Technical Physics, Vol. 48, No. 12, 2003, pp. 1501–1510. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 1–8.

ATOMS, SPECTRA, RADIATION

Angular Distribution of Auger Electrons for ($M_3 \longrightarrow N_{2,3}N_{2,3}$), ($M_4 \longrightarrow N_{4,5}N_{4,5}$), ($M_4 \longrightarrow N_1N_3$), and ($M_{4,5} \longrightarrow O_{2,3}O_{2,3}$) in a Xenon Atom

A. Yu. Elizarov* and I. I. Tupitsyn**

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

e-mail: a.elizarov@mail.ioffe.ru

*** St. Petersburg State University, Universitetskaya nab. 7/9, St. Petersburg, 198904 Russia*

e-mail: tup@tup.usr.pu.ru

Received March 19, 2003; in final form, June 11, 2003

Abstract—The angular distribution of Auger electrons is considered, and the angular distribution anisotropy parameter α_2 for ($M_3 \longrightarrow N_2N_3$), ($M_3 \longrightarrow N_3N_3$), ($M_4 \longrightarrow N_1N_3$), ($M_4 \longrightarrow N_4N_5$), ($M_4 \longrightarrow N_5N_5$), and ($M_{4,5} \longrightarrow O_{2,3}O_{2,3}$) transitions in a xenon atom is calculated. The matrix elements are evaluated by the nonrelativistic Hartree–Fock method with *LS* coupling, as well as by the relativistic Hartree–Fock method with *jj* coupling (single-configuration approximation) and intermediate coupling (multiconfiguration approximation).

INTRODUCTION

The anisotropy of the Auger electron angular distribution in atoms was first predicted in [1]. Since then, this effect has been the object of much theoretical and experimental investigation. Of special interest is the distribution of Auger electrons that arise upon filling deep vacancies resulting from interaction between atoms and synchrotron radiation [2]. The general theory of Auger electron angular distribution anisotropy, which is based on the density matrix formalism, was developed in [3–6]. Later, the coefficients of the angular distributions were calculated [7–9]. However, the existing discrepancy between theoretical predictions and experimental data [6, 9, 10] necessitates further investigation in this field. The reasons for the discrepancy still remain unclear. One may be the roughness of the theory (specifically, it ignores interference between the final states of the ion–Auger electron system [11]). Various approximations used in calculating the wave functions of the initial and final states of the ions and wave functions of the continuous spectrum may also be a factor.

In this work, we use the theory of Auger electron angular distribution developed in [6–8]. This theory is generalized for the case of atoms with two unclosed valence shells and includes both *LS* and *jj* couplings. In addition, we derive expressions for the angular distribution anisotropy coefficient α_2 in the case of intermediate coupling, i.e., for the multiconfiguration relativistic wave functions of the initial and final states of the ions. The wave functions of the ions are calculated by the Hartree–Fock method and by the multiconfiguration relativistic Hartree–Fock–Dirac method.

The parameters of the Auger electron angular distribution depend on the wave function of the continuous spectrum. In early works, this function did not take into account exchange interaction between an electron of the continuous spectrum and core-shell electrons and ignored off-diagonal Lagrangean factors, which render the wave function of the continuous spectrum orthogonal to occupied single-electron states of the ion. In this work, we consider the effect of these factors on the angular distribution parameter α_2. In relativistic calculations, the wave function of the continuous spectrum was found by solving the Hartree–Fock–Dirac relativistic equations. The influence of relativistic effects on α_2 may be considerable, since the behavior of the continuous spectrum function near the core makes a major contribution to the value of α_2.

In the next section, we outline the basic principles underlying the calculation of the Auger electron angular distribution parameter α_2. Then, we give the values of α_2 calculated for the ($M_3 \longrightarrow N_{2,3}N_{2,3}$), ($M_4 \longrightarrow N_{4,5}N_{4,5}$), ($M_4 \longrightarrow N_1N_3$), and ($M_{4,5} \longrightarrow O_{2,3}O_{2,3}$) transitions in a xenon atom in different approximations.

GENERAL THEORY

The scattering of electrons or photons by an atom $A(J_0)$ may produce a vacancy $A^+(J_1)$ in the inner shell. The occupation of the vacancy is accompanied by the emission of a photon or electron (Auger process). Let

us consider the Auger process

$$A(J_0) + e \longrightarrow e_s + e_k + A^+(J_1) \searrow A^+(J_1) \longrightarrow A^2(J) + e_A. \quad (1)$$

Here, e_s and e_k are scattered and knocked-out electrons, respectively. The Auger process produces a doubly charged ion $A^{2+}(J)$ and an Auger electron e_A in the continuous spectrum. Since the lifetime of the excited state is much longer than the time of collision, the scattering process may be considered as proceeding in two steps: the formation of a vacancy and Auger decay [1]. To exclude interference between the states of the electrons e_s and e_A, we consider a process where the energies of the scattered and Auger electrons differ. The quantum states of the electrons e_s and e_k and those of the doubly charged ion are disregarded. Also, we assume that neither the electron (e) nor the atom $A(J_0)$ are polarized in the initial state $A(\gamma_0 J_0)$.

When electrons are scattered by an atom, a preferential direction appears in the system of colliding particles. This causes [12] anisotropy in the angular distribution of the Auger electron flux intensity $I(\Theta)$. An expression for the angular distribution $I(\Theta)$ of Auger electrons was obtained in [1, 4–7]. In [6], this distribution was described by using parameters $A(KkQ)$, which bear information on the dynamics and geometry of the Auger process. In [6], a general expression for the parameters $A(KkQ)$ was derived:

$$A(KkQ) = \sqrt{(2K+1)(2k+1)} \times \sum_{M, M_1, M_1', m_s, m_s'} (-1)^{J_1 - M_1 + \frac{1}{2} - m_s} \times \begin{pmatrix} J_1 & J_1 & K \\ M_1 & -M_1' & -Q \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & k \\ m_s & -m_s' & -Q \end{pmatrix} \times \langle JM, p^{(-)} m_s | V | J_1 M_1 \rangle \langle JM, p^{(-)} m_s' | V | J_1 M_1' \rangle^*. \quad (2)$$

Here, V is the operator of electron–electron Coulomb interaction:

$$V = \frac{1}{2} \sum_{i \neq j} v(\mathbf{r}_i, \mathbf{r}_j), \quad v(\mathbf{r}, \mathbf{r}') = \frac{1}{|\mathbf{r} - \mathbf{r}'|}. \quad (3)$$

(i) Intermediate coupling and *jj* coupling. Expanding the wave function $\langle p^{(-)} m_s|$ of the Auger electron continuous spectrum in spherical waves in view of spin–orbit interaction [7], we come to an expression for the parameters $A(KkQ)$:

$$A(KkQ) = \frac{1}{4\pi p} \sqrt{(2K+1)(2k+1)} \times \sum_{l, l'} i^{(l'-l)} e^{i(\sigma_l^j - \sigma_{l'}^{j'})} \sum_{j, j'} (-1)^{J + J_1 + j + Q + l'} \times \sqrt{(2l+1)(2l'+1)(2j+1)(2j'+1)} \begin{Bmatrix} J & J_1 & j \\ K & j' & J_1 \end{Bmatrix} \quad (4)$$

$$\times \sum_X (2X+1) \begin{pmatrix} X & l' & l \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} K & X & k \\ -Q & 0 & Q \end{pmatrix} \begin{Bmatrix} \frac{1}{2} & \frac{1}{2} & k \\ j' & j & K \\ l' & l & X \end{Bmatrix} \times \langle (J, \varepsilon j) J_1 \| V \| J_1 \rangle \langle (J, \varepsilon j') J_1 \| V \| J_1 \rangle,$$

where $l + l'$ is even, σ_l^j is the phase shift for an electron of the continuous spectrum in a state $\langle lj|$, and the 6j and 9j symbolic notation of Clebsch–Gordan coefficients is used [13].

The coefficients α_K of the angular distribution are related to the parameters $A(KkQ)$ as [6]

$$\alpha_K(J) = \frac{A(K00)}{A(000)}. \quad (5)$$

It is easy to check that

$$\alpha_K = \left(\sum_{lj} \langle J_1 \| V \| (J_j) J_1 \rangle^2 \right)^{-1} \sqrt{(2K+1)(2J_1+1)} \times \sum_{ll'} \sum_{jj'} (-1)^{J + J_1 - 1/2 + l'} i^{l + l'} C(K)_{jj'} \cos(\sigma_{l'} - \sigma_l) \quad (6)$$

$$\times \langle J_1 \| V \| (J_j) J_1 \rangle \langle J_1 \| V \| (J_{j'}) J_1 \rangle,$$

where

$$C(K)_{jj'} = -(-1)^{K + 2j'} \sqrt{\frac{(2j+1)(2j'+1)}{2K+1}} C^{K0}_{j' - \frac{1}{2}, j + \frac{1}{2}} \times \begin{Bmatrix} J & J_1 & j \\ K & j' & J_1 \end{Bmatrix}. \quad (7)$$

The matrix elements $\langle J_1 \| V \| (J_j) J_1 \rangle$ are determined for the initial and final multielectron states of an arbitrary atom. In the general case, they are derived with the Wigner–Eckart theorem [13, 14] if the multielectron wave function Ψ_{J_1, M_1} for the initial state $A^+(J_1)$ of the ion, the wave function $\Psi_{J, M}$ for the final state $A^{2+}(J)$ of the ion, and the single-electron wave function ψ_{jm} for

the Auger electron are known:

$$\langle J_1\|V\|(J, j)J_1\rangle = \frac{\sqrt{2J_1+1}}{C^{J_1M_1}_{JM,\,jm}}\langle J_1M_1|V|JM, jm\rangle. \quad (8)$$

The wave functions Ψ_{J_1, M_1} and $\Psi_{J, M}$ can be calculated by the relativistic Hartree–Fock–Dirac method in the one-configuration approximation. In this case, the coefficients α_K are determined for pure *jj* coupling. For closed-shell atoms, such an approach is equivalent to that worked out in [6–8] and applies only to closed-shell heavy atoms, for which the final state $A^{2+}(J)$ of the ion has vacancies on deep core levels. More correct results can be obtained by the multiconfiguration Hartree–Fock method. With this method, one can, in particular, take into account the superposition of all relativistic configurations that correspond to one nonrelativistic configuration of the ion, i.e., configurations that have the same occupancies of nonrelativistic shells (*nl*) and different occupancies of the relativistic shells (*nlj*). Such an approximation will be referred to as coupling of the intermediate type. Intermediate coupling passes to *LS* coupling, as it should, in the nonrelativistic limit. However, this is not true for pure *jj* coupling in closed-shell systems, specifically, for ions with two vacancies on inner levels.

In the particular case of atoms with closed valence shells, the expression for the reduced matrix elements $\langle J_1\|V\|(J_j)J_1\rangle$ in hole representation for pure *jj* coupling has the form

$$\begin{aligned}&\langle(J, \varepsilon j)J_1\|V\|J_1\rangle\\ &= (-1)^{\Theta+j}\langle(l_fj_f, l'_fj'_f)J\|v\|(l_ij_i, \varepsilon lj)J\rangle,\end{aligned} \quad (9)$$

where l_fj_f and $l'_fj'_f$ are the quantum numbers of the two vacancies in the ion $A^{2+}(J)$, and l_ij_i are the quantum numbers of one vacancy in the initial state of the ion $A^{2+}(J_1)$.

The parameter Θ takes half-integer values and depends on the phase factors in the wave functions for the initial and final states of the ion. Substituting (9) into (4) and (6) eliminates the dependence of Θ. In this case, the additional phase factor is $(-1)^{j+j'+1}$. The expressions for $A(KkQ)$ and α_K obtained in hole representation are similar to (9) and (25) in [7].

The reduced matrix element in hole representation for *jj* coupling was obtained in [15]. Its final form is given in [7]. In our notation,

$$\begin{aligned}&\langle(l_fj_f, l'_fj'_f)J\|v\|(l_ij_i, \varepsilon lj)J\rangle\\ &= \tau(-1)^{j'_f+j_i}\sqrt{(2j_f+1)(2j'_f+1)(2J+1)}\\ &\times\Bigg[(-1)^J\sum_k C^{j_i\frac{1}{2}}_{j_f\frac{1}{2},k0}C^{j\frac{1}{2}}_{j'_f\frac{1}{2},k0}\begin{Bmatrix} j'_f & j & k\\ j_i & j_f & J\end{Bmatrix}\\ &\times R^k(n_fl_fj_f, n'_fl'_fj'_f, n_il_ij_i, \varepsilon lj)\\ &+\sum_k C^{j_i\frac{1}{2}}_{j'_f\frac{1}{2},k0}C^{j\frac{1}{2}}_{j_f\frac{1}{2},k0}\begin{Bmatrix} j_f & j & k\\ j_i & j'_f & J\end{Bmatrix}\\ &\times R^k(n_fl_fj_f, n'_fl'_fj'_f, \varepsilon lj, n_il_ij_i)\Bigg],\end{aligned} \quad (10)$$

where the coefficient τ depends on whether the two vacancies of the ion $A^{2+}(J)$ are equivalent or not:

$$\tau = \begin{cases}1 & (n_fl_fj_f)\neq(n'_fl'_fj'_f)\\ \dfrac{1}{\sqrt{2}} & (n_fl_fj_f) = (n'_fl'_fj'_f).\end{cases} \quad (11)$$

Radial integrals R^k in (10) coincide with the standard radial integrals in the Hartree–Fock–Dirac method [16]

$$\begin{aligned}&R^k(A, B, C, D)\\ &= \int_0^\infty dr_1\int_0^\infty dr_2[P_A(r_1)P_C(r_1)+Q_A(r_1)Q_C(r_1)]\gamma_k(r_1, r_2)\\ &\times[P_B(r_1)P_D(r_1)+Q_B(r_1)Q_D(r_1)],\end{aligned} \quad (12)$$

where *A*, *B*, *C*, and *D* are the numbers of relativistic shells; *P* and *Q* are the major and minor components of the radial wave function, respectively; and

$$\gamma_k(r_1, r_2) = \frac{r^k_<}{r^{k+1}_>}. \quad (13)$$

(ii) *LS* coupling. If relativistic effects are weak, the ion states can be described in terms of *LS* coupling. To find an expression for the parameters $A(K00)$ of the angular distribution in the case of *LS* coupling, it is necessary to pass from *jj* coupling to *LS* coupling. Then, for the reduced matrix element with *LS* coupling, we arrive at (see, e.g., [17])

$$\begin{aligned}&\langle(J, \varepsilon lj)J_1\|V\|J_1\rangle\\ &= \sqrt{(2L_1+1)(2S_1+1)(2J+1)(2j+1)}\\ &\times\begin{Bmatrix} L & S & J\\ l & \frac{1}{2} & j\\ L_1 & S_1 & J_1\end{Bmatrix}\left\langle\left(\left(LS, \varepsilon l\frac{1}{2}\right)L_1S_1\right)J_1\|V\|(L_1S_1)J_1\right\rangle.\end{aligned} \quad (14)$$

Using the Wigner–Eckart theorem, one easily finds

$$\left\langle \left(\left(LS, \varepsilon l \frac{1}{2} \right) L_1 S_1 \right) J_1 \| V \| (L_1 S_1) J_1 \right\rangle = \sqrt{\frac{(2J_1+1)}{(2L_1+1)(2S_1+1)}} \left\langle \left(LS, \varepsilon l \frac{1}{2} \right) L_1 S_1 \| V \| (L_1 S_1) \right\rangle. \quad (15)$$

In this case, instead of (14), we have

$$\langle (J, \varepsilon l j) J_1 \| V \| J_1 \rangle = \sqrt{(2J_1+1)(2J+1)(2j+1)} \times \begin{Bmatrix} L & S & J \\ l & \frac{1}{2} & j \\ L_1 & S_1 & J_1 \end{Bmatrix} \left\langle \left(LS, \varepsilon l \frac{1}{2} \right) L_1 S_1 \| V \| L_1 S_1 \right\rangle. \quad (16)$$

Then, the parameters $A(K00)$ in LS coupling are given by

$$A(K00) = \frac{1}{4\sqrt{2}\pi p} \times \sum_{l, l'} i^{(l'-l)} e^{i(\sigma_l - \sigma_{l'})} \sqrt{(2l+1)(2l'+1)} C^{K0}_{l_0, l'_0} \times \sum_{j, j'} (-1)^{\frac{1}{2} - J - J_1 + l + l'} (2J_1+1)(2J+1)(2j+1)(2j'+1) \times \begin{Bmatrix} \frac{1}{2} & j' & l' \\ K & l & j \end{Bmatrix} \begin{Bmatrix} J & J_1 & j \\ K & j' & J_1 \end{Bmatrix} \begin{Bmatrix} J_1 & L_1 & S_1 \\ J & L & S \\ j & l & \frac{1}{2} \end{Bmatrix} \times \begin{Bmatrix} J_1 & L_1 & S_1 \\ J & L & S \\ j' & l' & \frac{1}{2} \end{Bmatrix} \left\langle \left(LS, \varepsilon l \frac{1}{2} \right) L_1 S_1 \| V \| L_1 S_1 \right\rangle \times \left\langle \left(LS, \varepsilon l' \frac{1}{2} \right) L_1 S_1 \| V \| L_1 S_1 \right\rangle. \quad (17)$$

Here, L_1, S_1, and J_1 are the quantum numbers describing the state of a singly charged ion A^{2+}; L, S, and J are the quantum numbers describing the state of a doubly charged ion A^{++}; and l and l' are the orbital quantum numbers of an electron from the continuous spectrum (Auger electron). In the nonrelativistic approximation, the wave function of an Auger electron is independent of j; therefore, reduced matrix elements in (17) are independent of j and j'. Then, after summation over j and j', formula (17) takes the form

$$A(K00) = \frac{1}{\sqrt{2}} \frac{1}{4\pi p} (2J+1)(2J_1+1) \times \sum_{l, l'} i^{(l'-l)} e^{i(\sigma_l - \sigma_{l'})} \sqrt{(2l+1)(2l'+1)} C^{K0}_{l_0, l'_0} \times \sum_x (-1)^{x + J_1} (2x+1) \begin{Bmatrix} x & J_1 & l \\ K & l' & J_1 \end{Bmatrix} \times \begin{Bmatrix} L & S & J \\ \frac{1}{2} & x & S_1 \end{Bmatrix}^2 \begin{Bmatrix} L & L_1 & l \\ J_1 & x & S_1 \end{Bmatrix} \begin{Bmatrix} L & L_1 & l' \\ J_1 & x & S_1 \end{Bmatrix} \times \left\langle \left(LS, \varepsilon l \frac{1}{2} \right) L_1 S_1 \| V \| L_1 S_1 \right\rangle \left\langle \left(LS, \varepsilon l' \frac{1}{2} \right) L_1 S_1 \| V \| L_1 S_1 \right\rangle. \quad (18)$$

The reduced matrix elements can be found with the Wigner–Eckart theorem [13, 14] if the multielectron wave function $\Psi_{L_1, M_{L_1}, S_1, M_{S_1}}$ for the initial state A^+ of the ion, the wave function $\Psi_{L, M_L S, M_S}$ for the final state A^{2+} of the ion, and the single-electron wave function $\psi_{lm, \frac{1}{2} m_s}$ for the Auger electron are known:

$$\left\langle \left(LS, \varepsilon l \frac{1}{2} \right) L_1 S_1 \| V \| L_1 S_1 \right\rangle = \frac{\sqrt{(2L_1+1)(2S_1+1)}}{C^{L_1 M_{L_1}}_{L M_L, lm} C^{S_1 M_{S_1}}_{S M_S, lm}} \times \langle L_1 M_{L_1} S_1 M_{S_1} | V | L M_L, lm \rangle. \quad (19)$$

For atoms with closed valence shells, the reduced matrix elements in (17) and (18) can be calculated in hole representation [15] with pure jj coupling by an expression similar to (9):

$$\left\langle \left(LS, \varepsilon l \frac{1}{2} \right) L_1 S_1 \| V \| L_1 S_1 \right\rangle = \left\langle \left(l_f \frac{1}{2}, l'_f \frac{1}{2} \right) LS \| v \| \left(\varepsilon l \frac{1}{2}, l_i \frac{1}{2} \right) LS \right\rangle = \tau (-1)^{l_f + l_i} \sqrt{(2l_f+1)(2l'_f+1)} \times \left[(-1)^L \sum_k R^k (n_f l_f, n'_f l'_f, n_i l_i, \varepsilon l) \times C^{l_i 0}_{l_f 0, k0} C^{l0}_{l'_f 0, k0} \begin{Bmatrix} l'_f & l_f & L \\ l_i & l & k \end{Bmatrix} \right] \quad (20)$$

$$+(-1)^S\sum_k R^k(n_f l_f, n'_f l'_f, \varepsilon l, n_i l_i)$$
$$\times C^{l0}_{l_f 0,\,k0} C^{l_i 0}_{l'_f 0,\,k0}\begin{Bmatrix} l'_f & l_f & L \\ l & l_i & k \end{Bmatrix}\Bigg],$$

where $R^k(nl, n'l', n_1 l_1, n'_1 l'_1)$ is the radial integral:

$$R^k(A, B, C, D) = \int_0^\infty dr_1 \int_0^\infty dr_2 P_A(r_1) P_B(r_2) \times \gamma_k(r_1, r_2) P_C(r_1) P_D(r_2). \quad (21)$$

(iii) Calculation of the wave functions for the continuous spectrum. In the nonrelativistic case, the wave functions of the continuous spectrum were determined in the Hartree–Fock approximation by solving the equation

$$-\frac{1}{2}\frac{d^2}{dr^2}P_{\varepsilon l}(r) + \left[\frac{l(l+1)}{2r^2} + V_C(r)\right]P_{\varepsilon l}(r) + W_{ex}(r) = \varepsilon P_{\varepsilon l}(r) + \sum_{nl}\lambda_{\varepsilon l,\,nl} P_{nl}(r), \quad (22)$$

where n and l are the quantum numbers of occupied atomic shells of the ion A^{2+}, $\lambda_{\varepsilon l,\,nl}$ are the off-diagonal Lagrangean factors that render the continuous spectrum function $P_{\varepsilon l}$ orthogonal to the atomic radial functions P_{nl} of the same symmetry, $V_C(r)$ is the Coulomb potential, and W_{ex} is the result of action of the nonlocal exchange operator on the radial wave function of the continuous spectrum.

The continuous spectrum function $P_{\varepsilon l}$ is normalized to the δ function in terms of energy: $\langle P_{\varepsilon l}|P_{\varepsilon' l}\rangle = \delta(\varepsilon - \varepsilon')$. Then, the asymptotics of $P_{\varepsilon l}$ has the form

$$P_\varepsilon(r) \simeq \sqrt{\frac{2}{\pi p}}\sin(\tau + \sigma_l), \quad \tau = pr + \frac{Z}{p}\ln(2pr) - l\frac{\pi}{2}, \quad (23)$$

where $p = \sqrt{2\varepsilon}$, Z is the ion charge, and σ_l is the scattering phase.

Joining an arbitrary unnormalized regular-at-zero solution $\tilde{P}_\varepsilon(r) = NP_\varepsilon(r)$ to Eq. (22) and its derivative with asymptotics (23), one can determine the normalizing factor N and the phase of the solution $\tilde{P}_\varepsilon(r)$. However, to provide a high accuracy, joining should be accomplished over a distance of 10 000 to 100 000 a.u.; in other words, a solution $\tilde{P}_\varepsilon(r)$ to (22) that is regular at zero must be found over a wide range of the radial variable r. A much more efficient procedure is joining the solution and its derivative with a linear combination of Coulomb functions F and G that are, respectively, regular and irregular at zero [18]:

$$\tilde{P}_\varepsilon(R_{jnt}) = AF(R_{jnt}) + BG(R_{jnt}). \quad (24)$$

Here, R_{jnt} is the position of the joint point. The Coulomb functions can be effectively calculated with the continued fraction technique [19]. In this case, the joint point can be taken in a domain where all radial atom wave functions are negligibly small and the atomic potential may be approximated by a Coulomb potential with a high accuracy; that is, $R_{jnt} \sim$ (20–50) a.u. The normalizing factor N and the phase σ_l can be found from the coefficients A and B:

$$\begin{cases}\cos\sigma_l = \frac{1}{N}[A\cos(\sigma_l^0) - B\sin(\sigma_l^0)] \\ \sin\sigma_l = \frac{1}{N}[A\sin(\sigma_l^0) + B\cos(\sigma_l^0)]\end{cases} \quad N = \sqrt{A^2 + B^2}. \quad (25)$$

Here, σ_l^0 is the phase of the Coulomb functions [17]:

$$\sigma_l^0 = \arg\Gamma(l + 1 + i\eta), \quad \eta = -\frac{Z}{\sqrt{2\varepsilon}}. \quad (26)$$

The relativistic wave functions of the continuous spectrum were found in the Hartree–Fock–Dirac approximation by solving the equation [16]

$$\begin{cases} c\left(-\frac{d}{dr} + \frac{k}{r}\right)Q_{\varepsilon k} + V_C P_{\varepsilon k} + W^P_{ex} = \varepsilon P_{\varepsilon k} + \sum_{nl}\lambda_{\varepsilon l,\,nl} P_{nk} \\ c\left(\frac{d}{dr} + \frac{k}{r}\right)P_{\varepsilon k} + [-2c^2 + V_C]Q_{\varepsilon k} + W^Q_{ex} \\ \quad = \varepsilon Q_{\varepsilon k} + \sum_{nl}\lambda_{\varepsilon l,\,nl} Q_{nk}. \end{cases} \quad (27)$$

Here, $P_{\varepsilon k}$ and $Q_{\varepsilon k}$ are the major and minor components of the radial wave function for the continuous spectrum, P_{nk} and Q_{nk} are the major and minor components of the wave functions for occupied shells in the ion A^{2+}, W^P_{ex} and W^Q_{ex} are the major and minor components of the effect of the nonlocal exchange operator on the two-component radial wave function for the continuous spectrum, and k is the relativistic quantum number.

