hollow cathode, such discharges are widely used as plasma sources of charged particles [1, 2]. In this paper, we study a discharge system with an annular hollow cathode and two peripheral discharge chambers (Penning cells) [3], one of which is supplied with power, whereas the anode of the other chamber is connected to the cathode (Fig. 1). The penetration of the plasma into the hollow cathode was established by the ion current from a single probe, which was set at the end of the hollow cathode near the aperture of the working cell. The discharge was ignited at a residual pressure of air in the vacuum chamber.

Experiments show that the transition from a dark discharge to a glow discharge occurs in a jump. At a pressure of 13.3 Pa and voltage of 70 V, the discharge current I changes rapidly from 1–50 μA to 1–10 mA and the discharge voltage V falls by 5–10 V. This transition is accompanied by penetration of the plasma into the hollow cathode, the probe current being 1 μA. The absence of the effect of a hollow cathode at low current is related to the fact that, in this case, the charged particles do not enter the hollow cathode from the Penning cell. The plasma and electrons do not penetrate into the hollow cathode because of the drop in the cathode potential in front of the cathode aperture, whereas the penetration of ions is hindered by their scattering in front of the aperture due to the specific shape of the boundary of the cathode dark space.

As the current increases, the length L of the region of the cathode voltage drop decreases. At certain critical parameters of the reflex discharge, the cathode dark space shrinks so that the ion shell in front of the hollow cathode aperture breaks and the plasma penetrates into the cathode. The condition for penetrating is $L \ll R$, where R is the radius of the hollow cathode [1].

The maximum current of the dark discharge attained 200 μA at a discharge voltage of 950 V and a pressure of 8 Pa; in this case, the probe current was zero. The dependence of the discharge current on the voltage at a pressure of 11.7 Pa is shown in Fig. 2. When the glow discharge was ignited at the minimum value of the discharge current required for this type of a discharge to exist (1–1.4 mA), there was plasma in the hollow cathode and the probe current attained 8 μA. As the current

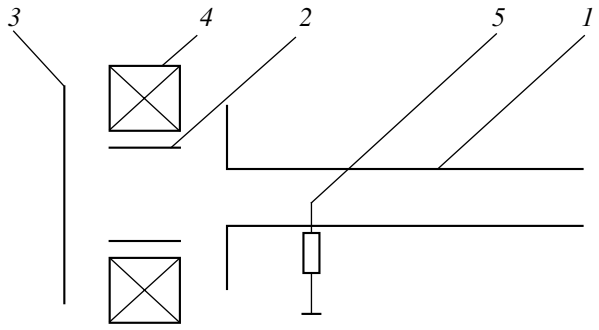

Fig. 1. Diagram of the discharge system: (*1*) hollow cathode, (*2*) anode, (*3*) flat cathode, (*4*) permanent magnet, and (*5*) probe.

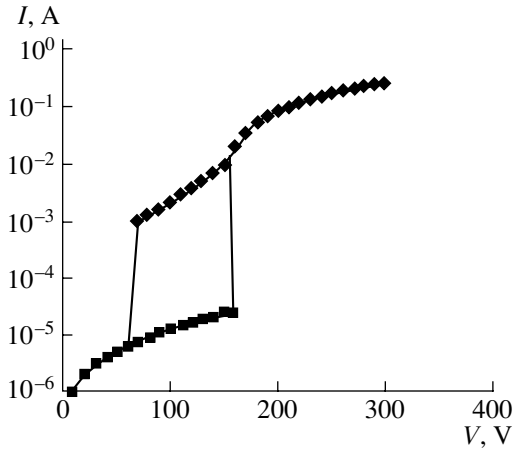

Fig. 2. The discharge current I vs. voltage V at a pressure of 11.7 Pa.