Energy normalization for relativistic wave functions has the form [20]

$$\int_0^\infty dr[P_\varepsilon(r)P_{\varepsilon'}(r) + Q_\varepsilon(r)Q_{\varepsilon'}(r)] = \delta(\varepsilon - \varepsilon'). \quad (28)$$

When normalized to the δ function in terms of energy, the radial wave function of the continuous spec-

Table 1. Coefficients α_2 for several Auger transitions in Xe (*LS* coupling)

Term		α_2^{11}	α_2^{00}	α_2 [10]	α_2 [9]	α_2 [25]	α_2 [26]
$Xe(M_4N_{4,5}N_{4,5})$	$^1S_0$	–1.0000	–1.0000		–1.000		–1.000
$Xe(M_4N_{4,5}N_{4,5})$	$^3P_0$	–1.0000	–1.0000		–1.000		–1.000
$Xe(M_4N_{4,5}N_{4,5})$	$^3P_1$	–0.8000	–0.8000		–0.800		–0.800
$Xe(M_4N_{4,5}N_{4,5})$	$^3P_2$	0.0	+0.7100		0.0		0.0
$Xe(M_4N_{4,5}N_{4,5})$	$^1D_2$	–0.2240	–0.1917		–0.167		–0.189
$Xe(M_4N_{4,5}N_{4,5})$	$^3F_2$	+0.5817	–0.5867		+0.558	+0.55	+0.607
$Xe(M_4N_{4,5}N_{4,5})$	$^3F_3$	+0.4597	+0.4659		+0.43	+0.42	+0.493
$Xe(M_4N_{4,5}N_{4,5})$	$^3F_4$	–0.7513	–0.7390		–0.806	–0.82	–0.608
$Xe(M_4N_{4,5}N_{4,5})$	$^1G_4$	–0.6203	–0.6144		–0.640		–0.499
$Xe(M_5N_{4,5}N_{4,5})$	$^1S_0$	–1.0690	–1.0690	–1.069	–1.069		–1.069
$Xe(M_5N_{4,5}N_{4,5})$	$^3P_0$	–1.0690	–1.0690	–1.069	–1.069		–1.069
$Xe(M_5N_{4,5}N_{4,5})$	$^3P_1$	–0.7483	–0.7483	–0.749	–0.748		–0.748
$Xe(M_5N_{4,5}N_{4,5})$	$^3P_2$	–0.3818	–0.3818	–0.371	–0.382		–0.382
$Xe(M_5N_{4,5}N_{4,5})$	$^1D_2$	–0.2394	–0.2050	–0.124	–0.178		–0.202
$Xe(M_5N_{4,5}N_{4,5})$	$^3F_2$	+0.5157	–0.7134	+0.738	+0.0056	–0.02	+0.115
$Xe(M_5N_{4,5}N_{4,5})$	$^3F_3$	+0.3695	+0.3338	+0.336	+0.322	+0.32	+0.412
$Xe(M_5N_{4,5}N_{4,5})$	$^3F_4$	+0.4658*	+0.3774	+0.386	+0.435	+0.420	+0.506
$Xe(M_5N_{4,5}N_{4,5})$	$^1G_4$	–0.6631	–0.6568	–0.710	–0.685		–0.533

* For this line, the experimental value of α_2 equals $+0.4312 \pm 0.12$ [23].

trum has (at large r) the asymptotics [20]

$$\begin{cases} P(r) \simeq \frac{1}{c}\left(\frac{\varepsilon + 2c^2}{\pi p}\right)^{1/2} \sin(\tau + \sigma_k) \\ Q(r) \simeq \frac{1}{c}\left(\frac{\varepsilon}{\pi p}\right)^{1/2} \cos(\tau + \sigma_k), \end{cases} \tag{29}$$

$$\tau = pr - \eta \log 2pr - \frac{\pi l^*}{2},$$

where

$$l^* = \begin{cases} \gamma, & k > 0 \\ \gamma - 1, & k < 0, \end{cases} \tag{30}$$

$$\gamma = \sqrt{k^2 - Z^2/c^2}, \quad p = \frac{1}{c}\sqrt{(\varepsilon + c^2)^2 - c^4}, \tag{31}$$

$$\eta = -\frac{Z(\varepsilon + c^2)}{c^2 p}.$$

As in the nonrelativistic case, σ_k here is the phase of the wave function for the continuous spectrum. Also, as in the nonrelativistic case, the phase and normalizing factor of an arbitrary unnormalized regular-at-zero solution $\tilde{P}_\varepsilon$, $\tilde{Q}_\varepsilon$ to the Dirac equation were found by joining this solution with a linear combination of relativistic Coulomb functions that are regular (P_r, Q_r) and irregular (P_i, Q_i) at zero:

$$\begin{cases} \tilde{P}_\varepsilon(R_{jnt}) = AP_r(R_{jnt}) + BP_i(R_{jnt}) \\ \tilde{Q}_\varepsilon(R_{jnt}) = AQ_r(R_{jnt}) + BQ_i(R_{jnt}). \end{cases} \tag{32}$$

The asymptotics of the relativistic Coulomb functions may be taken in the form

$$\begin{cases} P_r(r) \simeq \frac{1}{c}\left(\frac{\varepsilon + 2c^2}{\pi p}\right)^{1/2} \sin(\tau + \sigma_k^0) \\ Q_r(r) \simeq \frac{1}{c}\left(\frac{\varepsilon}{\pi p}\right)^{1/2} \cos(\tau + \sigma_k^0), \end{cases}$$

$$\begin{cases} P_i(r) \simeq \frac{1}{c}\left(\frac{\varepsilon + 2c^2}{\pi p}\right)^{1/2} \cos(\tau + \sigma_k^0) \\ Q_i(r) \simeq -\frac{1}{c}\left(\frac{\varepsilon}{\pi p}\right)^{1/2} \sin(\tau + \sigma_k^0). \end{cases} \tag{33}$$

Here, σ_k^0 is the phase of the relativistic Coulomb functions. Using the standard expression [21], one easily finds

$$\sigma_k^0 = \arg\Gamma(l^* + 1 + i\eta) + \frac{1}{2}\arg\left(\frac{k + i\eta c^2/(\varepsilon^2 + c^2)}{\gamma k/|k| + i\eta}\right). \tag{34}$$

Table 2. Coefficients α_2 for several Auger transitions in Xe for *LSJ* coupling and intermediate coupling (I)

Term	${}_{(00)}\alpha_2$ [7]	${}_{(11)}\alpha_2^{HF(LSJ)}$	${}_{(11)}\alpha_2^{HFD}$	$\alpha_2^{HFD(I)}$
$Xe(M_3N_2N_3)^3P_2\ \alpha_2$	−0.0905	0.0	−0.0836	−0.0000
$Xe(M_3N_3N_3)^3P_2\ \alpha_2$	+0.5431	+0.8000	+0.5332	+0.6212
$Xe(M_4N_4N_5)^3F_4\ \alpha_2$	−0.6805	−0.7948	−0.6706	−0.8266
$Xe(M_4N_5N_5)^3F_4\ \alpha_2$	+0.4161	+0.4409	+0.2703	+0.3837
$Xe(M_5N_4N_5)^3F_4\ \alpha_2$	−0.6041	−0.8004	−0.6034	−0.8314
$Xe(M_5N_5N_5)^3F_4\ \alpha_2$	+0.1544	+0.4370	+0.1688	+0.3796
$Xe(M_4N_1N_3)^3P_2\ \alpha_2$	+0.4760	+0.6818	+0.5058	+0.5034

Note: α_2^{HF} and α_2^{HFD} are the anisotropy parameters of the photoelectron angular distribution for the wave functions calculated by the Hartree–Fock and Hartree–Fock–Dirac methods.

Table 3. Coefficients α_2 for several Auger transitions in Xe for *LS* coupling and intermediate coupling

Term	Experiment [27]	*MHFD* [10]	*HF*(*LS*)	*HFD*(*I*)
$Xe(N_4O_{2,3}O_{2,3})^1S_0$		−1.000	−1.000	−1.0000
$Xe(N_4O_{2,3}O_{2,3})^3P_2$	0.72 ± 0.13	0.231	0.000	−0.1674
$Xe(N_4O_{2,3}O_{2,3})^3P_1$	−0.73 ± 0.11	−0.837	−0.800	−0.8321
$Xe(N_4O_{2,3}O_{2,3})^3P_0$		−1.000	−1.000	−1.0000
$Xe(N_4O_{2,3}O_{2,3})^1D_2$	0.05 ± 0.06	−0.116	0.5160	0.3634
$Xe(N_5O_{2,3}O_{2,3})^1S_0$		−1.069	−1.069	−1.0690
$Xe(N_5O_{2,3}O_{2,3})^3P_2$	0.47 ± 0.13	−0.385	−0.382	−0.2017
$Xe(N_5O_{2,3}O_{2,3})^3P_1$	−0.77 ± 0.17	−0.743	−0.748	−0.7309
$Xe(N_5O_{2,3}O_{2,3})^3P_0$	−1.07 ± 0.10	−1.069	−1.069	−1.0690
$Xe(N_5O_{2,3}O_{2,3})^1D_2$	0.24 ± 0.10	0.094	0.551	0.6167

For relativistic wave functions with asymptotics (33), the Wronskian has the form

$$W = (P_i Q_r - P_r Q_i) = \frac{1}{c\pi}. \tag{35}$$

To calculate the relativistic Coulomb functions, we applied a transformation [21, 22] that allows one to reduce the radial Coulomb-like Dirac equation to two differential equations that formally coincide with nonrelativistic Schrödinger equations. This transformation can be written as

$$\begin{pmatrix} P' \\ Q' \end{pmatrix} = U\begin{pmatrix} P \\ Q \end{pmatrix}, \quad U = \begin{pmatrix} 1 & X \\ X & 1 \end{pmatrix}, \qquad X = -\frac{Z}{c}\frac{k}{|k|}\frac{1}{|k|+\gamma}. \tag{36}$$

Using the transformation U, one can obtain the second-order differential equations

$$\left[-\frac{d^2}{dr^2} + \frac{l_1^*(l_1^*+1)}{r^2} - \frac{2Z^*}{r}\right]P' = 2\varepsilon^* P',$$
$$\left[-\frac{d^2}{dr^2} + \frac{l_2^*(l_2^*+1)}{r^2} - \frac{2Z^*}{r}\right]Q' = 2\varepsilon^* Q', \tag{37}$$

where

$$\varepsilon^* = \frac{(\varepsilon + c^2)^2 - c^4}{2c^2} = \varepsilon\left(1 + \frac{\varepsilon}{2c^2}\right),$$
$$Z^* = \frac{Z(\varepsilon + c^2)}{c^2} = Z\left(1 + \frac{\varepsilon}{c^2}\right), \tag{38}$$
$$l_1^* = \begin{cases} \gamma, & k > 0 \\ \gamma - 1, & k < 0 \end{cases}, \quad l_2^* = \begin{cases} \gamma - 1, & k > 0 \\ \gamma, & k < 0 \end{cases}.$$

Table 4. Coefficients α_2 for the Xe($M_{4,5}N_{4,5}N_5$) Auger transition, where the orbital moment l_1' of the Auger electron takes values 2, 4, and 6, for *jj* coupling

Term	*jj*	α^{00}	α^{01}	α^{11}
Xe($M_5N_4N_5$)	$\left(\frac{3}{2},\frac{3}{2}\right)_0$	–1.0690	–1.0690	–1.0690
Xe($M_5N_4N_5$)	$\left(\frac{3}{2},\frac{3}{2}\right)_2$	–0.3059	+0.2914	–0.7881
Xe($M_4N_4N_5$)	$\left(\frac{3}{2},\frac{5}{2}\right)_1$	–0.8000	–0.8000	–0.8000
Xe($M_4N_4N_5$)	$\left(\frac{3}{2},\frac{5}{2}\right)_2$	+0.408	+0.0258	+0.0247
Xe($M_4N_4N_5$)	$\left(\frac{3}{2},\frac{5}{2}\right)_3$	+0.4616	+0.4574	+0.4553
Xe($M_4N_4N_5$)	$\left(\frac{3}{2},\frac{5}{2}\right)_4$	–0.6398	–0.6644	–0.4553
Xe($M_4N_5N_5$)	$\left(\frac{5}{2},\frac{5}{2}\right)_0$	–1.0000	–1.0000	–1.0000
Xe($M_4N_5N_5$)	$\left(\frac{5}{2},\frac{5}{2}\right)_2$	–0.7806	–0.7605	–0.7591
Xe($M_4N_5N_5$)	$\left(\frac{5}{2},\frac{5}{2}\right)_4$	–0.1992	+0.1426	–0.1441
Xe($M_5N_4N_5$)	$\left(\frac{3}{2},\frac{3}{2}\right)_0$	–1.0690	–1.0690	–1.0690
Xe($M_5N_4N_5$)	$\left(\frac{3}{2},\frac{3}{2}\right)_2$	–0.7877	–0.7873	–0.7881
Xe($M_5N_4N_5$)	$\left(\frac{3}{2},\frac{5}{2}\right)_1$	–0.7483	–0.7483	–0.7483
Xe($M_5N_4N_5$)	$\left(\frac{3}{2},\frac{5}{2}\right)_2$	–0.1198	–0.1346	–0.1372
Xe($M_5N_4N_5$)	$\left(\frac{3}{2},\frac{5}{2}\right)_3$	+0.3719	+0.3665	+0.3639
Xe($M_5N_4N_5$)	$\left(\frac{3}{2},\frac{5}{2}\right)_4$	–0.5858	–0.5951	–0.5972
Xe($M_5N_5N_5$)	$\left(\frac{5}{2},\frac{5}{2}\right)_0$	–1.0690	–1.0690	–1.0690
Xe($M_5N_5N_5$)	$\left(\frac{5}{2},\frac{5}{2}\right)_2$	–0.3255	–0.3376	–0.3401
Xe($M_5N_5N_5$)	$\left(\frac{5}{2},\frac{5}{2}\right)_4$	+0.1994	+0.1874	+0.1839

Differential equations (37) differ from those given in [22] in that they are written in a form that is more suitable for the solution of the problem stated. Regular (F_1) and irregular (G_1) solutions to the first equation in (37), as well as regular (F_2) and irregular (G_2) solutions to the second equation in (37), can be found by the same procedure [19] as in the nonrelativistic case. Then, for the relativistic Coulomb functions, it is easy to obtain

$$\begin{cases} P_r = \dfrac{N_0}{1-X^2}(F_1 - XF_2) \\ Q_r = \dfrac{N_0}{1-X^2}(F_2 - XF_1), \end{cases} \qquad \begin{cases} P_i = \dfrac{N_0}{1-X^2}(G_1 - XG_2) \\ Q_i = \dfrac{N_0}{1-X^2}(G_2 - XG_1), \end{cases} \tag{39}$$

where the normalizing factor N_0 is given by

$$N_0^2 = \frac{(1-X^2)}{c\pi\gamma p}\left[c\left(\gamma + |k| + \frac{\varepsilon|k|}{c}\right)\right]. \tag{40}$$

RESULTS AND DISCUSSION

We calculated the anisotropy parameter α_2 of the Auger electron angular distribution for ($M_3N_{2,3}N_{2,3}$), ($M_4N_1N_3$), ($M_4N_{4,5}N_{4,5}$), and ($M_{4,5}O_{2,3}O_{2,3}$) transitions in a xenon atom. Our results are in satisfactory agreement with the only known experimental value of α_2 that has been obtained for the Xe($M_5N_{4,5}N_{4,5}$) transition: $\alpha_2 = 0.431 \pm 0.12$ [23] (Table 1). In the relativistic case with intermediate coupling, the calculated value of α_2 is 0.3796 (Table 2). For ($M_{4,5}O_{2,3}O_{2,3}$) Auger transitions, agreement with experiment is much worse. For some of the transitions, α_2 differs even in sign (Table 3), which indicates a need for further research.

In calculations, we took into account exchange interaction and orthogonalized the wave function of the Auger electron partial wave to the core functions of A^{++} with Lagrangean factors. The computational results for *jj* coupling are given in Table 4 with the following notation: α_2^{00}, orthogonalization and exchange are disregarded; α_2^{01}, without orthogonalization; and α_2^{11}, both orthogonalization and exchange are included. From Table 4 it follows that α_2 greatly depends on whether exchange and orthogonalization are taken into consideration. Data listed in Tables 1 and 2 are consistent with calculations in [7, 9–11], where the atom wave functions were obtained by solving the relativistic Hartree–Fock–Dirac equation with the procedure described in [24]. Our calculations of the parameter α_2 for the ($M_{4,5}N_{4,5}N_{4,5}$) Auger transitions in Xe in the case of *LS* and *jj* coupling showed that the inclusion of exchange and orthogonalization (of the Auger electron wave function to core orbitals) in combination affects the value of α_2 almost to the same extent as relativistic effects. The parameter α_2 varies most significantly in the multiconfiguration case with intermediate coupling. As was noted [7], in the *LS* coupling approximation, the orbital moment l of the Auger electron partial waves may take values $l = 2$ and 4, while for *jj* coupling, the Auger electron state with $l = 6$ is also taken into account. The values of α_2 for $l = 2$, 4, and 6 are listed in Table 4. Note that partial waves with $l = 2$ and/or 4 make a major contribution to α_2.

To conclude, we performed "first-principle" calculations of the angular anisotropy parameter for Auger electrons produced by electron scattering by atoms. The *LS*, *jj*, and intermediate coupling approximations for single-electron and multielectron wave functions were used. The parameter α_2 was found to be extremely sensitive to a computational technique. This may be used as a test for computational methods that are applied in particle–atom scattering problems.

ACKNOWLEDGMENTS

This work was supported by the program "Integration" (project no. L-01-02).

REFERENCES

1. W. Mehlhorn, Phys. Lett. **26**, 166 (1986).
2. H. Winick and S. Doniach, *Synchrotron Radiation Researches* (Wiley, New York, 1980).
3. J. Eichler and W. Fritsch, J. Phys. B **9**, 1451 (1976).
4. E. G. Berezhko and N. M. Kabachnik, J. Phys. B **10**, 2467 (1977).
5. H. Klar, J. Phys. B **13**, 4741 (1980).
6. K. Blum, B. Lohmann, and E. Taute, J. Phys. B: At. Mol. Phys. **19**, 3915 (1986).
7. B. Lohmann, J. Phys. B **23**, 3147 (1990).
8. N. M. Kabachnik, H. Aksela, and S. Ritz, Phys. Rev. A **49**, 4653 (1994).
9. M. H. Chen, Phys. Rev. A **45**, 1684 (1992).
10. J. Tulkki, N. M. Kabachnik, and H. Aksela, Phys. Rev. A **48**, 1277 (1993).
11. N. M. Kabachnik, J. Tulkki, H. Aksela, and S. Ritz, Phys. Rev. A **49**, 4653 (1994).
12. B. Cleff and W. Mehlhorn, J. Phys. B **7**, 593 (1974).
13. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).
14. I. I. Sobel'man, *Introduction to the Theory of Atomic Spectra* (Fizmatgiz, Moscow, 1963; Springer-Verlag, Berlin, 1979).
15. W. N. Assad, Nucl. Phys. **44**, 415 (1963).

16. I. P. Grant, Adv. Phys. **19**, 747 (1970).
17. R. Karazija, *The Theory of X-ray and Electronic Spectra of Free Atoms* (Mokslas, Vilnius, 1987).
18. M. Abramovitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1971; Nauka, Moscow, 1979).
19. A. R. Barnett, Comput. Phys. Commun. **24**, 141 (1981).
20. M. E. Rose, *Relativistic Electron Theory* (Wiley, New York, 1961).
21. R. A. Swainson and G. W. F. Drake, J. Phys. A **24**, 79 (1991).
22. L. Infeld, Phys. Rev. **59**, 737 (1941).
23. U. Hahn, J. Semke, H. Merc, and J. Kessler, J. Phys. B **18**, L417 (1985).
24. I. P. Grant, B. J. McKenzie, P. H. Norrington, *et al.*, Comput. Phys. Commun. **21**, 207 (1980).
25. N. M. Kabachnik and I. P. Sazhina, J. Phys. B **9**, 1681 (1976).
26. N. M. Kabachnik, I. P. Sazhina, I. S. Lee, and O. V. Lee, J. Phys. B **18**, L417 (1985).
27. B. Kammerling, V. Schmidt, W. Mehlhorn, *et al.*, J. Phys. B **22**, L597 (1989).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 12, 2003, pp. 1511–1521. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 9–19.

GASES AND LIQUIDS

Nonlinear Vibrations of a Charged Drop in a Third-Order Approximation in the Amplitude of Multimode Initial Deformation

A. N. Zharov, S. O. Shiryaeva, and A. I. Grigor'ev

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

e-mail: shir@uniyar.ac.ru

Received April 15, 2003

Abstract—A solution to the problem of nonlinear surface vibration of a charged ideal liquid drop is found in a third-order approximation in initial multimode deformation of the equilibrium spherical shape by the method of many scales. It is shown that the spectrum of modes that are responsible for the shape of the drop at an arbitrary time instant depends considerably on the spectrum of modes governing the initial deformation of the drop. The latter spectrum also has an effect on nonlinear corrections to the vibration frequencies and, consequently, on a nonlinear correction to the critical Rayleigh parameter, which specifies the stability of the drop against self-charge. © *2003 MAIK "Nauka/Interperiodica".*

(1) The nonlinear dynamics and surface instability of a variously configured charged liquid is of interest for geophysical application, technical physics, scientific instrumentation, and chemical engineering [1–3].

One of the early works that investigated the nonlinear surface vibration of a charged drop of an ideal liquid is that by Tsamopolous and Brown [4]. They gave a solution to the problem of nonlinear vibration of the surface of a charged drop for single-mode initial deformation, when the initial shape of the drop in the spherical coordinate system (r, ϑ, φ) is described by the equation

$$r = R + \xi_0 P_0(\cos\vartheta) + \varepsilon P_m(\cos\vartheta).$$

Here, ε is an arbitrary small parameter determining the initial deformation amplitude, $P_m(\cos\vartheta)$ is an mth-order Legendre polynomial, and ξ_0 is a constant taken in such a way that the volume of the drop at the given initial deformation remains equal to the volume of a spherical drop with a radius R. Also, Tsamopolous and Brown [4] derived an analytical expression for the generatrix of a nonlinearly vibrating drop that is accurate to the second order of smallness in initial deformation amplitude. Finally, they analytically obtained nonlinear third-order corrections to the vibration frequencies at fixed initial deformations. These corrections fitted experimental data well [5]. However, these authors considered a limited spectrum of initial deformations of the drop: the initial deformation was associated with the second ($n = 2$), third ($n = 3$), or fourth ($n = 4$) mode.

The study initiated in [4] was pursued in [6], where the initial excitation of an arbitrary mth mode was analyzed in a second-order approximation in ε. It was also shown [6] that the spectrum of modes that are excited in the second order of smallness because of nonlinear interaction contains only even modes with numbers from the range $[0, 2m]$. It turned out that the drop surface nonlinearly vibrates in the vicinity of a prolate spheroid and not a sphere, as follows from the linear analysis.

The situation where the initial shape of the surface is described by the expression

$$r = R + \xi_0 P_0(\cos\vartheta) + \xi_1 P_1(\cos\vartheta) + \varepsilon(h_{n_1} P_{n_1}(\cos\vartheta) + h_{n_2} P_{n_2}(\cos\vartheta)),$$

where ξ_1 is a constant that is found from the condition of the drop's center-of-mass immovability at nonlinear vibrations and h_{n_1} and h_{n_2} are constants accounting for the partial contribution of each mode to the initial deformation of the spherical surface, was considered in a quadratic approximation in ε in [7]. In that work, the mechanism of nonlinear resonance energy exchange between modes, which takes place when

$$\omega_{n_1} = 2\omega_{n_2},$$

where $\omega_n = (\sigma/\rho R^3)\sqrt{n(n-1)(n+2-W)}$ is the frequency of an nth mode of capillary vibrations of the drop and $W = Q^2/(4\pi\sigma R^3)$ is the Rayleigh parameter, was also studied.

The case when the initial deformation of the drop is due to the superposition of an arbitrary finite number of modes was analyzed in [8]. In such a situation, the initial shape of the drop surface is described by the equation

$$r = R + \xi_0 P_0(\cos\vartheta) + \xi_1 P_1(\cos\vartheta) + \varepsilon \sum_{m \in \Omega} h_m P_m(\cos\vartheta),$$

where Ω is a set of the numbers of initially excited modes, h_m is a constant accounting for the partial contribution of an mth mode to the initial deformation of the drop's spherical shape. The results obtained in [8] were accurate to the second order of smallness in ε, which made it possible to find nonlinear corrections to the mode amplitudes. It appeared that the spectrum of modes of the second order of smallness may contain both even and odd modes. For example, when modes with numbers n_1 and n_2 are excited, the second-order spectrum contains only even modes with numbers from the range $[0, \max\{2n_1, 2n_2\}]$ if n_1 and n_2 are simultaneously even or odd. If n_1 is even and n_2 is odd, the second-order spectrum contains even modes from the range $[0, \max\{2n_1, 2n_2\}]$ and odd modes from the range $[|n_1 - n_2|, n_1 + n_2]$.

In [9] the nonlinear vibration of a charged drop was calculated in the third order of smallness in initial deformation amplitude for an arbitrary initial single-mode deformation, and analytical expressions for the drop generatrix and nonlinear corrections to the frequencies were obtained. In [10], it was shown that internal nonlinear resonances may occur in a charged drop under four-mode interaction, when the initial deformation of the drop is due to the superposition of several modes.

In this work, elaborating upon [9, 10], we study the nonlinear vibration of a drop in the third order of smallness in amplitude of the initial multimode deformation and find third-order corrections to the vibration frequencies.

(2) Let a drop of a perfectly conducting ideal liquid have a density ρ and surface tension coefficient σ. The drop has a radius R and a charge Q. We assume that the flow of the liquid (capillary vibrations) in the drop is potential with a velocity potential ψ. The electrostatic potential of the self-charge near the drop is ϕ. The shape of the drop is assumed to be axisymmetric at any (including initial) time instant. In the dimensionless variables $R = \rho = \sigma = 1$, the equation for the drop surface at an any time t has the form

$$F(r, \vartheta, t) = r - 1 - \xi(\vartheta, t) = 0. \tag{1}$$

The initial deformation of the shape of the drop is given as a superposition of several modes,

$$t = 0: \xi = \xi_0 P_0(\cos\vartheta) + \xi_1 P_1(\cos\vartheta) + \varepsilon \sum_{m \in \Omega} h_m P_m(\cos\vartheta); \tag{2}$$

and the initial velocity of all points on the surface is set equal to zero:

$$t = 0: \partial_t \xi = 0, \tag{3}$$

where ∂_t denotes partial derivative with respect to t.

The complete mathematical statement of the problem of charged drop capillary vibrations includes, along with Eq. (1) for the drop surface and initial conditions (2) and (3), the following equations and conditions [11, 12].