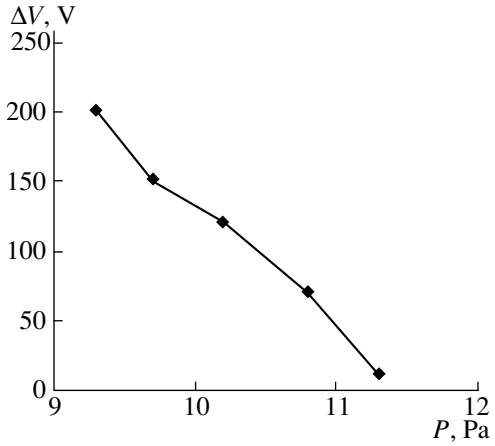

Fig. 3. The decrease in the voltage after the ignition of a glow discharge vs. pressure.

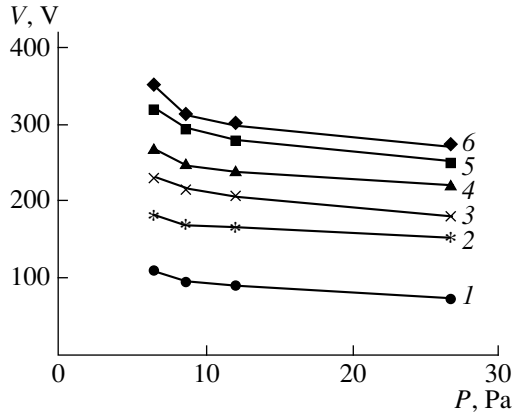

Fig. 4. Dependence of the discharge voltage V on the argon pressure P for different values of the discharge current: I = (*1*) 10, (*2*) 50, (*3*) 100, (*4*) 150, (*5*) 200, and (*6*) 250 mA.

decreased further, the glow discharge transformed into the dark discharge; in this case, there was no plasma in the hollow cathode and the probe current was zero.

The decrease in the voltage after the ignition of a glow discharge versus pressure is shown in Fig. 3. The higher the ignition voltage, the larger the decrease in the voltage; the ratio $\Delta V/V$ ranges within 0.5–1.

Figure 4 presents the dependence of the discharge voltage on the argon pressure for different values of the discharge current in the case when only one cell is supplied with power, whereas the other cell is insulated. At a constant discharge current, the discharge voltage decreases with increasing pressure.

Thus, it is experimentally shown that, in the discharge system under study, plasma penetrates into the hollow cathode when the discharge switches from the dark mode to the glow one.

REFERENCES

1. Yu. E. Kreĭndel', *Plasma Sources of Electrons* (Atomizdat, Moscow, 1977).
2. A. S. Metel', in *Proceedings of the 1st All-Union Conference on Plasma Emission Electronics, Ulan-Ude, 1991*, p. 77.
3. E. I. Gyrylov and A. P. Semenov, Zh. Tekh. Fiz. **65** (1), 189 (1995) [Tech. Phys. **40**, 106 (1995)].

Translated by N. Ustinovskiĭ

Technical Physics, Vol. 46, No. 2, 2001, pp. 268–269. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 139–140.

BRIEF COMMUNICATIONS

Energy Filter in the Form of a Three-Electrode Plane Capacitor with End Diaphragms

L. P. Ovsyannikova and T. Ya. Fishkova

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

e-mail: L.Ovsyannikova@pop.ioffe.rssi.ru

Received April 27, 2000

Abstract—An electrostatic system based on a plane capacitor is suggested to provide monochromatization of the beams of charged particles. In addition to monochromatization, this system also straightens the beam trajectory. The parameters of the system are calculated numerically within a broad range of variation of its strength. © *2001 MAIK "Nauka/Interperiodica".*

Two-electrode plane capacitors are widely used for the energy filtration of charged particles. For this purpose, the capacitor should be operated in the mirror mode with the beam inlet and outlet through its plate. A number of electrodes with linear potential variation are placed along the edges to provide field homogeneity. To simplity the system design, these electrodes can be replaced by end diaphragms, with a potential equal to the potential of any of the capacitor plates. Although both the linear dispersion and energy resolution of such a device are on par with similar characteristics of a classical plane capacitor, the former has a much simpler design than the latter [1]. To ensure angular beam focusing in the mirror mode, both the source and the detector of the charged particles should be near the capacitor. However, in some cases, this requirement cannot be met.