(i) The Laplace equations for the liquid velocity potential and electric field potential:

$$\Delta\psi = 0; \quad \Delta\phi = 0; \tag{4}$$

(ii) boundedness conditions for the potential:

$$r \longrightarrow 0: \psi \longrightarrow 0; \tag{5}$$

$$r \longrightarrow +\infty: \nabla\phi \longrightarrow 0; \tag{6}$$

(iii) kinematic and dynamic boundary conditions

$$r = 1 + \xi(\vartheta, t): \frac{dF}{dt} = 0; \tag{7}$$

$$\partial_t \psi + \frac{1}{2}(\nabla\psi)^2 = p + p_q - p_{at} - p_\sigma; \tag{8}$$

(iv) constancy condition for the volume of the drop:

$$\int_V r^2 \sin\vartheta \, dr d\vartheta d\varphi = \frac{4\pi}{3}; \tag{9}$$

$$V = \{r, \vartheta, \varphi | 0 \le r \le 1 + \xi; \ 0 \le \vartheta \le \pi; \ 0 \le \varphi \le 2\pi\};$$

(v) immovability condition for the center-of-mass of the drop:

$$\int_V \vec{\mathbf{r}} r^2 \sin\vartheta \, dr d\vartheta d\varphi = 0; \tag{10}$$

(vi) constancy condition for the total charge:

$$r = 1 + \xi(\vartheta, t): \int_S \vec{\mathbf{n}} \cdot \nabla\phi \, dS = -4\pi Q; \tag{11}$$

$$S = \{r, \vartheta, \varphi | r = 1 + \xi; \quad 0 \le \vartheta \le \pi; \ 0 \le \varphi \le 2\pi\};$$

(vii) constancy condition for the electric potential over the drop surface:

$$r = 1 + \xi(\vartheta, t): \phi = \phi_S(t). \tag{12}$$

In expressions (4)–(12), p is the equilibrium pressure in the drop; p_q and p_σ are the electric field pressure and capillary pressure, respectively; p_{at} is the atmospheric pressure; $\mathbf{n}$ is the unit vector normal to the drop surface; ϕ_S is the electric potential on the drop surface; and Δ is the Laplacian.

For convenience, we complement the set h_m of constants so that $h_m \equiv 0$ at any $m \in \Omega$.

(3) We will solve problem (1)–(12) by the method of many scales [13]. For this purpose, we introduce three different time scales $T_m = \varepsilon^m t$ $(m = 0, 1, 2)$ and represent the desired parameters in the form of the expansions

$$\phi(r, \vartheta, t) = \sum_{n=0}^{3} \varepsilon^n \phi^{(n)}(r, \vartheta, T_0, T_1, T_2) + O(\varepsilon^4), \tag{13}$$

$$\phi_S(r, t) = \sum_{n=0}^{3} \varepsilon^n \phi_S^{(n)}(r, T_0, T_1, T_2) + O(\varepsilon^4), \tag{14}$$

$$\psi(r, \vartheta, t) = \sum_{n=1}^{3} \varepsilon^n \psi^{(n)}(r, \vartheta, T_0, T_1, T_2) + O(\varepsilon^4),$$
$$\xi(\vartheta, t) = \sum_{n=1}^{3} \varepsilon^n \xi^{(n)}(\vartheta, T_0, T_1, T_2) + O(\varepsilon^4); \quad (15)$$

where $\phi^{(0)} = Q/r$ and $\phi_S^{(0)} = Q$ are zero-order solutions, which correspond to the equilibrium (spherical) shape.

Substituting (13)–(15) into (1)–(12), we obtain problems of different orders of smallness, which are discussed in Appendix A.

Since Laplace equation (4) is linear, the liquid-velocity and electric-field potentials of any order of smallness are solutions to Laplace equations (1A), (10A), and (19A); hence, in view of the boundedness conditions, they can be written in the form

$$\psi^{(m)}(r, \vartheta, T_0, T_1, T_2) = \sum_{n=1}^{\infty} r^n D_n^{(m)}(T_0, T_1, T_2) P_n(\cos\vartheta); \quad m = 1, 2, 3; \quad (16)$$

$$\phi^{(m)}(r, \vartheta, T_0, T_1, T_2) = \sum_{n=0}^{\infty} \frac{F_n^{(m)}(T_0, T_1, T_2)}{r^{n+1}} P_n(\cos\vartheta); \quad m = 1, 2, 3. \quad (17)$$

Note that in expression (16) summation starts with $n = 1$, because, as is well known, the potential is determined accurate to an arbitrary function of time. Hence, we may set $D_0^{(m)} = 0$.

At any time instant, a function describing the deviation of the shape of the drop from spherical can be represented in the form of an expansion in Legendre polynomials:

$$\xi^{(m)}(\vartheta, T_0, T_1, T_2) = \sum_{n=0}^{\infty} M_n^{(m)}(T_0, T_1, T_2) P_n(\cos\vartheta); \quad m = 1, 2, 3. \quad (18)$$

Substituting expressions (16)–(18) into Eqs. (1A)–(9A) yields explicit dependences of first-order quantities on T_0:

$$M_n^{(1)}(T_0, T_1, T_2) = a_n^{(1)}(T_1, T_2)\cos(\omega_n T_0 + \tau_n^{(1)}(T_1, T_2)); \quad (19)$$

$$D_n^{(1)}(T_0, T_1, T_2) = \partial_{T_0} M_n^{(1)}(T_0, T_1, T_2)/n; \quad (20)$$

$$F_n^{(1)}(T_0, T_1, T_2) = Q M_n^{(1)}(T_0, T_1, T_2). \quad (21)$$

In expression (19), the amplitude factor $a_n^{(1)}(T_1, T_2)$ and the nonlinear correction $\tau_n^{(1)}(T_1, T_2)$ to the frequency are functions depending only on the time scales T_1 and T_2.

When the problem in the first approximation in initial surface deformation is solved, the functions $a_n^{(1)}(T_1, T_2)$ and $\tau_n^{(1)}(T_1, T_2)$ should be taken to be constants that are determined from initial conditions (9A):

$$a_n^{(1)} = h_n, \quad \tau_n^{(1)} = 0, \quad n \in \Omega. \quad (22)$$

From expressions (22), it follows that the parameters $a_n^{(1)}(T_1, T_2)$ are other than zero only if $n \in \Omega$.

When the problem in the third-order approximation in initial surface deformation is solved, $a_n^{(1)}(T_1, T_2)$ and $\tau_n^{(1)}(T_1, T_2)$ as functions of T_1 and T_2 are found from the condition that secular terms in the second- and third-order problems, respectively, vanish in view of initial conditions (9A).

Substituting expressions (16)–(21) into Eqs. (13A)–(18A) and eliminating the secular terms, we find that the functions $a_n^{(1)}(T_1, T_2)$ and $\tau_n^{(1)}(T_1, T_2)$ do not depend on the time scale T_1. The explicit dependences of second-order quantities on the time scale T_0 in view of (22) are given by

$$M_0^{(2)}(T_0) = -\sum_{m \in \Omega} \frac{(a_m^{(1)})^2 \cos^2(\omega_m T_0)}{2m+1};$$

$$M_1^{(2)}(T_0) = \sum_{m \in \Omega} \chi_m a_m^{(1)} a_{m+1}^{(1)} \cos(\omega_m T_0)\cos(\omega_{m+1} T_0);$$

$$M_n^{(2)}(T_0, T_1) = a_n^{(2)}(T_1)\cos(\omega_n T_0 + \tau_n^{(2)}(T_1)) + \sum_{l, m \in \Omega} \frac{a_l^{(1)} a_m^{(1)}}{2} (\lambda_{lmn}^{(+)} \cos((\omega_l + \omega_m) T_0) + \lambda_{lmn}^{(-)} \cos((\omega_l - \omega_m) T_0)); \quad (23)$$

$$F_0^{(2)} = 0; \quad F_n^{(2)}(T_0, T_1, T_2) = Q M_n^{(2)}(T_0, T_1) + Q \sum_{l, m \in \Omega} l K_{lmn} a_l^{(1)} a_m^{(1)} \cos(\omega_l T_0)\cos(\omega_m T_0); \quad n \geq 1; \quad (24)$$

$$D_n^{(2)}(T_0, T_1) = \frac{1}{n}\left\{ \partial_{T_0} M_n^{(2)}(T_0, T_1) + \sum_{l, m \in \Omega} (l(l-1) K_{lmn} - \alpha_{lmn}) \times \frac{\omega_l}{l} a_l^{(1)} a_m^{(1)} \sin(\omega_l T_0)\cos(\omega_m T_0) \right\}; \quad n \geq 1; \quad (25)$$

where χ_m, $\lambda_{lmn}^{(+)}$, $\lambda_{lmn}^{(-)}$, K_{lmn}, and α_{lmn} are the coefficients determined in Appendix B. The expressions for $a_n^{(2)}$ and $\tau_n^{(2)}$ that satisfy initial conditions (18A) have the

form

$$a_n^{(2)} = -\sum_{l,m\in\Omega}\frac{h_l h_m}{2}(\lambda_{lmn}^{(+)}+\lambda_{lmn}^{(-)}),\quad \tau_n^{(2)} = 0. \qquad (26)$$

Substituting (16)–(21) and (23)–(25) into Eqs. (22A)–(28A) and eliminating secular terms from the solutions, we find that the functions $a_n^{(1)}(T_2)$, $a_n^{(2)}(T_1)$, and $\tau_n^{(2)}(T_1)$ are independent of the time scales T_1 and T_2 and equal their initial values given by (22) and (26). For the function $\tau_n^{(1)}(T_2)$, we have

$$\begin{aligned}\tau_n^{(1)}(T_2) &= T_2 b_n\\ &= \frac{T_2}{2\omega_n}\Bigg\{\sum_{k\in\Omega}\frac{h_k^2\Xi_n}{2(2k+1)}+\frac{h_n^2(2(n-1)\omega_n^2+\Xi_n)}{4(2n+1)}\\ &-\frac{\chi_{n-1}h_{n-1}^2}{4}(\beta^{2(+)}_{n-1,n,1,n-1,n}+\beta^{2(-)}_{n-1,n,1,n-1,n})\\ &-\frac{\chi_n h_n^2}{4}(\beta^{1(-)}_{n+1,n+1,1,n,n}+\beta^{2(+)}_{n+1,n+1,1,n,n})\\ &-\sum_{k\in\Omega}\frac{h_k^2}{4}[H^{1(-)(+)}_{nkkn}+H^{2(+)(+)}_{knkn}+H^{2(-)(-)}_{knkn}+(1-\delta_{kn})\\ &\times(H^{1(-)(+)}_{kknn}+H^{2(+)(+)}_{kknn}+H^{2(-)(-)}_{nkkn})]\Bigg\}.\end{aligned} \qquad (27)$$

The coefficients of expansions (16)–(18) are defined as

$$\begin{aligned}M_0^{(3)}(T_0) &= -\sum_{k\in\Omega}\frac{2M_k^{(2)}(T_0)}{2k+1}h_k\cos(\omega_k T_0)\\ &-\sum_{k,m,l\in\Omega}\frac{K_{kml}h_k h_m h_l}{3(2l+1)}\cos(\omega_k T_0)\cos(\omega_m T_0)\cos(\omega_l T_0);\\ M_1^{(3)}(T_0) &= -\frac{6}{5}M_1^{(2)}(T_0)h_2\cos(\omega_2 T_0)\\ &-3\sum_{m\in\Omega}\sum_{k=0}^{\infty}K_{kml}M_k^{(2)}(T_0)h_m\cos(\omega_m T_0)\\ &-\sum_{g=0}^{\infty}\sum_{k,m,l\in\Omega}K_{kmg}K_{gl1}h_k h_m h_l\cos(\omega_k T_0)\\ &\times\cos(\omega_m T_0)\cos(\omega_l T_0);\end{aligned}$$

$$\begin{aligned}M_n^{(3)}(T_0) &= -\sum_{k\in\Omega}\frac{h_n h_k^2(2(n-1)\omega_n\omega_k-\Xi_n)}{16(2k+1)\omega_k(\omega_n+\omega_k)}\\ &\times(\cos((\omega_n+2\omega_k)T_0)-\cos(\omega_n T_0))\\ &-\sum_{k\in\Omega}\frac{h_n h_k^2(1-\delta_{nk})(2(n-1)\omega_n\omega_k+\Xi_n)}{16(2k+1)\omega_k(\omega_n-\omega_k)}\\ &\times(\cos((\omega_n-2\omega_k)T_0)+\cos(\omega_n T_0))\\ &+\sum_{k=n-1}^{n+1}\sum_{l\in\Omega}\frac{\chi_l h_k h_l h_{l+1}}{4}\\ &\times\Bigg\{\frac{\beta^{1(+)}_{k,l+1,1,l,n}(\cos(\psi^{(+)(+)}_{k,l,l+1}T_0)-\cos(\omega_n T_0))}{(\omega_n^2-(\omega_k+\omega_l+\omega_{l+1})^2)}\\ &+\frac{\beta^{1(-)}_{k,l+1,1,l,n}D^{ln}_{k,l+1}(\cos(\psi^{(+)(-)}_{k,l,l+1}T_0)-\cos(\omega_n T_0))}{(\omega_n^2-(\omega_k+\omega_l-\omega_{l+1})^2)}\\ &+\frac{\beta^{2(+)}_{k,l+1,1,l,n}D^{kl}_{l,n-1}D^{ln}_{k,l+1}(\cos(\psi^{(-)(-)}_{k,l,l+1}T_0)-\cos(\omega_n T_0))}{(\omega_n^2-(\omega_k-\omega_l-\omega_{l+1})^2)}\\ &+\frac{\beta^{2(-)}_{k,l+1,1,l,n}D^{kl}_{l,n-1}(\cos(\psi^{(+)(-)}_{k,l+1,l}T_0)-\cos(\omega_n T_0))}{(\omega_n^2-(\omega_k-\omega_l+\omega_{l+1})^2)}\Bigg\}\\ &-\sum_{g=2k}^{\infty}\sum_{m,l\in\Omega}\frac{h_k h_m h_l(\lambda^{(+)}_{lmg}+\lambda^{(-)}_{lmg})}{4}\\ &\times\Bigg\{\frac{H^{0(+)}_{kgn}(\cos((\omega_k+\omega_g)T_0)-\cos(\omega_n T_0))}{\omega_n^2-(\omega_k+\omega_g)^2}\\ &+\frac{H^{0(-)}_{kgn}(\cos((\omega_k-\omega_g)T_0)-\cos(\omega_n T_0))}{\omega_n^2-(\omega_k-\omega_g)^2}\Bigg\}\\ &+\sum_{k,m,l\in\Omega}\frac{h_k h_m h_l}{4}\Bigg\{\frac{H^{1(+)(-)}_{kmln}(\cos(\psi^{(+)(+)}_{klm}T_0)-\cos(\omega_n T_0))}{\omega_n^2-(\omega_k+\omega_l+\omega_m)^2}\\ &+\frac{H^{1(-)(+)}_{kmln}D^{kn}_{lm}D^{ln}_{km}(\cos(\psi^{(+)(-)}_{klm}T_0)-\cos(\omega_n T_0))}{\omega_n^2-(\omega_k+\omega_l-\omega_m)^2}\\ &+\frac{H^{2(+)(+)}_{kmln}D^{mn}_{kl}D^{ln}_{km}(\cos(\psi^{(-)(-)}_{klm}T_0)-\cos(\omega_n T_0))}{\omega_n^2-(\omega_k-\omega_l-\omega_m)^2}\\ &+\frac{H^{2(-)(-)}_{kmln}D^{mn}_{kl}D^{kn}_{ml}(\cos(\psi^{(+)(-)}_{klm}T_0)-\cos(\omega_n T_0))}{\omega_n^2-(\omega_k-\omega_l+\omega_m)^2}\Bigg\};\\ F_0^{(3)}(T_0) &= Q\sum_{k,m,l\in\Omega}\frac{k+1}{2l+l}\left(\alpha_{kml}-\frac{k(k+1)}{2}K_{kml}\right)\\ &\times h_k h_m h_l\cos(\omega_k T_0)\cos(\omega_m T_0)\cos(\omega_l T_0);\end{aligned} \qquad (28)$$

$$F_n^{(3)}(T_0) = QM_n^{(3)}(T_0) + \sum_{m\in\Omega}\sum_{k=1}^{\infty}(k+1)K_{kmn}F_k^{(2)}(T_0) \times h_m\cos(\omega_m T_0) + Q\sum_{m\in\Omega}\sum_{m=0}^{\infty}(k-1)K_{kmn}M_m^{(2)}(T_0) \times h_k\cos(\omega_k T_0) - Q\sum_{g=0}^{\infty}\sum_{k,m,l\in\Omega}\frac{k(k+3)}{2}K_{kmg}K_{gln} \times h_k h_m h_l\cos(\omega_k T_0)\cos(\omega_m T_0)\cos(\omega_l T_0); \quad n\geq 1; \tag{29}$$

$$D_n^{(3)}(T_0, T_2) = \frac{1}{n}\partial_{T_0}M_n^{(3)}(T_0) - \frac{1-\delta_{1n}}{n}h_n b_n\sin(\omega_n T_0) - \frac{1}{n}\sum_{m\in\Omega}\sum_{k=1}^{\infty}(k(k-1)K_{kmn} - \alpha_{kmn})D_k^{(2)}(T_0)h_m\cos(\omega_m T_0)$$

$$+\frac{1}{n}\sum_{k\in\Omega}\sum_{m=0}^{\infty}(k(k-1) - \alpha_{kmn})M_m^{(2)}(T_0)\omega_k h_k\sin(\omega_k T_0) \tag{30}$$

$$+\frac{1}{n}\sum_{k,m,l\in\Omega}\sum_{g=0}^{\infty}\left(\frac{k(k-1)}{2}K_{kmg} - \alpha_{kmg}\right)(k-2)K_{gln}\omega_k \times h_k h_m h_l\sin(\omega_k T_0)\cos(\omega_m T_0)\cos(\omega_l T_0), \quad n\geq 1.$$

Here, Ξ_n, $\beta_{kmgln}^{1(\pm)}$, $\beta_{kmgln}^{2(\pm)}$, $H_{kgn}^{0(\pm)}$, $H_{kmln}^{1(\pm)(\pm)}$, $H_{kmln}^{2(\pm)(\pm)}$, $\psi_{kml}^{(\pm)(\pm)}$, and D_{lm}^{kn} are the coefficients given in Appendix B and δ_{kn} is the Kronecker delta.

Substituting (18) into (1) yields an expression for the drop generatrix:

$$r(\vartheta, T_0, T_2) = 1 + \varepsilon\sum_{m\in\Omega}M_n^{(1)}(T_0, T_2)P_n(\cos\vartheta) + \varepsilon^2\sum_{n=0}^{\infty}(M_n^{(2)}(T_0) + \varepsilon M_n^{(3)}(T_0))P_n(\cos\vartheta). \tag{31}$$

(4) Prior to analyzing expression (31), we note that the amplitudes of deviation of the drop surface from the equilibrium (spherical) shape are proportional to the sums (see expressions (23) and (28))

$$M_g^{(2)} \sim \sum_{k,m\in\Omega}K_{kmg}, \quad M_n^{(3)} \sim \sum_{g=0}^{\infty}\sum_{k,m,l\in\Omega}K_{kmg}K_{gln},$$

where the coefficients K_{kmg} are other than zero only if $|k-m|\leq g\leq |k+m|$ and $k+m+g$ is an even number.

Thus, if only one mode is initially excited (that is, $\Omega = \{n_1\}$), only even modes with numbers from the range $0\leq g\leq 2n_1$ are excited in the second order of smallness. In the third order of smallness, even modes from the range $0\leq n\leq 3n_1$ and odd modes from the range $1\leq n\leq 3n_1$ are excited when n_1 is even and odd, respectively. Thus, with even n_1, the vibration of the surface consists of even modes from the range $[0, 3n_1]$; with odd n_1, it consists of all the modes from the range $[0, 2n_1]$ and odd modes from the range $[2n_1+1, 3n_1]$.

If two modes with the numbers n_1 and n_2 are initially excited (that is, $\Omega = \{n_1, n_2\}$), a set of modes involved in shaping the drop expands still further.

If n_1 and n_2 are even numbers, the spectrum of second-order modes comprises only even modes with indices from the range $0\leq g\leq \max\{2n_1, 2n_2\}$, while the third-order spectrum contains even modes with indices from the range $0\leq n\leq \max\{3n_1, 3n_2\}$. In other words, the drop surface is shaped by even modes from the range $[0, \max\{3n_1, 3n_2\}]$.

If the numbers n_1 and n_2 of initially excited modes are odd, even modes with numbers from the range $0\leq g\leq \max\{2n_1, 2n_2\}$ are excited in the second order of smallness. In the third order of smallness, only odd modes with numbers satisfying the condition $1\leq n\leq \max\{3n_1, 3n_2\}$ shape the drop surface. Accordingly, the surface is shaped by all modes from the range $[0, \max\{2n_1, 2n_2\}]$ and by modes with odd numbers from the interval $[\max\{2n_1+1, 2n_2+1\}, \max\{3n_1, 3n_2\}]$.

If the numbers of initially excited modes are such that n_1 is even and n_2 is odd, the second-order spectrum comprises modes with even numbers from the range $0\leq g\leq \max\{2n_1, 2n_2\}$ and odd modes with numbers satisfying the condition $|n_1-n_2|\leq g\leq n_1+n_2$. The third-order spectrum comprises even modes with numbers from the range $0\leq n\leq \max\{3n_1, n_1+2n_2\}$ and odd modes with numbers from $1\leq n\leq \max\{3n_2, 2n_1+n_2\}$. Eventually, the drop surface is shaped by even modes with numbers from the range $[0, \max\{3n_1, n_1+2n_2\}]$ and by odd modes with numbers from the interval $[1, \max\{3n_2, 2n_1+n_2\}]$. It is seen that taking into account quantities of the third order of smallness in initial deformation considerably expands the spectrum of modes shaping the drop surface.

Also, taking into account quantities of the third order of smallness leads to a nonlinear shift of the frequencies of initially excited modes that is proportional to the initial deformation amplitude squared ε^2. The sign of a frequency correction is always negative, and its value depends considerably on the spectrum of modes shaping the drop surface at the initial time

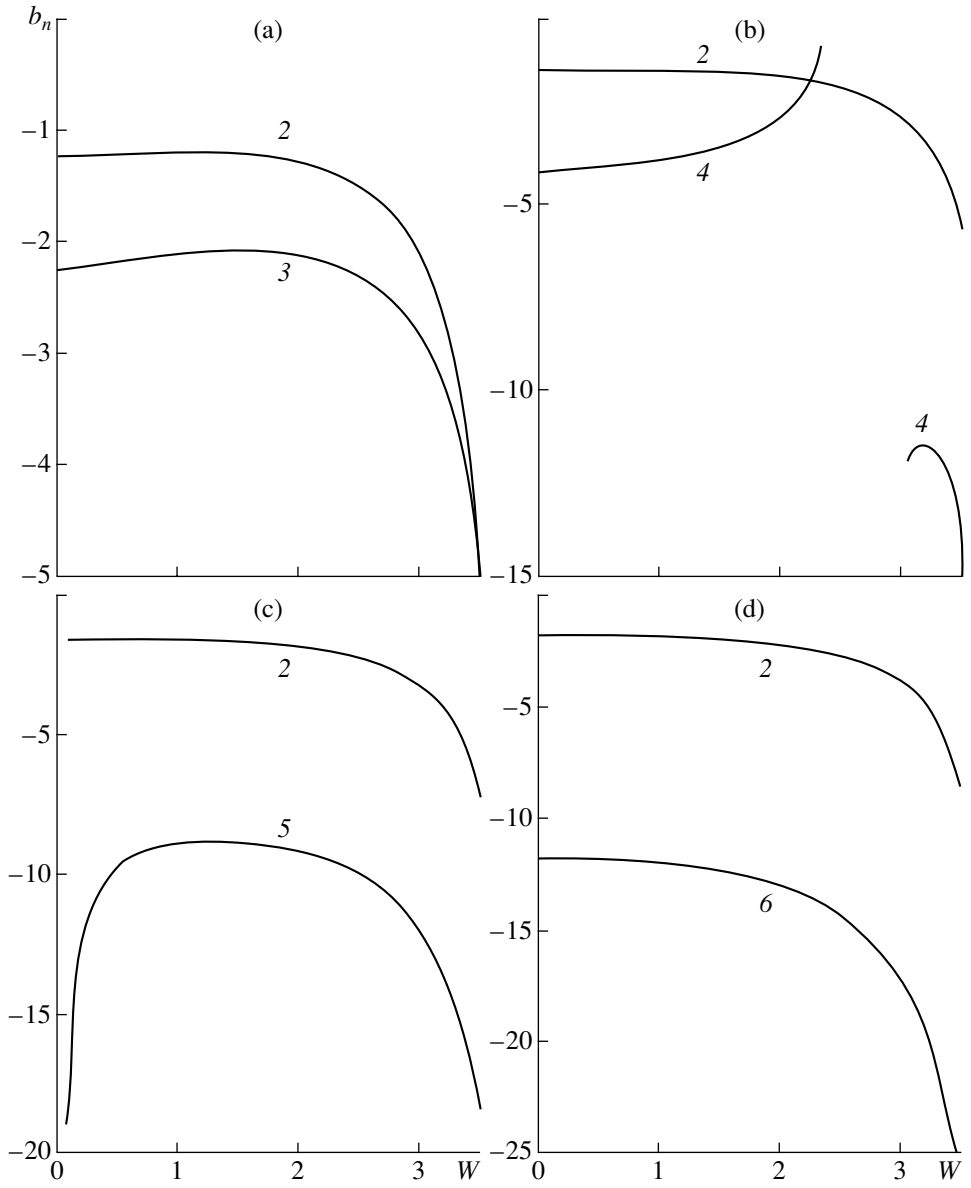


Fig. 1. Coefficient b_n as a function of the Rayleigh parameter $W = Q^2/4\pi$ for different pairs of initially excited modes: (a) the second and third modes, (b) the second and fourth modes, (c) the second and fifth modes, and (d) the second and sixth modes. The curve numbers coincide with the numbers of the initially excited modes.

instant and on the charge of the drop. For example, if two modes, one of which is the fundamental mode $n = 2$, are initially excited, frequency corrections increase compared with the situation of single-mode initial deformation [4]. Figure 1 plots frequency corrections for different pairs of modes excited at the initial time instant vs. the dimensionless parameter *W*. It is seen that the correction to the frequency of the fundamental mode depends on which of the modes is excited together with it at the initial time instant: the correction to the fundamental mode increases as the number of the mode excited simultaneously with the fundamental one grows. If it is remembered that the drop loses stability when the square of the fundamental mode frequency goes through zero as the parameter *W* grows [3, 9], it becomes clear that allowance for the nonlinear correction to the fundamental mode frequency reduces the critical value of the parameter *W* according to the relationship $\omega_2^2 + 2\varepsilon^2 b_2 = 0$ [9]. The higher the mode that is excited simultaneously with the fundamental one at the initial time instant, the larger the nonlinear correction to the critical condition for drop instability that follows from the above relationship.

A discontinuity of the curve for the correction to the fourth-mode frequency in Fig. 1b is associated with internal nonlinear resonant interaction between the fourth and sixth modes [4, 8, 9].

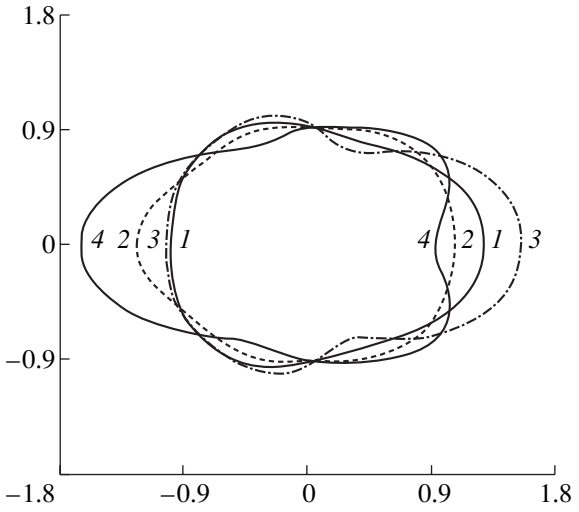


Fig. 2. Drop generatrix at the initial excitation of the second and third modes for $\varepsilon = 0.3$, $W = 3.7$, and $h_2 = h_3 = 0.5$. $t =$ (*1*) 0, (*2*) 1, (*3*) 3, and (*4*) 4.

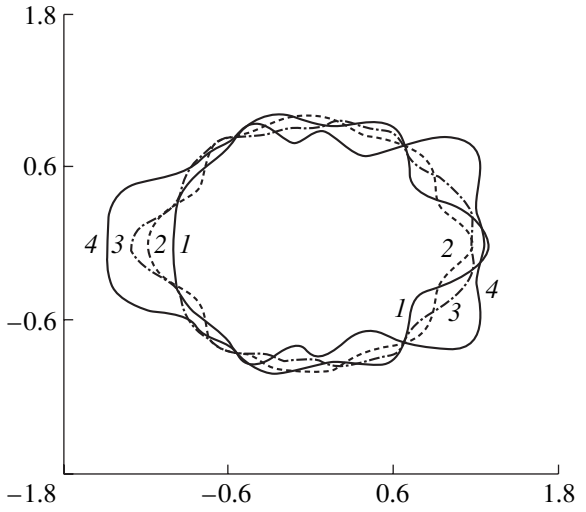


Fig. 3. Drop generatrix at the initial excitation of the sixth and seventh modes at $\varepsilon = 0.3$, $W = 3.4$, and $h_6 = h_7 = 0.5$. $t =$ (*1*) 0, (*2*) 0.5, (*3*) 1, and (*4*) 2.

Numerical analysis of expression (31) shows that surface elements adjacent to the symmetry axis deviate from equilibrium to the greatest extent (Figs. 2, 3). This is because individual modes add up only when ϑ is close to 0 and π. Away from these values, a smoother wavy surface is observed. This tendency is enhanced as the numbers of initially excited modes grow.

The electric field strength on the free surface of the drop is given by

$$E = E_n^{(0)} + \varepsilon \sum_{m \in \Omega} E_n^{(1)} P_n(\cos\vartheta) + \varepsilon^2 \sum_{n=0}^{\infty} (E_n^{(2)} + \varepsilon E_n^{(3)}) P_n(\cos\vartheta); \quad (32)$$

$$E_n^{(0)} = Q = 2\sqrt{\pi W}; \quad E_n^{(1)} = Q(n-1)M_n^{(1)};$$

$$E_n^{(2)} = (n+1)F_n^{(2)} - 2QM_n^{(2)} + Q \sum_{k,\, m \in \Omega} [(3-(m+1)(m+2))K_{kmn} + \alpha_{kmn}/2]h_k h_m \cos(\omega_k T_0)\cos(\omega_m T_0);$$

$$E_n^{(3)} = (n+1)F_n^{(3)} - 2QM_n^{(3)} + \sum_{\substack{k \in \Omega \\ m=0}}^{\infty} (\alpha_{kmn} - (m+1) \times (m+2)K_{kmn})h_k \cos(\omega_k T_0)F_m^{(2)} - Q \sum_{\substack{m=0 \\ k \in \Omega}}^{\infty} (k+4)(k-1) \times K_{kmn} h_k \cos(\omega_k T_0) M_m^{(2)} + Q \sum_{\substack{g=0 \\ k,\, m,\, n \in \Omega}}^{\infty} [((k+1)(k+2) \times (k+3)/2 - 4)K_{kmg} - ((l+1)/2 + k + l)\alpha_{kmg}]K_{gln} \times h_k h_m h_l \cos(\omega_k T_0)\cos(\omega_m T_0)\cos(\omega_l T_0).$$

Calculation by (32) shows that the self-charge field in the vicinity of the nonlinearly vibrating drop increases greatly at the poles when the drop is extended (Figs. 4, 5) and may trigger a corona discharge. This circumstance is of interest in the context of the initiation of lightning discharge [14, 15]. According to current concepts, lightning may originate from a corona near a large water-covered hailstone or water drop falling in a cloud. This mechanism of lightning initiation has not gained recognition, since the self-charge of drops that is detected in full-scale measurements (in storm clouds) is too small for a corona to be initiated at an undisturbed drop [16]. The fact that the electrostatic field at the poles of a nonlinearly vibrating drop is considerably enhanced allows one to consider the problem discussed from a new standpoint.

The calculations shown in Figs. 2–5 were made in the absence of mode resonant interaction. This issue calls for special consideration [17]. Nevertheless, resonance energy exchange between modes is a possibility.

It is easy to check that expressions (28) for third-order nonlinear corrections $M_n^{(3)}(t)$ to the vibration amplitudes have a resonance form: their denominators vanish under certain conditions. All resonances other than those obtained in the quadratic approximation [4, 8, 9] correspond to the four-mode interaction between drop capillary vibrations when the frequencies

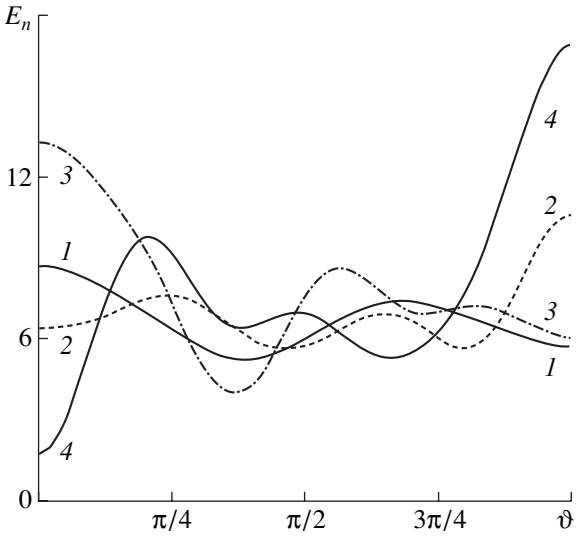


Fig. 4. Electric field strength E near the drop surface as a function of the polar angle ϑ for the same parameter values as in Fig. 2.

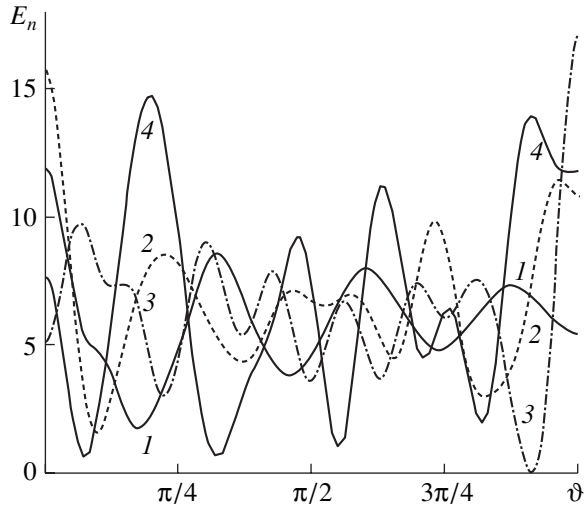


Fig. 5. Electric field strength E near the drop surface as a function of the polar angle ϑ for the same parameter values as in Fig. 3.

of resonantly interacting modes are related to each other by one of the relationships

$$\omega_n \pm \omega_k \pm \omega_l \pm \omega_m = 0.$$

Among the many internal nonlinear resonances taking place in the charged drop, those where the fundamental mode ($n = 2$) amplitude increases by means of energy transfer from higher modes at Rayleigh parameters $W < 4$ (which are subcritical in terms of self-charge) are of greatest interest for the problem of lighting initiation in storm clouds [14, 15]. According to second-order calculations [6, 8, 17, 18], when only three-mode resonances are realized, the lowest mode that gains energy from higher modes via resonant interaction is the third mode. In third-order calculations, when four-mode interaction occurs, the second mode also may resonantly build up. For example, if

$$\omega_n + \omega_k - \omega_l - \omega_m = 0; \quad W \le 4,$$

more than ten four-mode resonances take place in the range of mode numbers $2 \le n, k, l, m \le 30$ and seven of them include the second mode. The first ten of the possible four-mode resonances are listed in the table. It is seen that the first three are truly quadrimodal, while the rest of them are degenerate: one of the modes is involved in four-mode interaction twice.

Numbers of modes involved in resonances and the associated Rayleigh parameter

n	l	k	m	W
2	20	8	17	1.98141
2	29	12	24	1.39884
2	30	17	21	0.460245
2	9	6	6	0.0460245
2	17	11	11	1.35905
2	25	16	16	1.42339
2	28	18	18	2.9609
3	23	15	15	2.18618
3	28	18	18	0.450789
4	26	17	17	0.577818

CONCLUSIONS

Considering the initial multimode deformation of the drop in the third-order approximation in deformation amplitude allows one to obtain nonlinear corrections to the capillary vibration frequencies of the drop. These corrections depend considerably on the charge of the drop and on the spectrum of initially excited modes and, in turn, generate nonlinear corrections to the critical Rayleigh instability parameter. In calculating the generatrix of a nonlinearly vibrating drop, taking account of the third-order quantities makes it possible to discover the tendency of the drop to extend along its symmetry axis. This is an indirect indication that a large number of modes shape emitting protrusions on the drop surface [19].

APPENDIX A

Separation of Problems of Different Orders of Smallness

Substituting expansions (13)–(15) into the set of Eqs. (1)–(12) and collecting terms proportional to ε^1,

one readily obtains the first-order problem:

$$\Delta\Psi^{(1)} = 0; \quad \Delta\phi^{(1)} = 0; \tag{1A}$$

$$r \longrightarrow 0\colon \psi^{(1)} \longrightarrow 0; \tag{2A}$$

$$r \longrightarrow +\infty\colon \nabla\phi^{(1)} \longrightarrow 0; \tag{3A}$$

$$r = 1\colon \partial_{T_0}\xi^{(1)} = \partial_r\psi^{(1)}; \tag{4A}$$

$$\partial_{T_0}\psi^{(1)} = \frac{1}{4\pi}\partial_r\phi^{(0)}(\partial_r\phi^{(1)} + \xi^{(1)}\partial_{rr}\phi^{(0)}) + 2\xi^{(1)} + \Delta_\Omega\xi^{(1)}; \tag{5A}$$

$$\int_{-1}^{1}\xi^{(1)}d(\cos\vartheta) = 0; \quad \int_{-1}^{1}\xi^{(1)}P_1 d(\cos\vartheta) = 0; \tag{6A}$$

$$\int_{-1}^{1}\{\partial_r\phi^{(1)} + \xi^{(1)}(\partial_{rr}\phi^{(0)} + 2\partial_r\phi^{(0)})\}d(\cos\vartheta) = 0; \tag{7A}$$

$$\phi^{(1)} + \xi^{(1)}\partial_r\phi^{(0)} = \phi_S^{(1)}(t); \tag{8A}$$

$$t = 0\colon \xi^{(1)} = \varepsilon\sum_{m\in\Omega} h_m P_m(\cos\vartheta); \quad \partial_{T_0}\xi^{(1)} = 0. \tag{9A}$$

Terms proportional to ε^2 state the second-order problem:

$$\Delta\psi^{(2)} = 0; \quad \Delta\phi^{(2)} = 0; \tag{10A}$$

$$r \longrightarrow 0\colon \psi^{(2)} \longrightarrow 0; \tag{11A}$$

$$r \longrightarrow +\infty\colon \nabla\phi^{(2)} \longrightarrow 0; \tag{12A}$$

$$r = 1\colon \tag{13A}$$

$$\partial_{T_0}\xi^{(2)} + \partial_{T_1}\xi^{(1)} = \partial_r\psi^{(2)} + \xi^{(1)}\partial_{rr}\psi^{(1)} - \partial_\vartheta\xi^{(1)}\partial_\vartheta\psi^{(1)};$$

$$\partial_{T_0}\psi^{(2)} + \partial_{T_1}\psi^{(1)} + \xi^{(1)}\partial_{rT_0}\psi^{(1)} + \frac{1}{2}(\partial_r\psi^{(1)})^2 + \frac{1}{2}(\partial_\vartheta\psi^{(1)})^2 = \frac{1}{8\pi}\{2\xi^{(2)}\partial_r\phi^{(0)}\partial_{rr}\phi^{(0)} + (\xi^{(1)})^2((\partial_{rr}\phi^{(0)})^2 + \partial_{rrr}\phi^{(0)}\partial_r\phi^{(0)}) + (\partial_\vartheta\phi^{(1)})^2 + (\partial_r\phi^{(1)})^2 + 2\partial_r\phi^{(2)}\partial_r\phi^{(0)} + 2\xi^{(1)}(\partial_{rr}\phi^{(0)}\partial_r\phi^{(1)} + \partial_{rr}\phi^{(1)}\partial_r\phi^{(0)})\} + 2\xi^{(2)} + \Delta_\Omega\xi^{(2)} - 2(\xi^{(1)})^2 - 2\xi^{(1)}\Delta_\Omega\xi^{(1)}; \tag{14A}$$

$$\int_{-1}^{1}(\xi^{(2)} + (\xi^{(1)})^2)d(\cos\vartheta) = 0;$$
$$\int_{-1}^{1}(2\xi^{(2)} + 3(\xi^{(1)})^2)P_1 d(\cos\vartheta) = 0; \tag{15A}$$

$$\int_{-1}^{1}\left\{\partial_r\phi^{(2)} + \xi^{(1)}(\partial_{rr}\phi^{(1)} + 2\partial_r\phi^{(1)}) + \xi^{(2)}(\partial_{rr}\phi^{(0)} + 2\partial_r\phi^{(0)}) + (\xi^{(1)})^2\left(\frac{1}{2}\partial_{rrr}\phi^{(0)} + 2\partial_{rr}\phi^{(0)} + \partial_r\phi^{(0)}\right) - \partial_\vartheta\xi^{(1)}\partial_\vartheta\phi^{(1)}\right\} \tag{16A}$$
$$\times d(\cos\vartheta) = 0;$$

$$\phi^{(2)} + \xi^{(1)}\partial_r\phi^{(1)} + \xi^{(2)}\partial_r\phi^{(0)} + \frac{1}{2}(\xi^{(1)})^2\partial_{rr}\phi^{(0)} = \phi_S^{(2)}(t); \tag{17A}$$

$$t = 0\colon$$

$$\xi^{(2)} = -\sum_{m\in\Omega}\frac{h_m P_0(\cos\vartheta)}{2m+1} - \frac{3}{2}\sum_{l,m\in\Omega} h_l h_m K_{lml} P_1(\cos\vartheta); \tag{18A}$$

$$\partial_{T_0}\xi^{(2)} + \partial_{T_1}\xi^{(1)} = 0.$$

The third-order problem is defined by terms proportional to ε^3 and has the form

$$\Delta\psi^{(3)} = 0; \quad \Delta\phi^{(3)} = 0; \tag{19A}$$

$$r \longrightarrow 0\colon \psi^{(3)} \longrightarrow 0; \tag{20A}$$

$$r \longrightarrow +\infty\colon \nabla\phi^{(3)} \longrightarrow 0; \tag{21A}$$

$$r = 1\colon$$

$$\partial_{T_0}\xi^{(3)} + \partial_{T_1}\xi^{(2)} + \partial_{T_2}\xi^{(1)} = \partial_r\psi^{(3)} - \partial_\vartheta\xi^{(2)}\partial_\vartheta\psi^{(1)} - \partial_\vartheta\xi^{(1)}\partial_\vartheta\psi^{(2)} + \xi^{(2)}\partial_{rr}\psi^{(1)} + \xi^{(1)}(\partial_\vartheta\xi^{(1)}(2\partial_\vartheta\psi^{(1)} - \partial_{r\vartheta}\psi^{(1)}) + \partial_{rr}\psi^{(2)}) + \frac{1}{2}(\xi^{(1)})^2\partial_{rrr}\psi^{(1)}; \tag{22A}$$

$$\partial_{T_0}\psi^{(3)} + \partial_{T_2}\psi^{(1)} + \partial_{T_1}\psi^{(2)} + \xi^{(1)}\partial_{rT_1}\psi^{(1)} + \partial_\vartheta\psi^{(1)}\partial_\vartheta\psi^{(2)} + \partial_r\psi^{(1)}\partial_r\psi^{(2)} + \xi^{(2)}\partial_{rT_0}\psi^{(1)} + \xi^{(1)}(\partial_{rT_0}\psi^{(2)} + \partial_\vartheta\psi^{(1)}(\partial_{r\vartheta}\psi^{(1)} - \partial_\vartheta\psi^{(1)}) + \partial_r\psi^{(1)}\partial_{rr}\psi^{(1)}) + \frac{1}{2}(\xi^{(1)})^2\partial_{rrT_0}\psi^{(1)} = \frac{1}{8\pi}\Big\{2\xi^{(3)}\partial_r\phi^{(0)}\partial_{rr}\phi^{(0)}$$

$$+(\xi^{(1)})^3\left(\partial_{rr}\phi^{(0)}\partial_{rrr}\phi^{(0)}+\frac{1}{3}\partial_r\phi^{(0)}\partial_{rrrr}\phi^{(0)}\right)$$
$$+2(\partial_\vartheta\phi^{(1)}\partial_\vartheta\phi^{(2)}+\partial_r\phi^{(1)}(\xi^{(2)}\partial_{rr}\phi^{(0)}+\partial_r\phi^{(2)})$$
$$+\partial_r\phi^{(0)}\partial_r\phi^{(3)}+\xi^{(2)}\partial_r\phi^{(0)}\partial_{rr}\phi^{(1)}) \quad (23A)$$
$$+2\xi^{(1)}(\xi^{(2)}((\partial_{rr}\phi^{(0)})^2+\partial_r\phi^{(0)}\partial_{rrr}\phi^{(0)})+\partial_{rr}\phi^{(0)}\partial_r\phi^{(2)}$$
$$+\partial_\vartheta\phi^{(1)}(\partial_{r\vartheta}\phi^{(1)}-\partial_\vartheta\phi^{(1)})+\partial_r\phi^{(1)}\partial_{rr}\phi^{(1)}$$
$$+\partial_r\phi^{(0)}\partial_{rr}\phi^{(2)})+(\xi^{(1)})^2(\partial_{rrr}\phi^{(0)}\partial_r\phi^{(1)}+2\partial_{rr}\phi^{(0)}\partial_{rr}\phi^{(1)}$$
$$+\partial_r\phi^{(0)}\partial_{rrr}\phi^{(1)})\Big\}+(2+\Delta_\Omega)\xi^{(3)}+2\xi^{(1)}((\xi^{(1)})^2$$
$$-(2+\Delta_\Omega)\xi^{(2)})-2\xi^{(2)}\Delta_\Omega\xi^{(1)}+3(\xi^{(1)})^2\Delta_\Omega\xi^{(1)}$$
$$-(\partial_\vartheta\xi^{(1)})^2\partial_{\vartheta\vartheta}\xi^{(1)}-\frac{1}{2}(\partial_\vartheta\xi^{(1)})^2\Delta_\Omega\xi^{(1)};$$

$$\int_{-1}^{1}(3\xi^{(3)}+6\xi^{(1)}\xi^{(2)}+(\xi^{(1)})^3)d(\cos\vartheta)=0; \quad (24A)$$

$$\int_{-1}^{1}(\xi^{(3)}+3\xi^{(1)}\xi^{(2)}+(\xi^{(1)})^3)P_1(\cos\vartheta)d(\cos\vartheta)=0; \quad (25A)$$

$$\int_{-1}^{1}\Big\{\partial_r\phi^{(3)}+\xi^{(3)}(\partial_{rr}\phi^{(0)}+2\partial_r\phi^{(0)})+\xi^{(2)}(\partial_{rr}\phi^{(1)}+2\partial_r\phi^{(1)})$$
$$+(\xi^{(1)})^3\left(\frac{1}{6}\partial_{rrrr}\phi^{(0)}+\partial_{rrr}\phi^{(0)}+\partial_{rr}\phi^{(0)}\right)$$
$$+(\xi^{(1)})^2\left(\frac{1}{2}\partial_{rrr}\phi^{(1)}+2\partial_{rr}\phi^{(1)}+\partial_r\phi^{(1)}\right) \quad (26A)$$
$$+\xi^{(1)}(\xi^{(2)}(\partial_{rrr}\phi^{(0)}+4\partial_{rr}\phi^{(0)}+2\partial_r\phi^{(0)})+2\partial_r\phi^{(2)}$$
$$+\partial_{rr}\phi^{(2)}-\partial_\vartheta\xi^{(1)}\partial_{r\vartheta}\phi^{(1)})-\partial_\vartheta\xi^{(2)}\partial_\vartheta\phi^{(1)}$$
$$-\partial_\vartheta\xi^{(1)}\partial_\vartheta\phi^{(2)}\Big\}d(\cos\vartheta)=0;$$

$$\phi^{(3)}+\xi^{(1)}\partial_r\phi^{(2)}+\xi^{(2)}\partial_r\phi^{(1)}+\xi^{(3)}\partial_r\phi^{(0)}+\frac{1}{2}(\xi^{(1)})^2\partial_{rr}\phi^{(1)}$$
$$+\xi^{(1)}\xi^{(2)}\partial_{rr}\phi^{(0)}+\frac{1}{6}(\xi^{(1)})^3\partial_{rrr}\phi^{(0)}=\phi_S^{(3)}(t); \quad (27A)$$

$$t=0:\ \xi^{(3)}=-\sum_{k,m,l\in\Omega}\frac{h_kh_mh_l}{3(2l+1)}K_{kml}P_0(\cos\vartheta)$$
$$-\left(\frac{9}{5}h_2\sum_{k,m\in\Omega}h_kh_mK_{kml}\right.$$
$$\left.+\sum_{g=0}^{\infty}\sum_{k,m,l\in\Omega}h_kh_mh_lK_{kmg}K_{gl1}\right)P_1(\cos\vartheta); \quad (28A)$$
$$\partial_{T_0}\xi^{(3)}+\partial_{T_1}\xi^{(2)}+\partial_{T_2}\xi^{(1)}=0.$$

Here, $K_{mln}=(C_{m0l0}^{n0})^2$ and C_{m0l0}^{n0} are the Clebsch–Gordan coefficients [20].