In the preceding works [2, 3], we studied the working modes of an energy filter based on a cylindrical capacitor as charged particles entered through the end electrode. This system allows a more convenient arrangement of the source, detector, and other elements. If the beam intersects the capacitor axis at two points (double intersection), this system provides a straight beam-propagation trajectory. In this case, the system is of considerable interest for developing certain physical devices based on the monochromatic beams of charged particles.

The goal of this work was to solve similar problems using plane electrodes. The cross section of the suggested device is composed of three plane-parallel plates and two plane end diaphragms and is shown in Fig. 1. The diaphragms are placed close to the outer plane electrodes, and their holes are coated with grids to allow the beam to enter and exit. Fieldmaking poten-

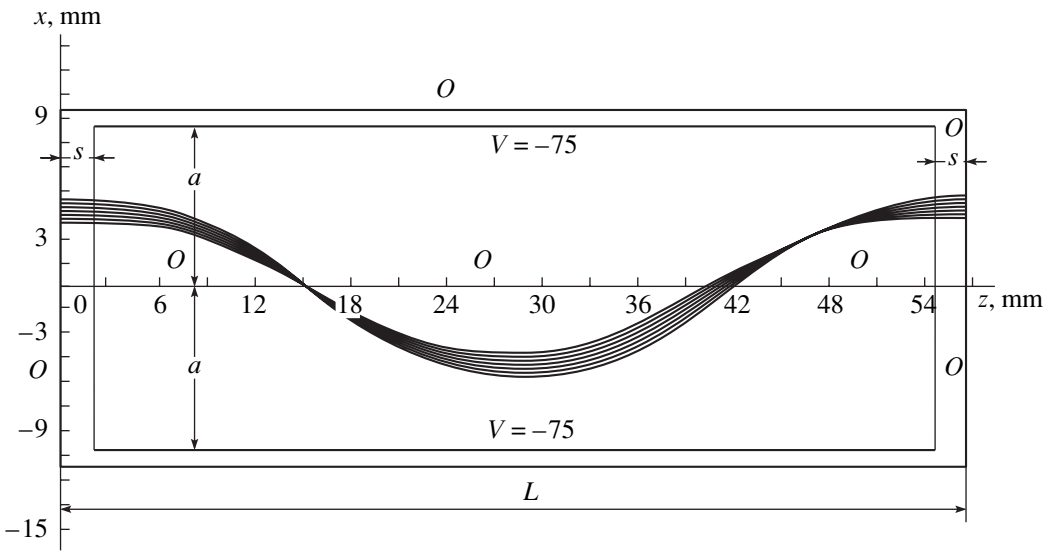

Fig. 1. A three-electrode plane capacitor with end diaphragms providing a straight trajectory of beam propagation.

1063-7842/01/4602-0268$21.00 © 2001 MAIK "Nauka/Interperiodica"

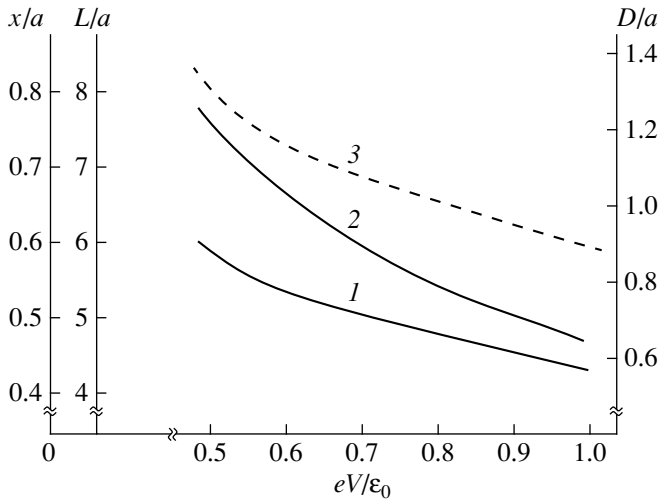

Fig. 2. Parameters of the system shown in Fig. 1 as functions of its strength: (*1*) distance from the axis to the beam entrance and exit in case of an axial beam trajectory; (*2*) system length; and (*3*) linear energy dispersion.