APPENDIX B

Expressions for the Coefficients of the Problem

$$H_{kmln}^{1(+)(-)}=\sum_{g=2}^{\infty}\beta_{kmgln}^{1(+)}\lambda_{lmg}^{(+)}+\sum_{g=1}^{\infty}\mu_{kmgln}^{1(-)}+\sum_{g=0}^{\infty}\mu_{kmgln}^{0(-)};$$
$$H_{kmln}^{1(-)(+)}=\sum_{g=2}^{\infty}\beta_{kmgln}^{1(-)}\lambda_{lng}^{(-)}+\sum_{g=1}^{\infty}\mu_{kmgln}^{1(+)}+\sum_{g=0}^{\infty}\mu_{kmgln}^{0(+)};$$
$$H_{kmln}^{2(+)(+)}=\sum_{g=2}^{\infty}\beta_{kmgln}^{2(+)}\lambda_{lmg}^{(+)}+\sum_{g=1}^{\infty}\mu_{kmgln}^{1(+)}+\sum_{g=0}^{\infty}\mu_{kmgln}^{0(+)};$$
$$H_{kmln}^{2(-)(-)}=\sum_{g=2}^{\infty}\beta_{kmgln}^{2(-)}\lambda_{lng}^{(-)}+\sum_{g=1}^{\infty}\mu_{kmgln}^{1(-)}+\sum_{g=0}^{\infty}\mu_{kmgln}^{0(-)};$$
$$H_{mgn}^{0(+)}=(\Pi_{mgn}^0-\Pi_{mgn}^1\omega_m\omega_g-\Pi_{mgn}^2\omega_g^2)(\lambda_{mmg}^{(+)}+\lambda_{mmg}^{(-)});$$
$$H_{mgn}^{0(-)}=(\Pi_{mgn}^0+\Pi_{mgn}^1\omega_m\omega_g-\Pi_{mgn}^2\omega_g^2)(\lambda_{mmg}^{(+)}+\lambda_{mmg}^{(-)});$$
$$\beta_{kmgln}^{1(+)}=\Pi_{kgn}^0-\Pi_{kgn}^1\omega_k(\omega_l+\omega_m)-\Pi_{kgn}^2(\omega_l+\omega_m)^2;$$
$$\beta_{kmgln}^{1(-)}=\Pi_{kgn}^0-\Pi_{kgn}^1\omega_k(\omega_l-\omega_m)-\Pi_{kgn}^2(\omega_l-\omega_m)^2;$$
$$\beta_{kmgln}^{2(+)}=\Pi_{kgn}^0+\Pi_{kgn}^1\omega_k(\omega_l+\omega_m)-\Pi_{kgn}^2(\omega_l+\omega_m)^2;$$
$$\beta_{kmgln}^{2(-)}=\Pi_{kgn}^0+\Pi_{kgn}^1\omega_k(\omega_l-\omega_m)-\Pi_{kgn}^2(\omega_l-\omega_m)^2;$$
$$\mu_{kmgln}^{1(-)}=\Lambda_{kmgln}^1-\Gamma_{kmgln}^1\omega_m\omega_k;$$
$$\mu_{kmgln}^{1(+)}=\Lambda_{kmgln}^1+\Gamma_{kmgln}^1\omega_m\omega_k;$$
$$\mu_{kmgln}^{0(-)}=\Lambda_{kmgln}^0-\Gamma_{kmgln}^0\omega_m\omega_k;$$
$$\mu_{kmgln}^{0(+)}=\Lambda_{kmgln}^1+\Gamma_{kmgln}^0\omega_m\omega_k;$$
$$\Lambda_{kmgln}^0=\frac{1}{2k}\{K_{gln}(\alpha_{kmg}(kn(l+3l^2-2(k+2)W)$$
$$+2(k-2)\omega_k^2)+K_{kmg}(kn(4-6k(k+1)$$
$$+(k^3-2(m+1)(m+2)-k^2(n-9)$$

$$-k(3n+2m(m+3)-22))W)-(k-1)k(k-n-2)\omega_k^2))$$
$$-2kn\alpha_{kmg}\sum_{\nu=1}^{[l/2]}(2l-4\nu+1)K_{g,\,l-2\nu,\,n}\};$$

$$\Lambda^1_{kmgln}=((g-n-1)K_{gln}-\alpha_{gln}/g)((m-1)K_{kmg}-\alpha_{kmg}/m)$$
$$\times\omega_m^2+Wnk((g+1)(l+n-g-2)K_{gln}+\alpha_{gln})K_{kmg};$$

$$\Gamma^0_{kmgln}=((k-1)(k-2(n+1))K_{kmg}/2$$
$$-((k-1)(m+n)-m)\alpha_{kmg}/(km))K_{gln}$$
$$+((k-1)(k-2)K_{klg}/2-(k-2)\alpha_{klg}/k)K_{gmn};$$

$$\Gamma^1_{kmgln}=-((g-n-1)K_{gkn}-(n+k)\alpha_{gkn}/(kg))$$
$$\times((m-1)K_{lmg}-\alpha_{lmg}/m)-((g-n-1)K_{gln}-\alpha_{gln}/g)$$
$$\times((m-1)K_{kmg}-\alpha_{kmg}/m);$$

$$\Pi^0_{kmn}=(\omega_k^2(n-k+1)+2kn(k+1)+2mn(m+1)-4n$$
$$+nW((n-k-5)(k-1)+(m+1)(k+n-m-2)))K_{kmn}$$
$$+(\omega_k^2/k+nW)\alpha_{kmn};$$

$$\Pi^1_{kmn}=(m+k-n-2)K_{kmn}-(n+k+m)\alpha_{kmn}/(mk);$$

$$\Pi^2_{kmn}=(m-n-1)K_{kmn}-\alpha_{kmn}/m;$$

$$\Xi_k=\omega_k^2+2k^2(k+1)-4k-5k(k-1)W;$$

$$\psi^{(+)(+)}_{kml}=\omega_k+\omega_m+\omega_l,\quad \psi^{(+)(-)}_{kml}=\omega_k+\omega_m-\omega_l;$$

$$\psi^{(-)(-)}_{kml}=\omega_k-\omega_m-\omega_l;$$

$$\lambda^{(\pm)}_{mln}=(\gamma_{mln}\pm\omega_m\omega_l\eta_{mln})/(\omega_n^2-(\omega_m\pm\omega_l)^2);$$

$$\alpha_{mln}=-C^{n0}_{m0l0}C^{n0}_{m(-1)l1}\sqrt{m(m+1)l(l+1)};$$

$$\gamma_{mln}=K_{mln}[\omega_m^2(n-m+1)+2n(l(l+1)-1)+(l(m+1)$$
$$-m(2m-2n+7)+3)nW/2]+\alpha_{mln}[\omega_m^2/m+nW/2];$$

$$\eta_{mln}=K_{mln}(n/2-m+1)+\alpha_{mln}(1+n/(2l))/m;$$

$$D^{kn}_{lm}=1-\delta_{lm}\delta_{kn}.$$

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (grant no. 03-01-00760).

REFERENCES

1. A. G. Bailey, Atomization Spray Technol. **2**, 95 (1986).
2. V. G. Dudnikov and A. L. Shabalin, Preprint No. 87-63 (Institute of Nuclear Physics, Siberian Division, USSR Academy of Sciences, Novosibirsk, 1987).
3. A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 3 (1994).
4. J. A. Tsamopolous and R. A. Brown, J. Fluid Mech. **127**, 519 (1983).
5. T. G. Wang, A. V. Anilkuma, and C. P. Lee, J. Fluid Mech. **308**, 1 (1996).
6. S. O. Shiryaeva, Zh. Tekh. Fiz. **71** (2), 27 (2001) [Tech. Phys. **46**, 158 (2001)].
7. S. O. Shiryaeva, Pis'ma Zh. Tekh. Fiz. **26** (22), 76 (2000) [Tech. Phys. Lett. **26**, 1016 (2000)].
8. S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 173 (2001).
9. A. N. Zharov, S. O. Shiryaeva, and A. I. Grigor'ev, Zh. Tekh. Fiz. **73** (6), 36 (2003) [Tech. Phys. **48**, 697 (2003)].
10. A. N. Zharov, A. I. Grigor'ev, and S. O. Shiryaeva, Pis'ma Zh. Tekh. Fiz. **29** (9), 75 (2003) [Tech. Phys. Lett. **29**, 388 (2003)].
11. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).
12. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).
13. A.-H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973; Mir, Moscow, 1976).
14. V. A. Dyachuk and V. M. Muchnik, Dokl. Akad. Nauk SSSR **248**, 60 (1979).
15. A. I. Grigor'ev and S. O. Shiryaeva, Phys. Scr. **54**, 660 (1996).
16. *Clouds and Cloudy Atmosphere: A Handbook*, Ed. by I. P. Mazin and A. Khrgian (Gidrometeoizdat, Leningrad, 1989).
17. S. O. Shiryaeva, Zh. Tekh. Fiz. **73** (2), 19 (2003) [Tech. Phys. **48**, 152 (2003)].
18. S. O. Shiryaeva, A. I. Grigor'ev, and D. F. Belonozhko, Pis'ma Zh. Tekh. Fiz. **28** (22), 45 (2002) [Tech. Phys. Lett. **28**, 945 (2002)].
19. A. I. Grigor'ev and S. O. Shiryaeva, Zh. Tekh. Fiz. **61** (3), 19 (1991) [Sov. Phys. Tech. Phys. **36**, 258 (1991)].
20. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Sci., Singapore, 1988).

Translated by N. Mende

Technical Physics, Vol. 48, No. 12, 2003, pp. 1522–1526. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 20–25.

GASES AND LIQUIDS

Transition Flow of Rarefied Gases: Current Models and Methods of Simulation

Yu. M. Pechatnikov

St. Petersburg State Polytechnical University, ul. Politekhnicheskaya 29, St. Petersburg, 195297 Russia

e-mail: yiriy@nm.ru

Received March 11, 2003; in final form, June 10, 2003

Abstract—Transition (molecular–viscous) isothermal channel flow of rarefied gases is considered. Present-day engineering physical models of transition gas flow and methods of simulation are analyzed and verified in terms of the kinetic theory on the micro- and macrolevels. © *2003 MAIK "Nauka/Interperiodica".*

Analysis of evacuation under a medium vacuum, which may be considered as a division of isothermal channel dynamics of rarefied gas, presents considerable difficulties. The basic problem in the description of the transition channel flow of rarefied gases is to match the random walk of molecules in the molecular regime with the laminar flow in the viscous regime.

The problem of qualitative and quantitative analysis of the transition flow of rarefied gas in the molecular–viscous range was stated by Knudsen and other scientists as early as 1910.

The available models of rarefied gas transition flow are based on experimental data for flows in openings, long rectangular ducts, or circular pipes. The choice of such configurations was dictated by the fact that, in the case of openings, interaction of the gas with the metal surface can be neglected, while for long channels, boundary conditions may be known or given. Numerical solutions to the Boltzmann equation with a collisional term that were obtained for Knudsen numbers Kn < 100 coincide with measurements within 5%. The hydraulic conductivity (hereafter, conductivity for brevity) of openings and short pipes increases monotonically with average pressure P_{av} throughout the transition flow range [1–6] (see Fig. 1, where J_{mv} is the ratio of the conductivities in the molecular–viscous and molecular regimes). At the boundary between the molecular–viscous and viscous regimes (Kn ≅ 0.01), the conductivities of an opening, short pipe, and long pipe of equal diameter tend to the same value (Figs. 2 and 3, where $\delta = 1/\mathrm{Kn}$) [6]. As Kn decreases from 100 to 1, the conductivity of long pipes decreases and then, with Kn diminishing from 1 to 0.01, monotonically rises (the well-known Knudsen paradox [6]), as demonstrated in Fig. 3.

The presence of a minimal conductivity for long channels and its absence in the case of openings has become a key question in developing a physical model of flow for arbitrary configurations of vacuum systems. A change of boundary conditions, even in simple cases, created insurmountable obstacles in taking the collision integral by integral kinetic techniques [4–7].

Today, the dynamics of rarefied gas is usually analyzed by simulating the physical process of momentum transfer for determinate molecules. It is assumed that the motion of molecules in rarefied gases is a random process obeying the laws of statistical physics. Random quantities and statistical errors of calculation are generated by the Monte Carlo method.

The fundamental difficulty in simulating the molecular–viscous regime is that a model must be adequate for the actual physical process of molecule collision both in rarefied flows and in continuous flows where the gas concentration reaches several moles. Another problem is that our knowledge of the physics of a medium vacuum and the behavior (interaction) of determinate molecules on the microlevel is approximate and superficial. It is difficult to simulate random quantities of a physical process, such as the velocity, collision rate,

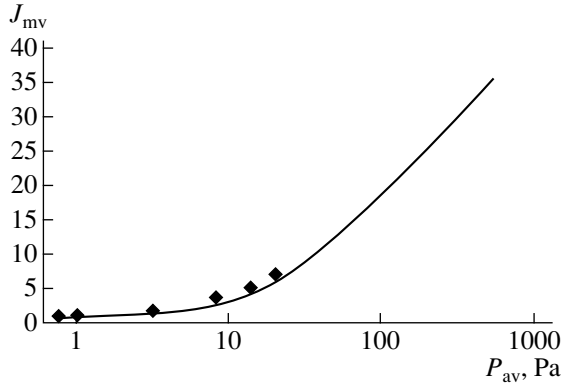


Fig. 1. Dimensionless conductivity J_{mv} vs. average pressure P_{av} for a 140-mm-long circular step adapter with inlet and outlet diameters of 400 mm and 25 mm, respectively, in the molecular–viscous regime. Solid line, experimental data; symbols, calculation by the method of probable directions.

1063-7842/03/4812-1522$24.00

and direction of motion of molecules, especially in many-particle collisions.

To simulate the flow in the near-molecular range, one may apply the Monte Carlo method (the probe particle method) if not only collisions with the inner surface but also primary intermolecular collisions are considered [4–6]. However, with Kn < 10, the analysis cannot be restricted to primary collisions and becomes much more complicated [7].

A new concept of the probe particle method was suggested by Haviland [8]. In his modification, the trajectory of one particle through the entire flow field is traced in view of pair elastic collisions with other particles distributed over the flow field. Uncertainty in the molecule distribution over the flow field somewhat detracts from the merit of this approach. Haviland's algorithm was applied in [9] to calculate the conductivity of short glass capillaries through which helium or argon flows with Kn > 0.3 under the assumption of mirror–diffuse reflection of molecules from the capillary wall. The authors of [9] report the gas flow rates for the capillaries and note that, as the molecular collision rate increases, the conductivity grows more slowly, while the fraction of mirror-reflected molecules rises. This fact is consistent with the results of [10], where the need for detailed consideration of the slip theory [11] was argued. Satisfactory qualitative agreement between computing and full-scale experiments in the limiting cases (opening and long pipe) was obtained in [9, 12]. The discrepancy observed in those works is explained by the insufficient accuracy of the method used and of the computational scheme, which is illustrated by a solution to the problem of gas self-diffusion in a short channel [12, 13].

For Kn between 100 and 0.1, the direct simulation method is widely used [14–16]. This method has been verified for a variety of problems [17, 18]. Note that the applicability domain of Bird's method is restricted to those Knudsen numbers at which only pair collisions are observed. It was shown [10] that pair intermolecular collisions dominate and, accordingly, Bird's method applies in the range $0.01 < \delta < 0.50$ ($\delta = 1/\mathrm{Kn}$). With $\delta > 0.5$, intermolecular collisions are not all pair; however, at $0.5 < \delta < 10$, the number of group collisions is relatively low (less than 10% of the total number of collisions). Hence, the accuracy of Bird's method remains sufficient for engineering purposes. For even lower Knudsen numbers, pair collisions constitute a minor part, so that one should take into account collective molecular interaction. At present, direct statistical simulation as a method of numerical calculation attracts little attention, although its accuracy (compared with a solution to the Boltzmann equation) and applicability to problems concerned with dynamics of rarefied gas remain unclear.

A probabilistic approach (the method of probable directions) to simulating the transition flow of rarefied gas has been suggested in [13]. This method is based on a physical engineering model of transition flow the basic advantage of which is that it is consistent with the concept of random walk of molecules in the molecular regime and of laminar flow in the viscous regime. In this model, it is assumed that the steady flow of a rarefied gas is the superposition of the molecular flow randomized by many-particle interaction and the molecular flow in the direction of the concentration gradient (directional flow) [19]. A gas flow is viewed as a cooperative motion of statistically independent similar determinate material (i.e., with a finite diameter and mass) molecules. The number of molecules involved in momentum transfer between the inlet and outlet of the channel increases with the D/λ ratio (D is the channel diameter and λ is the mean free path of a molecule). The integral characteristics of the gas flow are calcu-

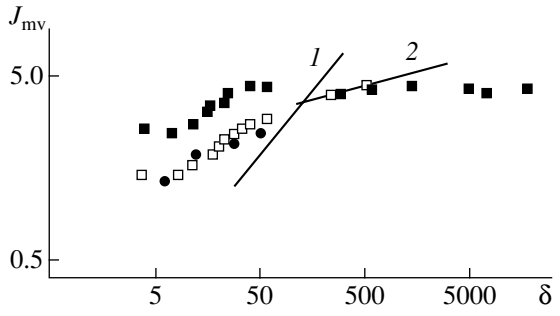


Fig. 2. Dimensional conductivity J_{mv} vs. the degree of rarefaction δ for circular pipes and openings in the case of slip flow. (●) The method of probable directions and (□) full-scale experiment for 55-mm-long pipelines with an inlet diameter of 25 mm. (*1*) Full-scale experiment for a long pipe of diameter 25 mm [2], (*2*) boundary layer model used in continuum mechanics [6], and (■) full-scale experiment for an opening of diameter 25 mm.

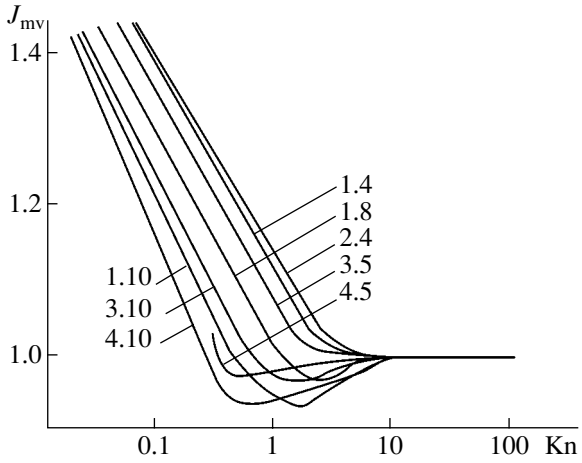


Fig. 3. Dimensionless conductivity J_{mv} vs. Kn for circular pipes in the molecular–viscous regime. The method of probable directions with L/D = (1.4) 4, (1.8) 8, and (1.10) 10; solution of the linearized Boltzmann equation for L/D = (3.5) 5 and (3.10) 10 [15]; calculation for L/D = (4.5) 5 and (4.10) 10 [15]; and experiment for L/D = (2.4) 3.7 [2, 6].

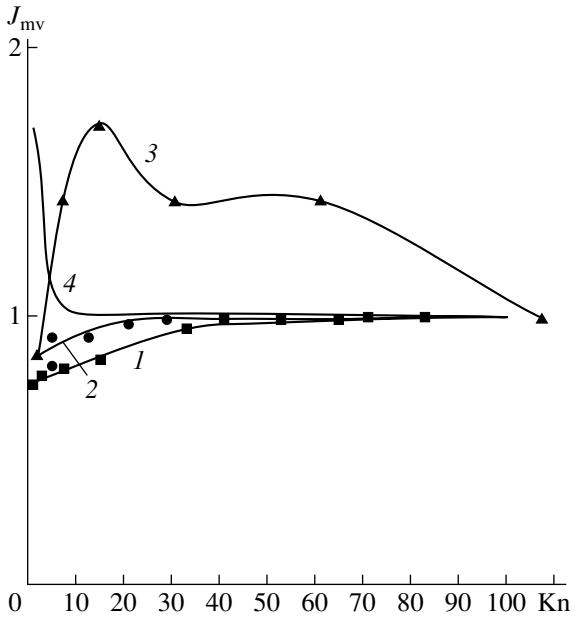


Fig. 4. Dimensionless conductivity J_{mv} vs. Kn for a rectangular slotted channel in the molecular–viscous regime. Full-scale experiments performed in (*1*) [1] and (*4*) [3], (*2*) the method of probable directions, and (*3*) direct simulation [14, 15].

lated by successively tracing the walk of individual probe molecules in the channel. A random sample of gas molecules that provides the desired accuracy of Monte Carlo statistical processing is considered. The trajectory of an individual molecule is described by a piecewise linear function that represents a broken line with segments equal to the free path λ_0 of the molecule:

$$\lambda_0 = -\lambda \ln(R), \quad (1)$$

where λ is the mean free path of the molecule and R is a random number ($R \in [0, 1]$) generated by the Monte Carlo method.

Molecules striking the channel wall first adhere to it and then leave according to the diffuse distribution. If a molecule travels a length λ_0 without collision with the wall, a collision with another molecule is simulated. As a result of this collision, the molecules change direction according to a model of molecular interaction.

A model of interaction (collision) in a statistical molecular ensemble is constructed in terms of the probability theory as applied to a large set of molecules [20]. The model is based on the assumption that the interaction potential is small. Then, (i) in describing the ensemble of molecules, the principle of superposition can be considered as a linear combination of pair interactions and (ii) the correlation length of pair intermolecular forces in the collisional range of the rarefied flow may exceed the mean free path of molecules (in view of the collective motion of a group of molecules).

Considering the probabilistic approach, we assume that the direction of each of the molecules after collision is random. In the range $0.01 < \delta < 0.50$, where pair collisions prevail, the directions of an individual statistical molecule are taken to be equiprobable in a complete solid angle of 4π sr [13]. In the range $0.5 < \delta < 100$, a statistical regularity that defines the postcollision direction of an individual molecule in a complete solid angle was found [10]. This regularity includes collective collisions of molecules [20] and random-to-laminar flow transformation and also defines the molecular motion directivity in the flow after collision as a function of the molecule concentration in a microvolume when Kn decreases [21–23].

Next, having determined the postcollision direction of the molecule, we simulate its free motion in this direction. Walks are simulated until the molecule escapes through the inlet or outlet of a vacuum element.

Analytical results for rarefied gas flow on the macrolevel have been supported by full-scale measurements made on circular and rectangular short and long ducts, rectangular slots, and step adapters (Figs. 1–4) [19–22]. The experiments have shown that probabilistic simulation cuts the machine time 100-fold compared with Bird's direct simulation method and at the same time needs moderate computational resources [24–27].

Results obtained by the method of probable directions on the microlevel level shed light on physical processes observed at the macrolevel [24, 25]. It should be noted that the fraction of collective interactions (including those at the channel walls) increases with δ. It has been shown [10] that, with $\delta \cong 0.5$, the tendency toward directional (drift) motion in the molecular–viscous regime appears.

Consideration of molecular motion on the macrolevel clarifies the reason why the conductivity of openings and short ($L/D < 4$, where L is the pipe length and D is the diameters of the pipe inlet) pipes increases throughout the transition flow range (Fig. 3). On the macrolevel, (i) the number of molecules passing through a vacuum element in the transition regime is larger than in the molecular one, (ii) the number of molecules involved in the transport process also increases, and (iii) the number of molecules passing through a vacuum element ballistically (i.e., without collisions with other molecules or walls) far exceeds the number of colliding particles. In the range $1 > \mathrm{Kn} > 0.01$, the conductivity grows at a higher rate. On the microlevel, this is explained by the fact that the flow becomes more and more laminar.

Consideration of the molecular flow on the microlevel clarifies the nature of the Knudsen paradox: a decrease in the conductivity of long pipes as Kn decreases from 100 to 1 and an increase in the conductivity when Kn diminishes from 1 to 0.01 (Fig. 3). On the microlevel, the fraction of molecular pair collisions grows, which makes the molecular motion along the pipe difficult. In this situation, the number of molecules

passing through a vacuum element decreases despite the fact that the number of molecules involved in the transfer process rises. With Kn between 1 and 0.01, the conductivity grows monotonically for two reasons: the number of molecules taking part in the transfer process increases and the flow becomes more directional [20–22].

Considering the motion of molecules on the microlevel, one sees that, as the length-to-diameter ratio L/D drops from 40 to 5, the Kn dependence of the conductivity smooths out (Fig. 3) because of a decrease in the number of pair collisions compared with that for long pipes. The minimum becomes less pronounced and shifts toward larger Knudsen numbers (Fig. 3). For pipes with $L/D < 4$, the dependence becomes smooth and monotonic and extrema are absent (Fig. 3).

Consideration of the molecular motion on the microlevel elucidates the effect observed at the boundary between the molecular–viscous and pure viscous regimes (Kn $\cong$ 0.01): the conductivities of short and long channels of the same diameter approach each other (Fig. 2). Under these conditions, most molecules travel through a vacuum element, since the flow laminarizes.

The method of probable directions is still being developed and defined [23–25].

Thus, qualitative and quantitative analysis of rarefied gas transition flow is such a complex problem that any step forward in this direction is of great value. It is related to the slip theory, the theory of gas diffusion, and Boltzmann statistics. It is noteworthy that one of the issues on the agenda of the international conference "Vacuum Gas Dynamics" held in Spain, July 2003, was "Can we agree on a way to model transition gas flow?"

The models and methods used in simulating the dynamics of rarefied gas were verified in [17, 18]. It was shown that, throughout the molecular–viscous range, solutions to equations of continuum mechanics with empirical coefficients taken from the slip theory [11] are in most cases matched to results of direct statistical simulation [14]. Such an approach is formal, since it cannot be given any physical interpretation. Physical models underlying both methods are inconsistent with each other.

Figure 3 compares the direct simulation method with the method of probable directions (relevant experimental data are also shown) by plotting the Kn dependence of the dimensionless conductivity I_{mv}. Such a representation is necessary when experimental data from various publications are verified or compared in terms of the theory of similarity [17]. Note that the Knudsen number is the decisive similarity test for the steady isothermal molecular–viscous flow of rarefied gas in geometrically similar structures that have a single characteristic length. This follows both from the kinetic theory [17], where the gas flow is viewed as the flow of determinate molecules, and from the continuous theory [9], where integral characteristics of rarefied gas flow are considered. It should be noted that such a physical meaning of the Knudsen number has been substantiated in many full-scale experiments. Physical quantities that characterize a gas of a given sort are constant quantities. Their effect should be taken into account when the properties of a gas of one sort are converted to those of a gas of another sort. Therefore, for geometrically similar elements, the dependence of J_{mv} on the properties of rarefied gas flow in the molecular–viscous regime may be represented as

$$J_{mv} = J_{mv}(A, \mathrm{Kn}, sort),$$

where J_{mv} is the ratio of the conductivities in the molecular–viscous and molecular regimes, A is the cross-sectional area of the inlet, and *sort* combines constant parameters of a gas of a given sort (the effective molecular diameter, etc.).

For example, for vacuum pipe fittings, we have

$$J_{mv} = J_{mv}(A, L/D, \delta, sort),$$

where L/D is the effective length-to-effective diameter ratio of a vacuum element and $\delta = 1/\mathrm{Kn}$.

The fact that the data (Figs. 2–4) obtained by the method of probable directions, the direct simulation method, and by solving the linearized Boltzmann equation qualitatively coincide indicates that the first two methods provide reliable results. The deviation from the full-scale measurements (curves *4*, *5*, Fig. 3; curve *3*, Fig. 4) may be associated with the extent to which the Boltzmann equation is approximated by these methods of numerical experiment, the accuracy of implementation of these methods, computer performance, and the accuracy of numerical and full-scale experiments.

CONCLUSIONS

The model underlying the method of probable directions seems the most adequate for describing the transition flow of rarefied gas [24, 25]. It is based on fundamental concepts of the kinetic theory of gases and, at the same time, is consistent with the diffusion theory of gases and the models of molecular and viscous flow on the macro- and microlevels. Second, this model takes into account not only pair but also collective molecular collisions, as well as the finite mass and effective diameter of molecules. Finally, it sheds light on the Knudsen paradox and accounts for the close values of hydraulic conductivity that are observed in vacuum channels of equal diameter but different length at the boundary between the molecular–viscous and viscous regimes.

REFERENCES

1. B. T. Porodnov, P. E. Suetin, S. F. Borisov, and N. D. Akinshin, J. Fluid Mech. **64**, 417 (1974).
2. D. J. Santeler, J. Vac. Sci. Technol. A **12**, 1744 (1994).
3. R. G. Livesey, J. Vac. Sci. Technol. A **19**, 1674 (2001).