tials are applied to the outer electrodes, whereas both the inner plate of the capacitor and the end diaphragms are grounded. The length of the system is several times longer than its width.

The geometrical electric parameters of the system were calculated numerically using the TEO computer program for a plane problem [3]. If the beam inlet direction is parallel to the longitudinal axis of the system and the inlet site is at a certain distance from the axis, the strength of the system and its total length should be chosen to meet the following condition: the focused beam spot size at the site of the first intersection should be as large as possible, whereas the output beam should be as parallel as possible. Therefore, this system provides a straight trajectory of beam propagation and its energy filtration according to the size of the first slit in the middle plate.

The results of numerical calculations of the parameters of the energy filter described above are given in Fig. 2. In these calculations, the distance between the plane electrodes and the gap between the outer and end electrodes were taken to be a = 10 mm and $s = 0.2a$, respectively. The trajectories of propagation of an electron beam with an initial energy of ε_0 = 100 eV in one of the working modes of the device are shown in Fig. 1.

The parameters of the system were chosen to make the inlet and outlet coordinates of the axial beam trajectory equal to one another: $x = x_0 = x_1$ (Fig. 2, curve *1*). In this case, dependence of the total length of the system L on its strength eV/ε_0 (where V is the potential difference between the outer and middle plates of the capacitor) is characterized by curve *2* (Fig. 2), whereas the dependence of the coefficient of linear dispersion $D = \Delta z\, \varepsilon_0/\Delta\varepsilon$ is characterized by curve *3* (Fig. 2). It follows from Fig. 2 that the stronger the system, the smaller its length and dispersion and the closer the beam is to its axis. If the inlet size of a parallel beam is $\Delta x_0 = 0.1a$, the size of the beam at the first intersection site does not exceed $0.02a$, whereas the outlet size of the beam is $\Delta x_j < 1.2\Delta x_0$. The exit angle of the axial trajectory in any working mode does not exceed 0.5°, and the maximum deviation of peripheral trajectories is less than ±3°.

The energy filter with straight beam trajectory described in this work is based on a three-electrode plane capacitor with end diaphragms. The parameters of the energy filter are close to those of a similar system based on a cylindrical capacitor. In terms of design characteristics, this energy filter is compatible with various optoelectronic systems composed of plane elements. It should be noted that such purely electrostatic systems are similar to a Wien filter with crossed electrostatic and magnetic fields.

REFERENCES

1. L. P. Ovsyannikova and T. Ya. Fishkova, Zh. Tekh. Fiz. **65** (3), 113 (1995) [Tech. Phys. **40**, 288 (1995)].
2. L. P. Ovsyannikova and T. Ya. Fishkova, Zh. Tekh. Fiz. **62** (5), 179 (1992) [Sov. Phys. Tech. Phys. **37**, 582 (1992)].
3. L. P. Ovsyannikova, S. V. Pasovets, and T. Ya. Fishkova, Zh. Tekh. Fiz. **62** (12), 171 (1992) [Sov. Phys. Tech. Phys. **37**, 1215 (1992)].

Translated by K. Chamorovskiĭ

Technical Physics, Vol. 46, No. 2, 2001, pp. 270–271. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 71, No. 2, 2001, pp. 141–142.

BRIEF COMMUNICATIONS

The Effect of Boundary Conditions on the Superconducting Transition Critical Temperature in Superlattices

I. N. Askerzade

Institute of Physics, Academy of Sciences of Azerbaijan, pr. Dzhavida 33, Baku, 370143 Azerbaijan

Received May 15, 2000

Abstract—For a superconducting superlattice consisting of alternating layers of two materials with thicknesses *a* and *b*, an analogue of the Cooper formula is obtained for boundary conditions in the general form. The effect of the boundary conditions on the critical temperature is studied, and the possibility of order parameter localization in the layers with a higher critical temperature is demonstrated. © *2001 MAIK "Nauka/Interperiodica".*

The phenomenon of high-temperature superconductivity has recently generated interest in experimental [1, 2] and theoretical [3–5] investigation of superconducting superlattices. The advanced deposition techniques enable one to produce artificial superlattices with smoothly varying parameters. Therefore, superlattices have a great potential for simulating properties of layered superconductors.

Let a superlattice consist of two alternating superconducting layers *A* and *B* of thicknesses *a* and *b* and have a total period $L = a + b$. In the classical statement of the Cooper problem, the boundary conditions imposed on the wave function (WF) of a Cooper pair are the continuity of the WF and its first derivative at the interface:

$$\Psi_A = \Psi_B, \quad \frac{1}{m_A}\left(\frac{\partial\Psi}{\partial z}\right)_A = \frac{1}{m_B}\left(\frac{\partial\Psi}{\partial z}\right)_B, \qquad (1)$$

where $m_{A,B}$ is the effective mass of charge carriers and subscripts *A* and *B* relate to the materials on each side of the interface.

Suppose that the critical temperature in layer *A*, T_{cA}, is higher than in the other, T_{cB}. In the general case, the WF and its derivative on one side of the interface should be joined with a linear combination of Ψ and $\partial\Psi/\partial z$ at its other side:

$$\begin{pmatrix}\Psi_A\\ \varphi_A\end{pmatrix} = \begin{bmatrix}t_{11} & t_{12}\\ t_{21} & t_{22}\end{bmatrix}\begin{pmatrix}\Psi_B\\ \varphi_B\end{pmatrix} = \hat{T}_{AB}\begin{pmatrix}\Psi_B\\ \varphi_B\end{pmatrix}, \qquad (2)$$

where

$$\varphi_A = l\left(\frac{\partial\Psi}{\partial z}\right)_A, \quad \varphi_B = l\frac{m_A}{m_B}\left(\frac{\partial\Psi}{\partial z}\right)_B, \qquad (3)$$

l is an arbitrary constant length, and $\hat{T}_{AB}$ stands for the transition matrix at the interface.

It is interesting to trace a relationship between the transition matrix and the superlattice critical temperature. For this purpose, we will study the effect of the boundary conditions upon the density of states on the Fermi surface and, hence, upon $T_c(a, b)$ in terms of the BCS theory in the effective mass approximation. Such a method was applied within the Eliashberg approach [6] to derive the critical temperature for an ellipsoidal Fermi surface.

We consider the Debye frequencies and the constants of four-fermion interaction in the BCS theory to be the same; this assumption is justified because of a weak dependence of T_c on ω_d and on the order of the electron–phonon interaction constants (the weak coupling model). The layers are considered to differ in mass only. Due to the periodic structure of superlattices, the carrier masses can be renormalized and the concept of longitudinal and transverse effective masses can be introduced [7]. Then, employing the transition matrix method, the electron spectrum is readily obtained from the solution of the Schrödinger equation for the wave function envelope satisfying the generalized boundary conditions. The effective masses at different points of a Brillouin zone can be found by expanding the energy in terms of the wave vector. Analysis shows that, at the top of a mini-Brillouin zone, longitudinal and transverse masses are expressed as

$$m_{\parallel} = h\frac{m_A \sinh\kappa b}{k(a+b)C_a^2}, \qquad (4)$$

where

$$h = \zeta(t_{22}^2 + \alpha^2 t_{12}^2) + \zeta^{-1}(t_{11}^2 + t_{21}^2\alpha^{-2}) - 2(\alpha t_{11}t_{12} + \alpha^{-1}t_{21}t_{22})\coth\kappa b,$$