4. Yu. A. Koshmarov and Yu. A. Ryzhov, *Applied Dynamics of Rarefied Gas* (Mashinostroenie, Moscow, 1977).
5. C. Cercignani, *Theory and Application of the Boltzmann Equation* (American Elsevier, New York, 1975; Mir, Moscow, 1976).
6. Yu. M. Pechatnikov, Vak. Tekh. Tekhnol. **6** (2), 5 (1996).
7. Yu. M. Pechatnikov, Inzh. Fiz., No. 3, 50 (2003).
8. J. K. Haviland, Methods Comput. Phys. **4**, 109 (1965).
9. B. T. Porodnov, L. M. Nusinzon, I. G. Neudachin, and V. V. Kalinin, Vopr. At. Nauki Tekh., Ser.: Fiz. Tekh. Vysokogo Vak. (Kharkov), No. 2, 82 (1977).
10. Yu. M. Pechatnikov, Inzh.-Fiz. Zh., No. 6, 673 (1992).
11. E. H. Kennard, *Kinetic Theory of Gases with an Introduction to Statistical Mechanics* (McGraw-Hill, New York, 1938).
12. L. M. Nusinzon, B. T. Porodnov, and P. E. Suetin, Mekh. Zhidk. Gaza, No. 1, 187 (1977).
13. Yu. M. Pechatnikov, Élektron. Tekh., Ser. 4: Élektrovak. Gazorazryadnye Prib., No. 4, 67 (1991).
14. G. A. Bird, *Molecular Gas Dynamics* (Clarendon Press, Oxford, 1976; Mir, Moscow, 1981).
15. G. Scherer-Abreu and R. A. Abreu, Vacuum, Nos. 8–10, 863 (1995).
16. A. I. Erofeev, M. N. Kogan, and O. G. Frilender, Mekh. Zhidk. Gaza, No. 5, 193 (1999).
17. V. N. Gusev, I. V. Egorov, A. I. Erofeev, and V. P. Provotorov, Mekh. Zhidk. Gaza, No. 2, 128 (1999).
18. A. K. Rebrov, in *Proceedings of the Seminar on the Modern State of Vacuum Technology, Kharkov, 2001* (Kontrast, Kharkov, 2001), pp. 6–15.
19. Yu. M. Pechatnikov, *Analysis of Complex Vacuum Elements in Molecular–Viscous Gas Flow Regime* (Svetlana, St.-Petersburg, 2002).
20. Yu. M. Pechatnikov, in *Proceedings of the Conference on Vacuum Science and Technology, Moscow, 2002* (Mosk. Inst. Élektron. Mashinostr., Moscow, 2002), pp. 51–54.
21. Yu. M. Pechatnikov, Inzh. Fiz., No. 2, 32 (2003).
22. Yu. M. Pechatnikov, Zh. Tekh. Fiz. **73** (8), 40 (2003) [Tech. Phys. **48**, 978 (2003)].
23. Yu. M. Pechatnikov, in *Proceedings of the International Conference on Vacuum Technology and Equipment* (Kontrast, Kharkov, 2003), pp. 64–67.
24. Yu. M. Pechatnikov, in *Proceedings of the Seminar on Vacuum Technology and Equipment* (UNIVAK, St.-Petersburg, 2003), pp. 32–37.
25. Yu. Pechatnikov, in *Proceedings of the 8th European Vacuum Congress on Vacuum Science and Technology, Berlin, 2003*.
26. Yu. M. Pechatnikov, in *Proceedings of the Seminar on Vacuum Technology and Equipment* (UNIVAK, St.-Petersburg, 2003), pp. 38–43.
27. Yu. Pechatnikov, in *Proceedings of the 4th International Symposium on Vacuum Technologies and Equipment, Kharkov, 2001*, pp. 128–129.

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 12, 2003, pp. 1527–1531. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 26–30.

SOLIDS

Electrical Breakdown in Ionic Crystals Exposed to Nanosecond Pulses

V. D. Kulikov

Tomsk Polytechnical University, pr. Lenina 30, Tomsk, 634050 Russia

e-mail: kulikov@list2.epd.tpu.edu.ru

Received January 10, 2003; in final form, May 20, 2003

Abstract—Electrical breakdown in alkali halide crystals subjected to ~10-ns-long electrical pulses is studied. Two, primary and basic, channels of the anodic discharge are noted. In the presence of the primary channel, the other arises at static breakdown voltages. Otherwise, the basic channel forms at voltages exceeding the static breakdown voltage by more than four times. The basic channel is assumed to form via a cascade of Auger transitions. The generation and migration of linear defects seem to play a significant role in the basic channel formation. The enhancement of the pulsed dielectric strength of the crystals is related to conditions of current passage through the metal–insulator interface. *© 2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The prevention of pulsed electrical breakdown in insulators remains a challenge for designers and users of any electrotechnical devices (from high-voltage equipment to microelectronic components). A specific feature of pulsed breakdown in alkali halide crystals [1–3] and complex ionic compounds (such as glass [1] and ammonium perchlorate [4]) is that in the breakdown voltage increases as the pulse width decreases. For pulse widths of ≈30 ns, a through breakdown channel in the crystals forms at voltages ≈2.5 times greater than the quasi-static breakdown voltage (i.e., when the field application duration is 1 μs or more) [2]. The structure of the anodic breakdown channel depends on the field strength. In NaCl crystals, a primary breakdown channel oriented along the "cathodic" ⟨100⟩ direction forms in the anode region at near-breakdown voltages under static conditions. The extension of this region varies from 50 to 500 μm. Then, the discharge channel propagates along the "allowable" ⟨110⟩ direction (basic channel). The primary channel propagated with a subsonic velocity, while the propagation velocity of the basic one in the allowable direction is ~10^7 cm/s. As the voltage grows, the primary channel shrinks and eventually becomes visually indiscernible [2, 3].

Thus, it may be supposed that the primary and allowable (basic) channels form by different breakdown mechanisms.

In this work, we study electrical breakdown in alkali halide crystals subjected to ≈10-ns-long pulses, examine the structure of the discharge channels, estimate breakdown voltages, and consider the breakdown mechanism in pulsed electric fields.

EXPERIMENTAL RESULTS

The objects investigated were 40 × 40-mm alkali halide crystal specimens with a thickness d = 0.3–6.0 mm. An electron beam from a GIN-400 accelerator was used as a pulse source (the maximal energy of electrons is ≈0.3 MeV; the beam current density, ≈300 A/cm^2; the pulse width, ≈18 ns). Aluminum electrodes ≈8 mm in diameter and ≈1 mm in thickness were made on both sides of the specimen (Fig. 1a). The electron beam fell on the lower electrode, thereby producing a capacitor with negative charge in the lower electrode and induced positive charge in the upper electrode. To prevent charge leakage, the lower electrode was mounted on an insulating support. Irradiation was performed at room temperature under a pressure of 0.13 Pa.

The field strength E was estimated by measuring the pulsed current passing in the charging circuit at the time of irradiation [5, 6]. The displacement current density i is related to E as $i = \varepsilon\varepsilon_0\partial E/\partial t$, where $\varepsilon\varepsilon_0$ is the absolute permittivity of the specimen. The current density i at the upper electrode is directly proportional to the permittivity of the material and inversely proportional to the specimen thickness in accordance with the plane capacitor law. The field strength $E(t)$ is found by integrating i over time. The peak (maximal) intensity E_m depends on the specimen thickness as $E_m \sim 1/d$.

Figure 1b shows the time waveforms of the beam current density I, displacement current density i, and field strength $E(t)$ for a 5.1-mm thick NaCl specimen. Here, the FWHM of the pulse is ≈10 ns and $E_m \approx 1.1 \times 10^6$ V/cm.

The formation of the channels was studied under different applied voltages (fields). At the time of field application, the discharge channel appeared as a glow between the electrodes. After the irradiation, it was

1063-7842/03/4812-1527$24.00

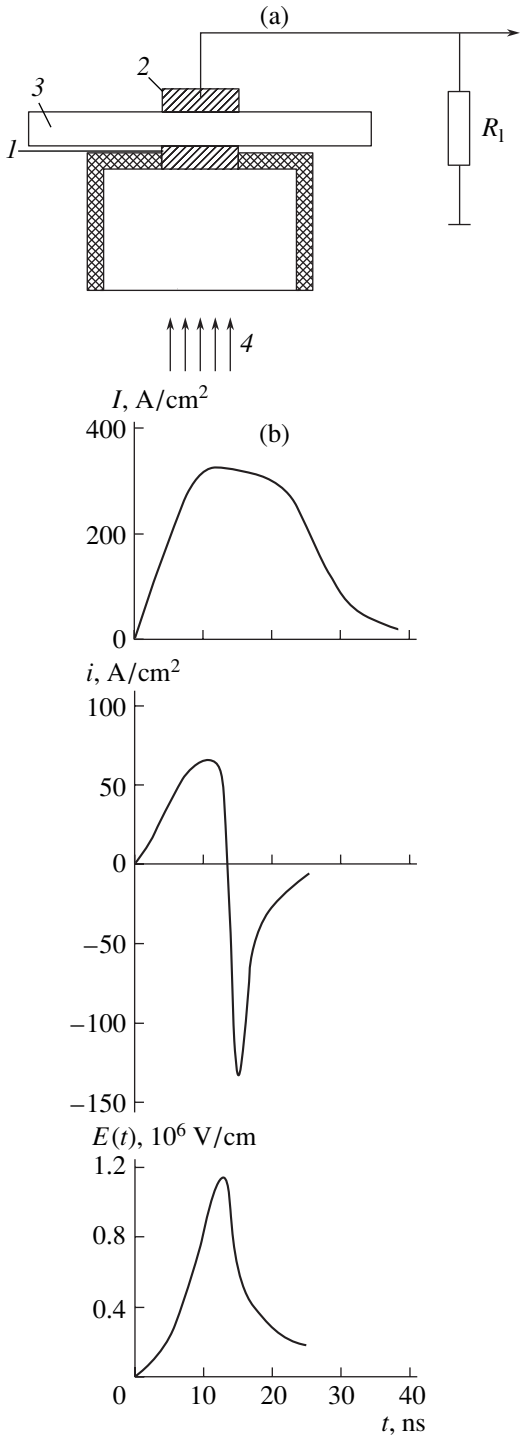


Fig. 1. (a) Experimental scheme: (*1*, *2*) electrodes, (*3*) insulator, (*4*) electron beam, and R_l is the load resistance. (b) I, electron beam current density; i, displacement current density; and $E(t)$, field pulse in the 5.1-mm-thick NaCl sample.

visualized by breakdown traces. The discharge channels propagate from the positive electrode (usually under its edges). NaCl plates 5- to 6-mm-thick (E_m = (0.9–1.12) × 10^6 V/cm) were broken down after ten to twelve pulses. However, after the initial five or six pulses, the glow, as well as material breakdown under the positive electrode, was absent. Then, the channel advanced from pulse to pulse up to the other electrode. Near the anode, the glow is not as bright as in the channel. It is essential that, for specimens 3.7 to 4.0 mm thick ($E_m \approx 1.5 \times 10^6$ V/cm is close to the static dielectric strength [1]), neither the glow nor breakdown under the anode is seen after the application of one or two initial pulses; the next pulse, however, breaks down the specimen throughout its thickness. We failed to observe a clear-cut primary channel in the ⟨100⟩ direction, although signs of it, a faint glow and minor damage to the material near the anode [3], were detected. The image of a typical discharge channel in the 5.1-mm-thick NaCl crystal after the application of ten pulses is shown in Fig. 2. The basic channel propagates from the upper surface (positive electrode) toward the lower one along the allowable ⟨110⟩ direction and noticeably damages the insulator. It seems that the primary channel is coincident with the basic one and therefore is indistinguishable. The structure of the basic channel at pulsed and static anodic discharges is the same. The primary channel is absent if a through channel appears after the first pulse. In NaCl, such a channel is observed at a thickness of ≈1 mm ($E_m \approx 5.6 \times 10^6$ V/cm).

For KI, NaCl, and LiF crystals, the basic discharge channel are aligned with the ⟨110⟩ direction; for KBr, RbCl, and KCl, with the ⟨100⟩ direction. For a pulse width of ≈10 ns, the mean dielectric strength corresponding to the one-pulse formation of a through channel equals ≈2.24 MV/cm for KI, ≈2.8 MV/cm for KBr, ≈3.1 MV/cm for RbCl, ≈3.73 MV/cm for KCl, ≈5.6 MV/cm for NaCl, and ≈18.5 MV/cm for LiF. According to these data, the pulsed dielectric strength exceeds the static value by a factor of 3.7–4.0 (for LiF, by a factor of 6). The field strength is kept at its peak value for ~10 ns, and the basic channel length is ~1 mm; hence, the mean velocity is on the order of ~10^7 cm/s, which is in satisfactory agreement with data obtained in [2, 3].

We studied the effect of X-ray radiation on the dielectric strength of the crystals. X-ray radiation arose when the electrons of the beam were decelerated in the Al target. The maximum of the bremsstrahlung spectrum lies in the low energy range. The absorption of radiation-induced F centers in the 3-mm-thick KBr sample was estimated with a He–Ne laser [7]. Without the field, absorption at the positive electrode was virtually absent and the absorption peak was observed at the negative Al electrode. With the lower Al electrode replaced by a 1-mm-thick lead electrode, the absorption of F centers diminishes down to zero. No variation in the dielectric strength of the samples was detected.

DISCUSSION

It seems reasonable to analyze the results taking into account the passage of through the metal–insulator interface.

In a gap of width s between a metallic electrode and insulator, the field strength is $E_1 = \varepsilon\varepsilon_0 E_m$. One should take into account that the electrode surface always has microtips of height $h < s$. At the top of the tips, the field strength is $E_2 = E_1 h/r = E_1\mu$, where r is the microtip radius and μ is the field enhancement factor. For $s = 10^{-3}$–10^{-2} cm, $\mu \approx 100$ [8, 9]. Hence, the actual field strength at the top of the microtips may be $\geq 10^8$ V/cm.

In an insulator, the Fermi level lies near the midgap [1]. When an insulator comes into contact with a metal, the Fermi levels are aligned (Fig. 3). The surface layer of the metal is depleted by electrons. An electric field on the order of 10^8 V/cm causes the electrons to tunnel from the valence band of the insulator to the metal.

The formation of a streamer in the bulk of an insulator may be explained in terms of a cascade of Auger transitions [10]. The basic ideas of this model are as follows.

(1) In the crystal lattice, electronic excitation is transferred from atom to atom according to the electronic configuration of a crystal. The table lists the binding energy for the highest energy levels of the cations in the valence band (measured from the conduction band bottom), the width W_{vs} of the upper valence subband, and the energy gap W_g for alkali halide crystals [11, 12]. Two ways of streamer formation may be singled out. One refers to crystals where the gap W_1 between the center of the upper valence subband and the highest cation energy level exceeds W_g (group I). Such are Li compounds (LiF, LiCl, LiBr, and LiI), Na compounds (NaF, NaCl, NaBr, and NaI), and KI. The other way applies to crystals for which $W_1 < W_g$ (group II). This condition is satisfied for K compounds (KF, KCl, and KBr) and Rb compounds (RbF, RbCl, RbBr, and RbI). An external field results in a significant bend of the energy bands. As a result of tunnel transition, holes accumulate at the insulator surface (for example, one or two holes on the 3p level of Cl^+ ions in NaCl; Figs. 3a, 3b). The travel of the streamer is associated with the relaxation of these holes and injection of electrons in the conduction band. In the crystals of group I, the holes on the low-lying 1s level of Li and 2p level of Na cannot relax. It is most likely that they relax on a nearby anion in the ⟨110⟩ direction, for example, through an interatomic Auger transition that transfers the charge from the 3p level of Cl^- (transition *1*) with the subsequent injection of the Auger electron into the conduction band (Fig. 3a).

In the crystals of group II, the hole relaxation in halogen ions takes place on cation high-lying levels. For example, in KCl, the hole relaxation in Cl^+ ions takes place on the 3p level of K^+. For resonant electron transfer to occur, it is necessary that the 3p level of K^+ in KCl be raised by ≈6.1 eV (Fig. 3b). Subsequent hole relaxation in K^{++} takes place as a result of the interatomic Auger decay on the 3p level of Cl^-. In these crystals, the streamer propagates in the ⟨100⟩ crystal direction. The fact that the breakdown is initiated at the anode and propagates in certain crystal directions is confined experimentally. For KBr, the values of W_g and W_1 are close to each other, which is corroborated by the change of breakdown direction from ⟨100⟩ to ⟨110⟩ at temperatures higher than 50°C [13].

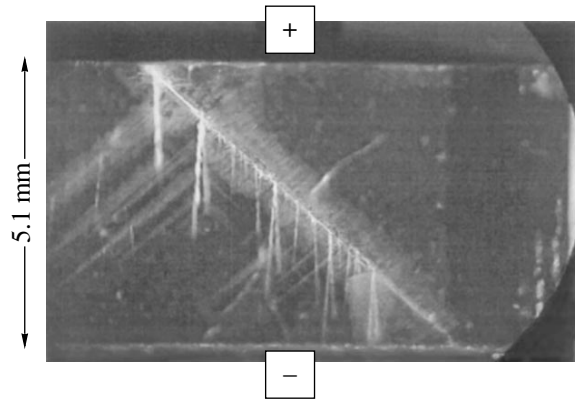


Fig. 2. Discharge channel in the NaCl sample.

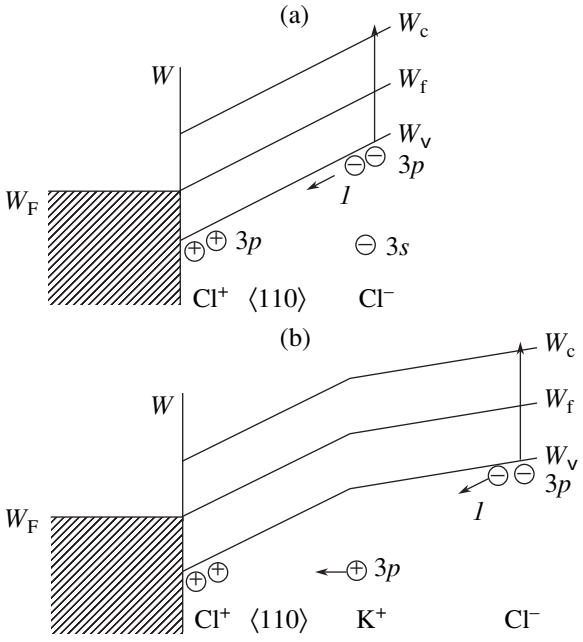


Fig. 3. Cascade Auger transitions in (a) NaCl and (b) KCl under a strong electric field. W is energy; W_c and W_v are, respectively, the conduction band bottom and the valence band top; and W_f and W_F are the Fermi level positions in the insulator and metal, respectively.

(2) Over a distance on the order of the interatomic spacing, the bend of the energy bands is comparable to the energy gap of the insulator. The probability that an Auger electron will fall into the conduction band is other than zero if the minimal energy gap between the 3p levels of neighboring chlorine ions in NaCl (Fig. 3a) (or between the 3p level of K^{++} and 3p level of Cl^- in KCl, Fig. 3b) is no less than the energy gap of the crys-

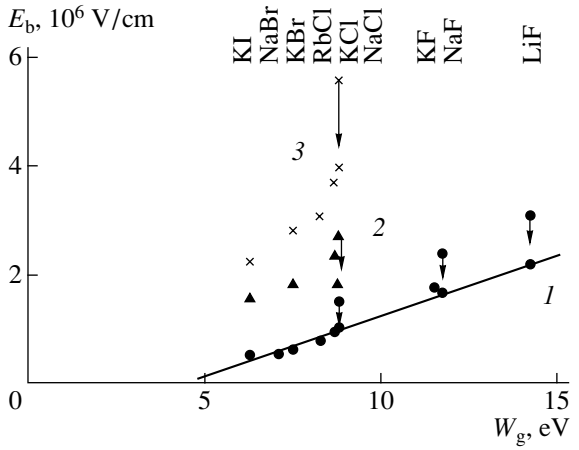


Fig. 4. Breakdown field E_b vs. energy gap W_g for various insulators. (*1*) Quasi-static conditions [1], (*2*) pulse width 30 ns (sample thickness ≈0.15 mm) [2], and (*3*) experimental data.

tal. For a mean interatomic distance in the lattice of ≈3 Å (the spacing between Cl^- ions is ≈4 Å), the external field strength must be $(2.5–3.0) \times 10^8$ V/cm. Such values can be reached only near irregularities on the electrode surface or at the end of the conducting channel. In high fields, electrons can be injected from the valence band into the conduction bands in two ways: by the tunnel or the Auger mechanism. In alkali halide crystals, the dependence of the breakdown field E_b on the energy gap is close to linear, which favors the Auger mechanism. In the case of tunnel transition, the E_b vs. W_g dependence must be near-exponential.

(3) Multiply charged ions generate high electric fields. Two holes in a chlorine ion produce the effective charge of Cl^+; one hole in a K^+ ion, the effective charge of K^{++}. The charge e^+ generates a field of $\approx 3 \times 10^7$ V/cm over a distance of 3–4 Å and $\approx 10^8$ V/cm over a distance of 1 Å. Such high fields are comparable to external ones.

(4) The rate of breakdown depends on the time of Auger transition. Bearing in mind that electrons in a cascade of Auger transitions are transferred from atom to atom, one can estimate the rate of breakdown as $\nu = 1\ \text{cm}/N\tau$, where N is the number of ions over a length of 1 cm and $\tau \approx 10^{-14}$ s is the Auger transition time [14]. For NaCl, $N \approx 3 \times 10^7$, yielding $\nu \approx 10^7$ cm/s, which is in satisfactory agreement with experimental data.

It is of interest to compare the static and pulsed dielectric strengths as a function of the energy gap for various insulators (Fig. 4). When analyzing these data, we took into consideration the mutual arrangement of the electric field and breakdown channel. In the crystals of group II, the field E and the breakdown channel are aligned with the ⟨100⟩ direction. In the crystals of group I, the breakdown propagates along the ⟨110⟩ direction. In this case, the field component in the ⟨110⟩ direction is less than E_b by a factor of 1.41 (in Fig. 4, the dielectric strengths of NaCl, NaF, and LiF that are decreased by a factor of 1.41 are indicated by arrows). Work [2], where the dielectric strength ratio for NaCl plates with the ⟨110⟩ and ⟨100⟩ directions running across was found to be ≈1/1.41, confirms our estimates. From Fig. 4, it follows that the E_b vs. W_g dependence under static conditions is near-linear but does not pass through the origin:

Electron binding energies and the parameters W_{vs}, W_g, and W_1 of alkali halide crystals (eV) [11, 12]

Crystal	Ion state	Binding energy	W_{vs}	W_g	W_1
LiF	$1sLi^+$	56.5	9.4	14.2	37.6
LiCl	$1sLi^+$	56.5	5.2	9.4	44.5
LiBr	$1sLi^+$	56.6	4.6	7.5	46.8
LiI	$1sLi^+$	55.8	3.1	6.2	48.05
NaF	$2pNa^+$	33.0	10.2	11.7	16.2
NaCl	$2pNa^+$	31.2	5.6	8.8	20.2
NaBr	$2pNa^+$	30.8	4.5	7.1	20.95
NaI	$2pNa^+$	30.9	4.5	5.9	23.75
KF	$3pK^+$	17.4	7.2	11.4	2.4
KCl	$3pK^+$	17.5	5.4	8.7	6.1
KBr	$3pK^+$	17.0	4.2	7.5	7.4
KI	$3pK^+$	18.0	4.2	6.3	9.6
RbF	$4pRb^+$	14.7	9.7	10.3	0.45
RbCl	$4pRb^+$	14.6	5.7	8.3	3.4
RbBr	$4pRb^+$	14.1	4.6	7.3	4.5
RbI	$4pRb^+$	13.9	3.4	6.3	5.9

$E_b = E_0 + kW_g$ (k is the proportionality coefficient and E_0 is the field strength at $W_g = 0$). It is likely that the bands are bent under the combined action of the external field and the field of a hole in the streamer channel. As a result, the breakdown voltage decreases by E_0. For a pulse width of 30 ns, E_b varies with W_g almost in the same way as under static conditions. For a 10-ns pulse, the E_b vs. W_g curve is steeper.

In KBr crystals subjected to prebreakdown fields, the generation of *F* centers near the anode was observed [7]. In alkali halide crystals, color centers are produced by the nonradiative decay of autolocalized excitons. The excitons, in turn, arise when conduction electrons are captured by the vacant state of autolocalized holes [12]. It seems most probable that the generation of *F* centers is associated with the generation of electron–hole pairs in a strong electric field. If so, the generation of free carriers counts in favor of the model of cascade Auger transitions. Unfortunately, the electrooptic method of estimating the field strength gave underestimated values of E_m [7]. The pulsed current method gives a field strength of $\approx 0.6 \times 10^6$ V/cm for a 3-mm-thick sample.

Thus, the properties of the basic discharge channel (the direction and rate of propagation and the growth of the dielectric strength with the energy gap of the insulator) are adequately described by the model of cascade Auger transitions under both static and pulsed conditions.

As was noted above, the primary breakdown channel is aligned with the field in the ⟨100⟩ direction and propagates with a subsonic velocity [2, 3]. It has been found that a prebreakdown field applied to thin layers of alkali halide crystals generates point and linear defects. Their concentration increases with temperature and duration of the pulse [15]. The formation of the primary channel is likely to be associated with the generation and migration of linear defects, which create channels and regions of easy charge transfer.

In our opinion, the increase in the breakdown voltage with decreasing pulse duration is related to the passage of current through the metal–insulator interface. It is conceivable that there exists a blocking bend of the bands on the insulator surface, which simultaneously prevents electron tunneling from the valence band and escape of the free electrons into the metal. In this case, a positive charge does not accumulate in the insulator. The effect of the bend can be reduced by either generating linear defects on the surface or rectifying the energy bands. The former case requires long pulses (quasi-static breakdown); the latter, nanosecond pulses and high fields.

CONCLUSIONS

Thus, pulsed electric fields applied to alkali halide crystals produce two discharge channels (primary and basic) propagating from the anode. In the basic channel, the carriers responsible for an impact avalanche are generated by the mechanism of cascade Auger transitions in the valence band of the insulator. The primary breakdown channel is most likely to be associated with the generation and migration of linear defects, which facilitate charge transfer. It appears that an increase in the dielectric strength under pulsed conditions is related to the fact that the current passage through the metal–insulator interface is difficult in this case presumably because of a blocking bend of the bands on the insulator surface. A decrease in the pulsed breakdown voltage to static values in the presence of the primary channel confirms this statement.

REFERENCES

1. G. I. Skanavi, *Physics of Insulators: High Field Range* (Fizmatgiz, Moscow, 1958).
2. A. A. Vorob'ev and G. A. Vorob'ev, *Electrical Breakdown and Damage of Solid Insulators* (Vysshaya Shkola, Moscow, 1966).
3. Yu. N. Vershinin, *Electrical Breakdown of Solid Insulators: Thermionic and Detonation Processes* (Ural. Otd. Ross. Akad. Nauk, Yekaterinburg, 2000).
4. I. G. Khaneft and A. V. Khaneft, Zh. Tekh. Fiz. **70** (4), 42 (2000) [Tech. Phys. **45**, 423 (2000)].
5. V. D. Kulikov, Izv. Akad. Nauk Lat. SSR, Ser. Fiz. Tekh. Nauk, No. 5, 97 (1990).
6. V. D. Kulikov and Yu. V. Lisyuk, Zh. Tekh. Fiz. **63** (7), 74 (1993) [Tech. Phys. **38**, 568 (1993)].
7. V. D. Kulikov, Pis'ma Zh. Tekh. Fiz. **28** (3), 36 (2002) [Tech. Phys. Lett. **28**, 99 (2002)].
8. R. P. Little and W. F. Whitney, J. Appl. Phys. **34**, 2430 (1963).
9. S. P. Bugaev and G. A. Mesyats, *Pulsed Electrical Discharge in Insulators*, Ed. by G. A. Mesyats (Nauka, Novosibirsk, 1985).
10. V. D. Kulikov, Pis'ma Zh. Tekh. Fiz. **26** (4), 77 (2000) [Tech. Phys. Lett. **26**, 170 (2000)].
11. V. V. Nemoshkalenko and V. G. Aleshin, *Electron Spectroscopy of Crystals* (Naukova Dumka, Kiev, 1976).
12. E. D. Aluker, L. Yu. Lusis, and S. A. Chernov, *Electron Excitations and Radioluminescence in Alkali Halide Crystals* (Zinatne, Riga, 1979).
13. M. E. Gaspari, Phys. Rev. **98**, 1679 (1955).
14. M. A. Elango, *Elementary Inelastic Radiation Processes* (Nauka, Moscow, 1988).
15. G. A. Vorob'ev, S. G. Ekhanin, and N. S. Nesmelov, Izv. Vyssh. Uchebn. Zaved. Fiz., No. 8, 26 (2000).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 12, 2003, pp. 1532–1539. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 31–38.

SOLIDS

Magnetic Field Penetration into Half-Space for Type-II Superconductors of the Second Kind

Yu. V. Medvedev and I. B. Krasnyuk

Galkin Physicotechnical Institute, National Academy of Sciences of Ukraine, Donetsk, 83114 Ukraine

Received October 1, 2002; in final form, April 11, 2003

Abstract—Equations that simulate the magnetic induction and current density distributions in half-space in view of the power I–V characteristic are derived. The magnetization front velocity is determined for a given mean rate of external magnetic field variation at the boundary of the sample. An integral condition for the electrical resistance (nonlinearly depending on the magnetic field) under which the magnetic flux penetrates into the sample with a finite rate is found. An analytical solution that simulates the power variation of the magnetic field at the boundary is given. The Bean generalized model describing the current density distribution near the critical current is considered. It is shown that solutions like shock waves may arise beyond the applicability domain of the Bean model. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

It is known that the magnetic field is fully expulsed from type-II superconductors in the Meissner phase at $H < H_{c1}$, where H_{c1} is the lower critical field. In the mixed state (Shubnikov phase) at $H_{c1} < H < H_{c2}$, where H_{c2} is the upper critical field, the magnetic field penetrates into a superconductor in the form of whirl lines (or vortices). The critical field H_{c1} depends on the London penetration depth λ, which defines the typical scale of electromagnetic response from a superconductor to any external disturbance. As the magnetic field increases, the density of whirl lines increases until the vortex cores start overlapping at $H = H_{c2}$.

Below, we consider the Maxwell equations, which simulate the dynamics of the vortex system at the macroscopic level, i.e., on space and time scales that far exceed both the London penetration depth and characteristic pinning-related scales. These equations should be complemented by the *I–V* characteristic $j(E, B)$, which simulates the superconductor's electromagnetic response, which, in turn, depends on the dynamic behavior of the vortex system.

Nonlinearity in the *I–V* characteristic may arise for various reasons: thermal creep, melting of the vortex lattice, pinning, etc. For example, in the case of magnetic flux creep, the vortex lattice (or at least its part) starts moving [1, 2]. As the current increases to the point where the Lorentz force is higher than the pinning force, the entire vortex lattice may be involved in motion and the *I–V* curve becomes linear. In this case, we are dealing with the viscous flow of the magnetic flux [2].

As has been mentioned in [3], equations for the *I–V* characteristic make it possible to analyze experimental data for thermal stability of the superconducting state of a composite superconductor. Stability analysis of the superconducting state is a basic challenge for engineering superconductivity [3]. In the theory of thermomagnetic instability, an external magnetic field or a current introduced into the specimen are considered as disturbing factors; this is precisely the situation that we consider below. We will restrict analysis to the case when the superconductor temperature is kept almost constant and equal to the coolant temperature.

We will show in the Introduction that the Maxwell equations in linear statement may be reduced to the Burgers equation with viscosity, where resistance $\rho_{flux}(B)$ plays the role of "viscosity." The laws of magnetic flux penetration into half-space are considered, and integral conditions are imposed on the function $\rho_{flux}(B)$. According to these conditions, (i) the undisturbed flux penetrates the half-space $x > 0$ an infinite distance with an infinite velocity and (ii) a flux disturbance penetrates a finite distance with a finite velocity. Obviously, property (i) is typical of solutions to linear models, while property (ii) is inherent in solutions to nonlinear problems. The associated model boundary-value problem can be considered as the generalization of the Bean model for the linear *I–V* characteristic. The generalization of this model for magnetic field induction to the power characteristic is impossible in the general case, since the equation includes the term on the order of $(\partial B/\partial x)^n$, where n is the exponent of the *I–V* characteristic. However, this can be done if the electrical resistance does not depend on the magnetic field induction.

Therefore, in Sect.1, we consider a model equation for transport current density, show the possibility for the occurrence of self-similar waves, and demonstrate that the velocity with which self-similar waves penetrate into the sample depends on the dimensionless parameter $\varepsilon = \langle b' \rangle_t^{-2}$. For example, at $n = 1$, the current

penetrates the sample with an infinite velocity and the Bean model can be applied. For sufficiently small $\varepsilon > 0$ and $n > 1$, the current penetrates into the sample in the form of a decaying density wave and, if $n \gg 1$, the Bean model can be used again. In Sects. 2 and 3, we determine the magnetization front velocity v and show that $v \propto n$ and $v \propto \langle b' \rangle_t^{-1}$, i.e., is proportional to the average rate of magnetic field increase on the sample surface.

This work was inspired by the results of [2], where a similar problem was considered but the dependence $\rho_{flux}(B)$ was not taken into account and the electromagnetic field E decreased linearly within a finite depth. In our paper, the magnetic induction and current density also propagate to a finite depth but decrease by a power law. As to the magnetic flux velocity, the results are basically identical: at $n \longrightarrow \infty$, the critical state model becomes valid. Exponential I–V curves were not considered in this paper.

The data above can be viewed as a kind of generalization of the Bean critical state model where the I–V slope is neglected. The results obtained in [2] can be considered in a similar way. Unlike [2], this work considers the explicit dependence $\rho_{flux}(B)$ for the linear I–V characteristic and generalizes the concept according to which the current with a density j_c is the response to any disturbance. It turns out that taking into account weak disturbances in the vicinity of the point $j = j_c$ is of minor importance if the field E penetrates with a high velocity (i.e., the Bean model is valid) and leads to current density damping at low velocities. The inclusion of strong disturbances in the vicinity of the critical value causes a many-valued density shock wave to arise; however, this issue calls for further consideration. As a result, taking into consideration even weak inhomogeneous disturbances of the current density in the vicinity of the critical current necessitates the use of a hydrodynamic model of current distribution, while taking account of the field dependence of the resistance makes us consider the variation of the magnetic field gradient, i.e., in essence, nonuniform current density ($j \neq j_c$) distributions.

1. PROBLEM DEFINITION

Let us consider the system of equations

$$\mathrm{div}B = 0, \quad \mathrm{curl}E = -\frac{1}{c}\frac{\partial B}{\partial t}, \quad \mathrm{curl}B = \mu_0 j, \quad (1)$$

where B is the magnetic field induction, E is the electromagnetic field, j is the transport current density, and μ_0 is the permeability of the medium.

The current dependence can be defined through the shape of the I–V characteristic $j(E, B)$. The I–V characteristic simulates the superconductor's electromagnetic response, which, in turn, depends on the dynamic behavior of the vortical system.

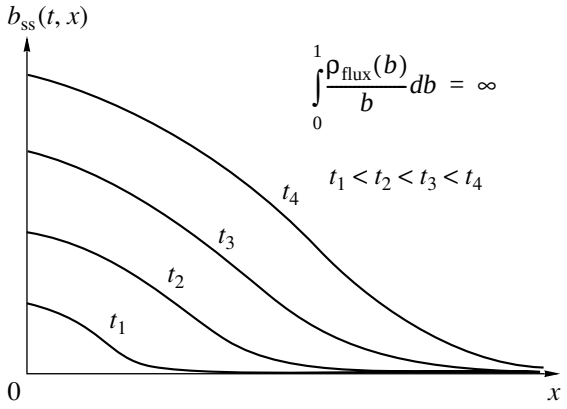


Fig. 1. Magnetic induction distribution for a finite propagation velocity of disturbances (for nonlinear I–V characteristics).

Let us consider a hard superconductor placed in an external magnetic field H that is directed along the Z axis parallel to the surface $x = 0$ of the sample, which occupies the half-space $x > 0$ [2, Fig. 1]. The magnetic field induction can then be expressed as [1]

$$B(x) = H\exp(-x/\lambda) + B_\gamma(x),$$

where the first term describes the Meissner state and the second term is determined by the vortex distribution.

For $B_\gamma(x) \equiv 0$, local coupling $B = N\Phi_0$ takes place, where N is the vortex density and Φ_0 is a fluxon under the condition that the London penetration depth is much less than the typical scale $\langle B' \rangle_t$ of magnetic induction variation [2]. This means that Eqs. (1) describe the behavior of the vortex system on the macrolevel, i.e., on space and time scales larger than those typical of pinning [4]. In other words, this means that inequality $j > j_c$ is satisfied. The reverse case $j < j_c$, where magnetic induction B at every point is determined by vortices localized in a domain on the order of λ, is not considered here.

In the simplest case $U < T$, where T is the temperature, the value of thermal activation barrier U that makes vortex motion difficult can be disregarded. Then, the properties of the sample can be simulated by the expression [4]

$$E = \rho_{flux}(B)j. \quad (2)$$

Here, $\rho_{flux} = \rho_n B/H_{c2}$ and ρ_n is the resistance in the normal state.

From Eqs. (1) and (2), it is easy to derive the Burgers equation

$$\mu_0 B_t' - \frac{\partial \rho_f(B)}{\partial B}(\nabla B)^2 = \rho_f \mathrm{low}(B)\Delta B,$$

where Δ is the Laplacian. Differentiating this equation with respect to x and making substitution $U = \partial B/\partial x$ brings this equation into the conventional Burgers equation with "viscosity."

In terms of the critical state model, this equation can be reduced to

$$\mu_0 B_t' = \rho_{\text{flux}}(B)\Delta B.$$

From the expressions

$$\partial_x E = -\partial_t B \quad \text{and} \quad -\partial_x B = \mu_0 j$$

one easily obtains the equation

$$\mu_0 \partial_t j = \Delta E,$$

which was studied in [2].

There are the following models of the I–V characteristic: the linear model

$$E = \begin{cases} \rho_f(B)(j - j_c) & \text{for} \quad j > j_c \\ 0 \quad \text{for} \quad j < j_c \end{cases} \tag{3}$$

and the nonlinear models [2]

$$E = E_c\left(\frac{j}{j_c}\right)^n;$$
$$E = E_c \exp\left(\frac{j - j_c}{j_\delta}\right) - E_0, \tag{4}$$

where the current density j_δ and phenomenological exponent n define the steepness of the curve.

The current density j_c is found at the electric field intensity E_c. In the Bean model of critical state, the induction dependences of the current density, $j_c(B)$, and resistance, $\rho_f(B)$, can be neglected. It was shown [2] that, for $n \longrightarrow \infty$ and $J_\delta \longrightarrow 0$, models (4) turn into the critical state model.

For linear model (3), one can write

$$E_t' = \frac{\rho_f}{\mu_0}\partial_{xx} E. \tag{5}$$

In the dimensionless variables $e = E/E_X$, $X = x/L_X$, and $\tau = t/t_X$, where

$$E_X = j_c \rho_f, \quad L_X = j_c \rho_f / \langle B_t' \rangle,$$
$$t_X = \mu_0 j_c^2 \rho_f / \langle B \rangle_t^2,$$

Eq. (5) takes the form

$$\partial_\tau e = \Delta_e. \tag{6}$$

Below, we will use the previous designation $X \longrightarrow x$.

Changing variables by averaging $\langle B_t' \rangle$ (see Appendix) allows one to separate typical scales of rate of change of the magnetic induction. In fact, in the dynamic regime, $\langle B_t' \rangle \propto \langle N \rangle$, where $\langle N \rangle$ is the mean number of vortices (on the scales L_X and t_X). Here, $B \propto B_\gamma$ according to the Bean generalized model [4]; i.e., the magnetic induction is the induction of a field averaged over a special scale $a \ll \langle B_t' \rangle < \lambda$, where a is the spacing between vortices. Since the inequality $a \ll \lambda$ is satisfied at $H \gg H_{c1}$, the quantity $\langle B_t' \rangle \propto N$ can be viewed as the rate of change of the density of vortices that leave or enter the sample through its surface [4, p. 1350] and the quantity $\omega \sim 1/t_x$ can be considered as a frequency.

Equation (5) is supplemented by the following boundary and initial conditions [2]:

$$\partial_x E(0, t) = -\partial_t B(0, t), \quad E(\infty, t) = 0, \quad E(x, 0) = 0,$$

where $t > 0$ and $x > 0$. In the dimensionless variables, these conditions are written as

$$\partial e_x(0, \tau) = -1, \quad e(\infty, \tau) = 0, \quad e(x, 0) = 0. \tag{6'}$$

An electric field that is a solution to the mixed initial- and boundary-value problem given by (6) and (6') is known to penetrate the half-space as a self-similar wave with a linear law $X_0(t, \langle B_t' \rangle)$ of magnetization front propagation [2, Fig. 3].

Consider the same boundary-value problem for the magnetic induction equation

$$\partial_t B = \frac{1}{\mu_0}\rho_f(B)\partial_{xx} B.$$

If $B \sim H_{c2}$, the dependence $\rho_f(B)$ can be neglected. Then, in the dimensionless variables introduced in [2], this equation can be written in the form

$$\frac{\partial b}{\partial t} = \frac{\partial^2 b}{\partial x^2} \quad \left(b = \frac{B}{B_{c_2}}\right) \tag{7}$$

with the boundary and initial conditions

$$\partial_t b(0, t) = 1 \quad \text{and} \quad b(x, 0) = 0. \tag{7'}$$

Solutions in the class of self-similar functions to the problem given by (6) and (6') were constructed in [2]. Let us show that the boundary-value problem given by (7) and (7') also have self-similar solutions. First, we note that this problem can be recast as

$$\partial_t b = \Delta b \tag{8}$$

with the boundary conditions written in the more general form

$$b(0, t) = (1 + t)^m \quad \text{and} \quad u(x, 0) = u_0(x), \tag{8'}$$

where $m = 1, 2, \ldots$.

In fact, for $m = 1$, differentiation of edge condition (8') with respect to time yields boundary condition (7'); hence, every solution to problem (8)–(8') is a solution to problem (7)–(7'). Generally speaking, a functional boundary condition is found from an associated differential condition up to a constant, which may be set equal to zero (since boundary and initial functions must continuously match each other).

For $m > 1$, the boundary-value problem considered (subject to $b(\infty, t) = 0$) also has a self-similar solution [5]:

$$b_{ss}(x, t) = (1 + t)^m \theta_{ss}(\zeta),$$
$$\zeta = x/(1 + t)^{1/2},$$

with

$$\theta_{ss}(\zeta) = 2^{2m+1}\frac{\Gamma(1+m)}{\pi^{1/2}}\exp\left(-\frac{\zeta^2}{4}\right)H_{-(2m+1)}\left(\frac{\zeta}{2}\right),$$

where $H_\nu(z)$ is a Hermitean function and $\Gamma(m)$ is the gamma function.

The form of the solution specifies the effective penetration depth of the wave:

$$x_{ef}^{ss} = \zeta_{ef}(1+t)^{1/2},$$

where $\zeta_{ef}(m)$ is determined from the condition $\theta_{ss}(\zeta_{ef}) = \theta_{ss}(0)/2 = 1/2$.

This analytical solution is of interest because it clearly demonstrates the dependence on the rate of electromagnetic field penetration into a superconductor (this electromagnetic field is induced by an increasing external magnetic field).

The analysis of self-similar solutions with a power condition at the boundary shows a natural regularity: the more stringent the boundary condition (the rate of increase of the external magnetic field), the higher the velocity of the arising magnetic wave. In the case of a power condition, the penetration depth of the magnetic field is also a power function. Schematically, the evolution of the self-similar penetration of the magnetic flux generated by the increasing external field is similar to that demonstrated in Fig. 1. It was shown [5, p. 52] that a self-similar solution is asymptotically stable against disturbances of edge data. Accordingly, the self-similar function $\theta_{ss}(\zeta)$ correctly describes the magnetic wave profile at $t \longrightarrow \infty$.

Similarly, for nonlinear *I–V* characteristics, Eq. (4) reduces to the equation

$$b_t' = ((\rho_{flux}(b), j)b_x)_x \quad (b = B/B_{c2}) \qquad (9)$$

with the boundary and initial conditions

$$b(0,t) = (1+t)^m \quad \text{and} \quad b(0,x) = u_0(x). \qquad (9')$$

Here, ρ_f means some dimensionless value of the resistance normalized to ρ_f, where $\rho_f = \langle\rho_{flux}\rangle$ is such that the variables L_x and t_x mentioned above are meaningful. In particular, one may assume, without loss in generality, that $\rho_f = \rho_n$ for $b = 1$ or $\rho_f = \rho_{flux}(B_0)$, where B_0 is a problem-dependent equilibrium value of the magnetic field. Based on Ohm's law $e = \rho_{flux} i$ and the equality $e = i^n$, we may assume that $\rho_{flux}(b, i) \propto bi^{n-1}$, where $i = j/j_c$.

Let $\rho_{flux}(b) \propto b^\alpha$, where $\alpha > 0$. (Below, we will show that this condition has meaning for high-temperature superconductors.) Then, it follows from [5, p. 67] that boundary-value problem (12)–(12') has self-similar solutions $b_{ss} = (1+t)^{1/\alpha}\theta_{ss}(\zeta)$, where θ_{ss} will be defined later. In this case, the magnetic wave penetration depth vs. time is given by

$$x_{ef}^{ss} = \zeta_{ef}(1 + i^{n-1}t)^{(1+m\alpha)/2},$$

$$\theta_{ss}(\zeta_{ef}) = 1/2.$$

In particular, for $m = 1$, boundary-value problem (9)–(9') admits a self-similar solution,

$$u_{ss}(x,t) = g_{ss}(\zeta), \quad \zeta = x/(1+t)^{1/2},$$

which meets boundary condition (9') for $m = 1$. Here, $\theta_{ss}(\zeta)$ is a solution to the boundary-value problem

$$(\rho_{flux}(\theta_{ss})\theta_{ss}')' + \frac{1}{2}\theta_{ss}' = 0,$$

$$\theta_{ss}(0) = 1, \quad \theta_{ss}(\infty) = 0, \quad \zeta > 0.$$

This solution (i.e., the penetration depth) is finite or infinite depending on whether or not Eq. (9) allows for a finite propagation velocity of disturbances.

Let us restrict our analysis to the case where the coefficient $\rho_{flux}(b)$ is such that a disturbance has a finite velocity [5]:

$$\int_0^1 \frac{\rho_{flux}(b)}{b}db < \infty.$$

For Eq. (9), this integral is equal to unity when $\rho_{flux}(b)$ is a linear function or tends to infinity if ρ_{flux} is b independent (solutions to the corresponding problems are shown in Figs. 1, 2). If the resistance does not depend on the magnetic induction, the propagation velocity of magnetic flux disturbances is infinite and the solution is likewise not finite (Fig. 1).

With flux creep taken into account, dissipation is described by the expression

$$\rho(T,H) = \rho_0(H)\exp[-(U_0(T,H)/T)].$$

The energy of activation was shown [6, 7] to have the following scaling form:

$$U_0 = A(H)(1 - T/T_c)^m,$$

where $A \propto 1/H^\alpha$.

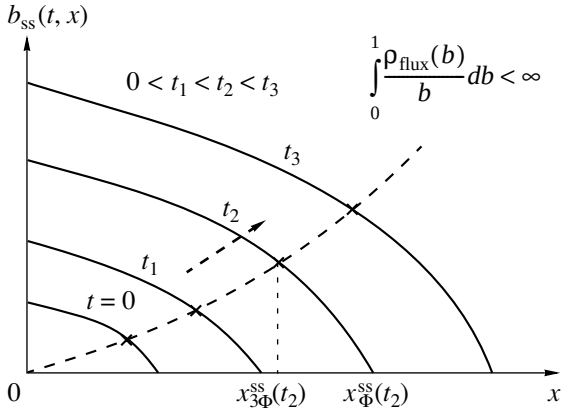


Fig. 2. Magnetic flux penetration for an infinite propagation velocity of disturbances (for linear I–V characteristics).

If the barrier U_0 is small enough, expanding the exponential into the Taylor series and taking into account that $H^\alpha/(H^\alpha + 1) \approx H^\alpha$ in a zero approximation leads us to the relationship

$$\rho(H) \propto \frac{B}{H_{c2}} H^\alpha, \quad \alpha > 0.$$

A similar expression was obtained in [8]:

$$U_{ef}(j, H) \propto F(j/j_i) H^{0.55} [1 - (T/T_x)]^2,$$

where T_x is close to the critical temperature T_c and $F(j/j_i) \sim (j/j_i)^{-n}$ in the case of collective flux creep [8]. Here, j_i is some value of the current that depends on j/j_c [9].

The model worked out in [8] predicts the following regimes: $j \ll j_c$, $j < j_c$, and $j \sim j_c$. We will consider only the last one. Since the authors of [6] and [8] considered the range $0.5 < \alpha < 0.6$ and the value of $\alpha = 0.55$, respectively, it will suffice to put $\alpha = 3/2$ in the above formulas in order to determine the penetration depth of the traveling magnetic wave. This abstract result may apparently be of value in considering specific problems (e.g., $YBa_2Cu_3O_7$ high-temperature ceramics [8, 9]). Note that, for $j < j_c$, the resistance ρ_{flux} depends on the current distribution; therefore, we will consider an equation that simulates the evolution of the current density in half-space taking into account an increase in the electromagnetic field at the superconductor boundary.

As follows from the above, the definition $\rho_{flux}(b) = b^\alpha$ is meaningful. Then, for $m = 1/\alpha$, the problem has the self-similar solution $b_{ss} = (1 + t)^{1/\alpha}\theta_{ss}(\zeta)$ (where $\theta_{ss}(\zeta) = [1 - \alpha^{1/2}\zeta]^{1/\alpha}$ and $\zeta = x/(1 + t)$), which is merely a traveling wave. In the general case, the solution has the form

$$u_{ss} = (1 + t)^m \theta_{ss}(\zeta),$$

where $\zeta = x/(1 + t)^{(1 + m\alpha)/2}$. This solution is plotted in Fig. 2.

In such a statement, a nonlinear *I–V* characteristic is impossible to take into consideration: to do this it would be necessary to analyze functional relationships like $e = \phi(i, b)$ concurrently with a set of equations for the functions i and b. However, if the *I–V* characteristic (Eq. (3)) is linear, the current density i in the diffusion equation may be viewed as a parameter. Then, with the current fixed, its effect can be taken into account by substituting it for t in the solution mentioned above. As a result, we find the magnetic wave penetration depth

$$x_{ef}^{ss} \propto (it)^{(1 + m\alpha)/2}.$$

2. CURRENT DENSITY DISTRIBUTION

For nonlinear *I–V* characteristics, we perform a change of variables

$$E_x = E_c, \quad L_x = E_c\langle B_t'\rangle, \quad t_x = \mu_0 j_c E_c \langle B_t'\rangle^2$$

in the equation that relates the current to the electrical field (see above). As a result, we obtain the equation

$$\frac{\partial^2 e}{\partial x^2} = \frac{1}{n} e^{\frac{1-n}{n}} \frac{\partial e}{\partial \tau},$$

with the boundary conditions

$$\frac{\partial e}{\partial x}(0, \tau) = -1, \quad e(\infty, \tau) = 0, \quad e(x, 0) = 0.$$

Since for nonlinear characteristics $e = \phi(i)$, where ϕ is a given function, we have $i = i/i_c$, this equation can be reduced to the equation

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{1}{n}\phi(i)^{\frac{1-n}{n}} \frac{\partial \phi(i)}{\partial \tau} = \frac{1}{n}\phi(i)^{\frac{1-n}{n}} \phi'(i) \frac{\partial i}{\partial \tau} \qquad (10)$$

with the boundary condition

$$\phi'(i(0, t)) \frac{\partial i}{\partial x}(0, t) = -1 \quad (\langle b_t'\rangle = 1). \qquad (11)$$

For power characteristics, boundary condition (11) is conveniently written in the form

$$\frac{\partial i}{\partial x} = -\frac{1}{n} i_{|x = 0, t\rangle 0}^{1-n}, \qquad (11')$$

and Eq. (11) may be represented in the form of the Burgers equation

$$\phi''(i) \frac{\partial i}{\partial x} + \phi'(i) \frac{\partial^2 i}{\partial x^2} = \frac{1}{n}\phi(i)^{\frac{1-n}{n}} \phi'(i) \frac{\partial i}{\partial \tau} \qquad (10')$$

with the additional conditions $i(\infty, \tau) = 0$ and $i(x, 0) = 0$ for $x > 0$.