1063-7842/01/4602-0270$21.00 © 2001 MAIK "Nauka/Interperiodica"

$$m_\perp^{-1} = m_A^{-1}\theta + m_B^{-1}(1-\theta),$$
$$\theta = \frac{C_a^2}{2(a+b)}a\left(1+\frac{\sin ka}{ka}\right), \quad (5)$$

$$C_a^2 = 2(a+b)\left\{a\left(1+\frac{\sin ka}{ka}\right)+b\left(\frac{\sinh \kappa b}{\kappa b}+1\right)\right.$$
$$\left.\times\frac{1+\cos ka}{\cosh \kappa b+1}\left[t_{11}-t_{12}\beta\tanh\frac{\kappa b}{2}\right]\right\}^{-1}.$$

The other designations follow [7]. The effective constant of interaction in the BCS expression for critical temperature

$$T_c = \frac{2\gamma}{\pi}\omega_D e^{-1/\lambda_{\text{eff}}} \quad (6)$$

can be evaluated from

$$\lambda_{\text{eff}} = N_{\text{eff}}(0)g. \quad (7)$$

Introducing the longitudinal and transversal effective mass means that the Fermi surface is an ellipsoid. For this spectrum, the density of states is given by

$$N_{\text{eff}}(0) = \frac{P_0}{\pi^2}\sqrt{m_\parallel m_\perp}. \quad (8)$$

A feature of our model is that the effective mass and, hence, the critical temperature depend on the generalized boundary conditions. Let us trace the behavior of the $T_c(a, b)$ function derived in view of the expressions for $m_\parallel$ and $m_\perp$. The generalized boundary conditions lead to the localization of the order parameter: as $a \Rightarrow 0$, T_c tends not to T_{cB} but to $T^* > T_{cB}$, where T^* is defined by elements of the transition matrix $\hat{T}_{AB}$. As this takes place, the high-T_c layers behave as planar defects, where the order parameter is localized. A similar situation is observed in bulk superconductors [8], where it is reasoned by the presence of Tamm levels. With ordinary boundary conditions (1), the order parameter localization does not occur and we come to the Cooper limit; i.e., T_c smoothly varies between T_{cA} and T_{cB}. These results are of interest for the interpretation of experimental data on order parameter localization on a dislocation network [1].

In conclusion, by choosing a transition matrix, one can control the minizone structure of an electron in a superlattice and, consequently, the density of states on the Fermi surface. The latter parameter, in turn, is responsible for the localization effect experimentally observed in superconducting superlattices.

REFERENCES

1. O. A. Mironov, S. V. Chistyakov, I. Yu. Skrylev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **50**, 300 (1989) [JETP Lett. **50**, 334 (1989)].
2. H. C. Yang, L. M. Wang, and H. E. Horng, Phys. Rev. B **59**, 8956 (1999).
3. S. Takahashi, T. Hirai, M. Machida, and M. Tachiki, Physica C (Amsterdam) **235–240**, 2585 (1994).
4. M. V. Baranov, A. I. Buzdin, and L. N. Bulaevskiĭ, Zh. Éksp. Teor. Fiz. **91**, 1063 (1986) [Sov. Phys. JETP **64**, 628 (1986)].
5. V. M. Gvozdikov, Fiz. Nizk. Temp. **15**, 636 (1989) [Sov. J. Low Temp. Phys. **15**, 358 (1989)].
6. M. Prohammer, Physica C (Amsterdam) **157**, 4 (1989).
7. N. F. Gashimzade and E. L. Ivchenko, Fiz. Tekh. Poluprovodn. (Leningrad) **25**, 323 (1991) [Sov. Phys. Semicond. **25**, 195 (1991)].
8. I. M. Suslov, Zh. Éksp. Teor. Fiz. **95**, 949 (1989) [Sov. Phys. JETP **68**, 546 (1989)].

Translated by A. Sidorova-Biryukova