Note that, if the magnetic flux penetrates into the sample by diffusion, it may be assumed that

$$-\frac{\partial \langle i\rangle_t}{\partial x} \propto -\frac{\partial^2 \langle b\rangle_t}{\partial x^2} = -\langle b_t'\rangle,$$

where averaging is carried out over the scales t_x and L_x. Since these scales are chosen such that $\langle b_t'\rangle = \text{const}$, boundary condition (11') is simplified to

$$\frac{\partial \langle i\rangle_t}{\partial x} \propto \frac{1}{n}\langle i\rangle_t^{1-n} = \langle b'\rangle_{t|x = 0, t\rangle 0}.$$

Without loss of generality, we may put $\tau \longrightarrow \text{const}\,\tau$ in such a way that the condition $n\langle b'\rangle = 1$ is satisfied. Then, the above relationship can be written as $i^{1-n}(0, t) = 1$; i.e.,

$$i(0, t) = 1, \quad t > 0. \qquad (12)$$

Thus, we have arrived at the problem with the same boundary conditions as for the linear *I–V* characteristic, but the linear diffusion equation here is replaced by nonlinear Burgers equation (1'). The Burgers equation

is conveniently written as

$$\frac{\partial i}{\partial x} + \frac{1}{(n-1)} i^2 \frac{\partial^2 i}{\partial x^2} = \frac{1}{n(n-1)} i^{\frac{1+n}{n}} \frac{\partial i}{\partial \tau}. \quad (13)$$

Equation (13) has a simple physical meaning: for $n = 1$, it turns into a nonlinear equation for current diffusion [2] (but in terms of electromagnetic field). For $n \longrightarrow \infty$, the limiting equation is $\partial i/\partial x = 0$. This equation has the solution $i = i_c$, which corresponds to the Bean model: $(-\nabla b) \propto i_c$. For large but finite n, this equation reduces to the equation

$$(n-1)\frac{\partial i}{\partial} + \frac{\partial^2 i}{\partial x^2} = \frac{1}{n} i^{\frac{n+1}{n}} \frac{\partial i}{\partial \tau},$$

from which it follows that the steady-state distribution of the current density has the form

$$i(x) \propto i(0)\exp(1-n)x + \text{const}.$$

Here, $i(\partial) = 1$ by virtue of boundary condition (12) and the constant of integration can be chosen from the condition $i(\infty) = \text{const}$.

In the general case, we must consider the equation

$$\frac{\partial i}{\partial \tau} - n(n-1) i^{-\frac{n+1}{n}} \frac{\partial i}{\partial x} = n i^{-\frac{n+1}{n}} \frac{\partial^2 i}{\partial x^2}.$$

Performing the compression of the spatial variable, $x \longrightarrow n\langle b_t' \rangle x$, we recast this equation in the form

$$\frac{\partial i}{\partial \tau} - \frac{n-1}{\langle b' \rangle_t} i^{-\frac{n-1}{n}} \frac{\partial i}{\partial x} = \frac{1}{\langle b' \rangle_t} i^{-\frac{n-1}{n}} \frac{\partial^2 i}{\partial x^2}. \quad (14)$$

Let us determine the parameter $\varepsilon = \langle b' \rangle_t^{-1}$, which plays the role of average viscosity for Burgers equation (14). For $\varepsilon \longrightarrow 0$, we have

$$\frac{\partial i}{\partial \tau} - \frac{n-1}{\langle b' \rangle_t} i^{-\frac{n-1}{n}} \frac{\partial i}{\partial x} = 0. \quad (14')$$

Equation (14') has the form of the conventional Burgers equation without viscosity (where the spatial variable plays the role of time). This equation is more convenient to write as

$$\frac{dj}{dt}(x(t), t) = 0 \quad \text{with} \quad \frac{dx(t)}{dt} = -\frac{n-1}{\langle b_t' \rangle} i^{-\frac{n-1}{n}}. \quad (15)$$

Equalities (15) mean that every solution to Eq. (14') is time-invariable along straight lines given by

$$x + \frac{n-1}{\langle b' \rangle_t} i(x, t)^{-\frac{n-1}{n}} t = \text{const}.$$

It is known that solutions to the Burgers equation may contain turning-over waves; that is, the solution becomes multivalued in the vicinity of the point $j = j_c$ (see, for example, [10, p. 189]). Such a situation may occur at any $n \neq 1$ but is impossible in the critical current regime $i = 1$. Otherwise, we find that the current penetrates into the sample in the form of a simple wave whose shape is completely specified by the boundary condition.

3. LAW OF MOTION OF THE MAGNETIZATION FRONT

Let us "freeze" the coefficients in Eq. (14'). In other words, we consider the equation

$$\frac{\partial i}{\partial \tau} - \frac{n-1}{\langle b' \rangle_t} \left\langle i^{-\frac{n-1}{n}} \right\rangle_x \frac{\partial i}{\partial x} = 0,$$

replace the current mean $\langle i \rangle_x$ by the induction mean $\langle b' \rangle_x$, and assume that the magnetic field has the property of ergodicity $\langle b \rangle_x = \langle b \rangle_t$ (see Appendix). Then, differentiating this equation with respect to x in view of the relationship $\langle i \rangle \propto (-\langle b \rangle_x)$ yields

$$\frac{\partial u}{\partial \tau} - \frac{n-1}{\langle b_t' \rangle} \langle u \rangle_x \frac{\partial u}{\partial x} = 0 \quad (u = \partial i/\partial x). \quad (16)$$

As follows from the properties of solutions to Eq. (16), the law of motion of the magnetization front is given by the relationship

$$X(t) \propto \frac{n-1}{\langle b' \rangle_t} \langle b' \rangle_x^{\frac{n-1}{n} t},$$

where $\langle b' \rangle_t \longrightarrow n\langle b' \rangle_t$ by virtue of the above substitution.

A similar relationship was obtained in [2]:

$$x(t) \propto \langle b' \rangle_t^{\frac{n-1}{n+1}} t. \quad (17)$$

The law of motion of the magnetization front is conveniently expressed by

$$x(t) \propto \frac{n-1}{n} \frac{\langle b' \rangle_x}{\langle b' \rangle_t} t.$$

This expression and representation (17) coincide up to the order of smallness of $(n+1)^{-1} \sim n^{-1}$. This is because the Burgers equation takes into account the current spatial distribution, which is determined by the gradient $\partial i/\partial x$. Neglecting this term yields Eq. (17), since the problem (after corresponding redefinitions) is reduced to that considered in [2].

4. CONCLUSIONS

Thus, based on the Maxwell equations, we derived model equations: a diffusion nonlinear equation for magnetic induction and a Burgers equation that simulates the current distribution in view of external magnetic field variation at the boundary of half-space. For the latter equation, two boundary-value problems are considered: (1) the rate of change of the magnetic flux

at the boundary is constant and (2) the rate of change of the flux increases with time by a power law. Two situations in case (1) are possible: (1a) the resistance ρ_{flux} is constant and (1b) the resistance depends on b^α linearly or nonlinearly. In case 1a, $\alpha = 0$ and the flux penetrates into the sample with an infinite velocity (Fig. 2); in case 1b, the magnetic wave penetrates into the sample with a finite velocity if $\alpha > 1$ (Fig. 1).

If condition 1b in case 2 is satisfied, the wave invariably moves with a velocity higher than in a constant-resistance medium with the same boundary regime, since the resistance is an increasing function of the magnetic field. The magnetic wave front (the point at which b_{ss} vanishes) moves with the same velocity: $x_f(t) = \zeta_f(1 + t)^{(1 + m\alpha)/2}$. Schematically, the evolution of the self-similar process is shown in Fig. 2 (the path of the magnetic wave half-width is shown by the dashed line).

When investigating the current density distribution, we consider the Bean boundary conditions and study the shape of current fluctuations only in the vicinity of the critical state. In this case, the rate of pumping by magnetic current and the exponent of the *I–V* characteristic enter into the coefficients of the equation in implicit form. It turns out that, with $n \longrightarrow \infty$, we come to solutions to the conventional Bean model. For $v \longrightarrow 0$, where v is the rate of pumping by magnetic field, we obtain the standard diffusion equation for current density, which was considered, for example, in [3]. For $v \longrightarrow \infty$, the equation degenerates into the Burgers wave equation. Eventually, we again obtain the Bean model of critical state; now, however, at $j = j_c$, a solution represents a traveling wave with an amplitude j_c rather than being a constant. For large fluctuations near j_c, such a wave becomes unstable: it steepens giving rise (if the Hugoniot conditions at discontinuity [10] are satisfied) to a shock wave. However, this issue was not considered in detail. To conclude, the equation for magnetic field induction and that for current density evolution, although written in different forms, are formally identical: they can be expressed in the same form by making a differential change of variables. The solution of the joint set of equations was not considered in this work; however, the current distribution for a plate [4] may be involved in the coefficient $\rho_{flow}(b)$ in specific problems for the magnetic diffusion equation.

APPENDIX

Let averaging operators be designated as

$$M_t[B(x, t)] = \langle B(z)\rangle_{t_0} = \lim_{\nu \to \infty} \frac{1}{\nu} \int_{t_0}^{t_0 + \nu} B(x, t)dt$$

and

$$M_x[B(x, t)] = \langle B(x, t)\rangle_x = \frac{1}{L_x}\int_{t_0}^{L_x} B(x, t)dx$$

if the function $B(x)$ is L_x-periodic in x. Such a function can be expressed by the Fourier series

$$B(x) = \sum_{k>0} B_k \exp ikx.$$

Then, the application of the averaging operator leads to the relationship

$$M_x[B(x)] = B_0;$$

i.e., averaging yields the free term of the Fourier series.

Consider the function

$$B(\omega t) = \sum_{k>0} B_k \exp ik\omega t$$

and apply the time averaging operator. If there is a vector $k = (k_1, k_2)$, taking the corresponding two-dimensional integral for $x = \omega_1 x$ and $t = \omega_2 t$ yields (if ω_1 and ω_2 are rationally incommensurable numbers) the same result after averaging over variable x and time t [11]:

$$M_x[B(x)] = M_t[B(\omega t)].$$

It is this observation that was used above: in statistical mechanics, this relationship is known as the ergodicity of trajectories [10].

The basic reason for using this procedure was the need to make nonlinear equations dimensionless. For example, the substitutions $L_x = E_c/B_t'$ and $t_x = \mu_0 j_c E_c/(B_x')^2$ allowed Romanovskii [2] to recast the diffusion equation with a power *I–V* characteristic in dimensionless form. Here, the question arises as to how the parameters t_x and L_x should be treated. Obviously, it will suffice to perform averaging $\langle L_x\rangle = E_c/\langle B'\rangle_t$ in any sense. The result of averaging over t depends, generally speaking, on the parameter t_0 (for details, see [12]). In most applied problems, the properties of solutions to equations do not depend on t_0; however, this is sometimes not the case. For example, the averaging of a random pinning potential [4, p. 1159] required the following operation:

$$\langle \ldots\rangle = \frac{1}{t_0}\int_0^{t_0} (\ldots)dt,$$

where it is assumed that the averaging operator includes taking the limit at $\eta \longrightarrow \infty$ implicitly.

Another important problem is to establish the fundamental possibility that such characteristic scales exist (see, e.g., [4]). In the case of viscous flux creep, the characteristic time t_0 for a plate may be taken in the

form

$$t_0 = \frac{\pi}{2}\frac{Td^2}{|\partial_j U| c v_0 H},$$

where $U(j)$ is the activation barrier, d is the plate width, c is the velocity of light, v_0 is the "microscopic" velocity of vortices, and H is the magnetic field intensity.

ACKNOWLEDGMENTS

This work was partially supported by the Ministry of Education and Science of Ukraine (project no. 2M/71-2000).

REFERENCES

1. L. M. Fisher, I. B. Voloshin, V. S. Gorbachev, S. E. Savel'ev, and V. A. Yampol'skii, in *Proceedings of the 7th International Workshop on Critical Current in Superconductors, Alpbach, Austria, 1994*, Ed. by H. W. Weber, pp. 101–104.
2. V. R. Romanovskii, Zh. Tekh. Fiz. **70** (12), 47 (2000) [Tech. Phys. **45**, 1557 (2000)].
3. V. R. Romanovskii, Pis'ma Zh. Tekh. Fiz. **23** (3), 15 (1997) [Tech. Phys. Lett. **23**, 94 (1997)].
4. G. Blatter, M. V. Feidel'man, V. B. Geshkenbein, *et al.*, Rev. Mod. Phys. **66**, 1125 (1994).
5. A. A. Samarskiĭ, V. A. Galaktionov, S. P. Kurdyumov, and A. P. Mikhaĭlov, *Blow-Up Regimes in Problems for Quasi-Linear Parabolic Equations* (Nauka, Moscow, 1987).
6. Shi Dohglu, H. E. Konrous, Ming Xu, and D. H. Kim, Phys. Rev. B **43**, 514 (1991).
7. Y. Yeshurun and A. P. Malozemoff, Phys. Rev. Lett. **60**, 2202 (1988).
8. P. J. Kung, M. P. Maley, M. E. McHenry, *et al.*, Phys. Rev. B **46**, 6427 (1992).
9. M. V. Feigel'man, V. B. Geshkenbein, A. I. Larkin, and V. M. Vinokur, Phys. Rev. Lett. **63**, 2303 (1989).
10. G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to Nonlinear Physics: From Pendulum to Turbulence and Chaos* (Nauka, Moscow, 1988).
11. E. A. Grebennikov, *Method of Averaging in Applied Problems* (Nauka, Moscow, 1986).
12. V. M. Volosov, Usp. Mat. Nauk **17** (6), 3 (1962).

Translated by M. Astrov

Technical Physics, Vol. 48, No. 12, 2003, pp. 1540–1545. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 39–44.

SOLIDS

On a Relationship between the Tangential Component of Stray Field and Elastic Bending of a Ferromagnetic Steel Tube

V. G. Kulejew and V. V. Lopatin

Institute of Metal Physics, Ural Division, Russian Academy of Sciences, ul. S. Kovalevskoĭ 18, Yekaterinburg, 620219 Russia

e-mail: kulejew@imp.uran.ru

Received January 29, 2003; in final form, April 17, 2003

Abstract—The tangential component of stray field on the surface of an elastically bent tube is evaluated under the assumption that the internal magnetic field is constant. This component bears more information than the normal one, since it allows one to determine the mean value of internal stresses of the first kind. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

When a ferromagnetic steel tube is bent (hereafter it is assumed that its ends are hinged), it takes the shape of a half-sinusoid [1]:

$$y = y_0 \cos\frac{\pi x}{l_0}; \quad -\frac{l_0}{2} \le x \le \frac{l_0}{2}. \tag{1}$$

Here, l_0 is the tube length and y_0 is the sag in the midsection $x = 0$, where the bending stresses are maximal (Fig. 1a). The sag specifies the curvature ρ and maximal value of bending stresses σ_b^m in the midsection of the tube along the generatrices $\alpha = 0$ (extension) and $\alpha = 180°$ (compression) of the surface $r_1 = r_2$ (r_1 and r_2 are the inner and outer radii of the tube, respectively, and α is the azimuth coordinate measured from the surface of bend [1]):

$$\sigma_b^m = \frac{Er_2}{\rho} = \frac{Nr_2}{J} = E\frac{\pi^2}{l_0^2}y_0 r_2, \tag{2}$$

where E is the Young's modulus; N is the bending moment; and J is the inertia moment of the tube section:

$$J = \frac{1}{4}\pi(r_2^4 - r_1^4), \quad \frac{1}{\rho} = \frac{N}{EJ} = -\frac{d^2y}{dx^2}. \tag{3}$$

As a result, bending stresses at an arbitrary point of the tube are given by [1]

$$\sigma(x, \alpha, r) = \sigma_b^m \frac{r}{r_2}\cos\alpha\cos\frac{\pi x}{l_0}; \quad r_1 \le r \le r_2. \tag{4}$$

If the tube is long, shear stresses can be neglected. Since the tube is in the weak terrestrial magnetic field H_0 ($H_0 \ll H_c$, where H_c is the coercive force of steel), this field induces a low magnetization $M(H_{0x})$. The action of high bending stresses ($\sigma_b^m > \bar{\sigma}_i$, where $\bar{\sigma}_i$ is the mean value of internal stresses [2]) causes irreversible displacements of 90° domain walls [2, 3] and, accordingly, a significant magnetoelastic increment of the magnetization, $\Delta M_\sigma \gg M(H_{0x})$.

Although the increment ΔM_σ depends on σ nonlinearly and tends to saturate [2, 3], it invariably increases with the stress. Therefore, the magnetization distribution in an elastically bent tube will copy, albeit with a

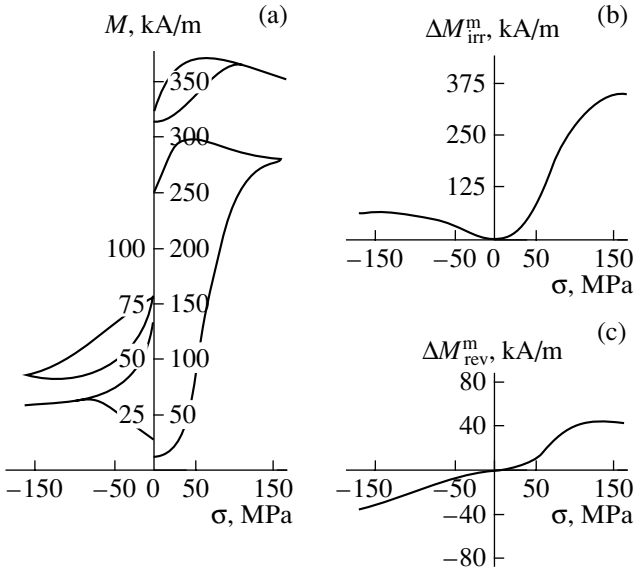


Fig. 1. Magnetoelastic variation of the magnetization in St 3 steel subjected to a permanent magnetic field of 40 A/m. (a) Application and relief of tensile and compressive stresses; the closed loops are their steady variations. (b, c) Amplitude dependences of the irreversible (ΔM_{irr}^m) and reversible (ΔM_{rev}^m) parts of the total magnetoelastic increment ΔM_σ.

nonlinear correction, the distribution $\sigma_b(x, \alpha, r)$ of bending stresses. Eventually, all the features of the function $\sigma_b(x, \alpha, r)$ (its extrema, their positions, the "wavelength" of the stress, and points where the function changes sign) will be involved in the function $\Delta M_\sigma(x, \alpha, r)$.

A stray magnetic field $\Delta\mathbf{H}(\Delta M_\sigma)$ around the tube will bear information on the magnetization distribution. Owing to the relationship between $\sigma_b(x, \alpha, r)$ and $\Delta M_\sigma(x, \alpha, r)$, one can judge the value and distribution of bending stresses in the tube by measuring the stray field.

In [4], we found the distribution of magnetization magnetoelastic increments $\Delta M_\sigma^\pm(x, \alpha, r)$ in an elastically bent tube (see (4)) subjected to a weak longitudinal magnetic field H_{0x}, knowing the stress distribution in the tube and the values of the increments $\Delta M_\sigma^\pm$ due to uniform extension (ΔM_σ^+) or compression (ΔM_σ^-) [2, 3]. From this distribution, we determined the volume magnetic charge distribution $\rho_m^\pm(x, \alpha, r)$ and then, using available solutions to the equations of magnetostatics [5, 6] and a number of approximations, the stray field $\Delta H_r(x, \alpha, r)$ normal to the tube surface [4].

At the same time, it is obvious that the tangential component $\Delta H_x(x, \alpha, r)$ of the stray field may also be of interest for nondestructive inspection, since it bears information on the distribution of the magnetizations $\Delta M_\sigma^\pm$ in an elastically bent tube and, accordingly, on the stress distribution in it. Therefore, in this work we seek a relationship between the stray field tangential component on the tube surface and elastically bending stresses.

DISTRIBUTION OF LONGITUDINAL MAGNETIZATION IN AN ELASTICALLY BENT TUBE

As is well known [2, 3], a magnetoelastic increment of magnetization may take place in the presence of at least a weak magnetic field and elastic stresses. It is the sum of the regular (reversible) change of magnetization $\Delta M_{rev}^\pm(H_0, \sigma_\pm)$, which is positive in the extended part of the tube (+) and negative in its compressed part (–), and the general irreversible increment $\Delta M_{irr}^\pm(H_0, \sigma_\pm)$, which is always positive and makes a major contribution to $\Delta M_\sigma^\pm$ [2, 3]. As the number of loading cycles (the application and relief of a stress of the same amplitude $\sigma_\pm^m$) grows, the values of $\Delta M_{irr}^\pm$ increase, tending to saturation: $\Delta M_{irr}^\pm(t \longrightarrow \infty) \longrightarrow \Delta M_{irr}^m(H_0, \sigma_\pm^m)$, where t is time. Considering that the tube may experience several loading cycles, the time variation of the stress at each point (x, α, r) of the tube is conveniently described by the expressions [2–4]

$$\sigma_\pm(t) = \sigma_\pm^m(x, \alpha, r)\sin^2\left(\frac{\omega t}{2}\right) = \pm\sigma_m\sin^2\left(\frac{\omega t}{2}\right), \quad (5)$$

where the circular frequency ω is assumed to be so small that the loading and unloading processes are quasi-static. According to (5), the first application of the load of amplitude σ_m corresponds to $\omega t = \pi$; the first relief, to $\omega t = 2\pi$; and so on. Eventually, we have [3]

$$\Delta M_\sigma^+(H_0, \sigma_\pm, t) = \Delta M_{irr}^\pm(H_0, \sigma_\pm, t) + \Delta M_{rev}^\pm(H_0, \sigma_\pm, t),$$

$$\Delta M_{irr}^\pm(t) = \Delta M_{irr}^m(H_0, \sigma_\pm^m)\left[1 - \exp\left(-\frac{\omega t}{4\pi}k^\pm\right)\right], \quad (6)$$

$$\Delta M_{rev}^\pm(t) = \pm\Delta M_{rev}^m(H_0, \sigma_\pm^m)\sin^2\left(\frac{\omega t}{2}\right).$$

The values of ΔM_{irr}^m and ΔM_{rev}^m depend on the elastic stress amplitude at a given point of the tube in a given field H_0, and the parameters $k^\pm$ specify the rate of approach of $\Delta M_{irr}^\pm(t)$ to saturation.

Figure 1a shows the magnetoelastic variations of the magnetization in St 3 steel subjected to a permanent inner magnetic field of 40 A/m and experiencing uniform load cycling (a stress amplitude of ±150 MPa). Also shown is the steady regular variation of the magnetization (tenth cycle).

As the amplitude σ_m grows, both components (6) first increase and then tend toward saturation (in the same field). Curves for $\Delta M_{irr}^m(\sigma_\pm^m)$ and $\Delta M_{rev}^m(\sigma_\pm^m)$ in the range $0 \le \sigma_m \le 150$ MPa are demonstrated in Fig. 1b. As was shown [2, 3], the amplitude dependences of ΔM_{irr}^m and ΔM_{rev}^m generally obey the exponential law

$$\Delta M_{irr}^m(\sigma_\pm^m) = \Delta M_1^\pm(H_0)\left\{1 - \exp\left[-\left(\frac{\sigma_\pm^m}{\sigma_1^\pm}\right)^2\right]\right\},$$
$$\Delta M_{rev}^m(\sigma_\pm^m) = \Delta M_2^\pm(H_0)\left\{1 - \exp\left[-\left(\frac{\sigma_\pm^m}{\sigma_2^\pm}\right)^2\right]\right\}, \quad (7)$$

where the parameters $\sigma_1^\pm$ and $\sigma_2^\pm$ should be determined from experiments with given steel [2, 3].

In a first approximation, the low-field dependences of $\Delta M_1^\pm$ and $\Delta M_2^\pm$ are linear:

$$\Delta M_1^\pm(H_0) = B_1^\pm H_0, \quad \Delta M_2^\pm(H_0) = B_2^\pm H_0. \quad (8)$$

From (8) and Fig. 1, one easily evaluates the magnetization magnetoelastic increments in St 3 steel for any other field if it is low: $H_0 \ll H_c$. Straightforward pro-

cessing of the data in Figs. 1a and 1c according to (6)–(8) yields the following model parameters (for St 3 steel): $\sigma_1^+ = \sigma_2^+ = \sigma_1^- = \sigma_2^- = \sigma_0 = 100$ MPa, $B_1^+ = 740$, $B_2^+ = 107$, $B_1^- = 150$, $B_2^- = 80$, $k^+ = 3$, and $k^- = 6$. From (6)–(8), it is easy to derive a general expression for the distribution of the magnetization $\Delta M_\sigma^\pm$ in a bent tube:

$$\Delta M_\sigma^\pm[H_0, t, \sigma_b(x, \alpha, r)] = A^\pm(H_0, t)\left[1 - \exp\left(-\left[\frac{\sigma_\pm^m}{\sigma_0}\right]^2\right)\right],$$
$$A^\pm(H_0, t) = \left\{B_1^\pm\left[1 - \exp\left(-\frac{\omega t}{4\pi}k^\pm\right)\right] \pm B_2^\pm \sin^2\frac{\omega t}{2}\right\}, \quad (9)$$

where only the stress amplitude is a function of (x, α, r): $\sigma_\pm^m$ always equals $\sigma_b(x, \alpha, r)$.

STRAY FIELD NEAR AN ELASTICALLY BENT TUBE

Expressions (10) in view of (4) yield volume magnetic charges of density [5]

$$\rho_m^\pm(x, \alpha, r) = -4\pi\,\mathrm{div}(\Delta M_\sigma^\pm) = -4\pi\frac{\partial}{\partial x}(\Delta M_\sigma^\pm). \quad (10)$$

Substituting (9) and (4) in (10), we have

$$\rho_m^+(x, \alpha, r) = A^\pm\frac{\pi}{l_0}4\pi\left(\frac{\sigma_{bent}^m}{r_2}\right)^2 r^2\cos^2\alpha \times \sin\frac{2\pi x}{l_0}\exp\left[-\left(\frac{\sigma_\pm^m}{\sigma_0}\right)\right]. \quad (11)$$

In this case, magnetic potentials $\varphi^\pm(x', \alpha', r')$ take the form [5]

$$\varphi^\pm = A^\pm\int_{-\frac{l_0}{2}}^{\frac{l_0}{2}}dx\int_{r_1}^{r_2}dr\int d\alpha^\pm\frac{r\rho^\pm(x, \alpha^\pm, r)}{R(x, \alpha^\pm, r, x', \alpha', r')}, \quad (12)$$

where

$$R^2 = (x' - x)^2 + r^2 + r'^2 - 2rr'\cos(\alpha' - \alpha^\pm). \quad (13)$$

The total magnetic potential $\varphi(x', \alpha', r')$ of an elastically bent tube is the sum of expressions (12) in view of (13). The stray field components are calculated by the formulas [5]

$$\Delta H_x(x', \alpha', r') = -\frac{\partial\varphi}{\partial x'} = A^+J_1^+ + A^-J_1^-, \quad (14)$$

$$\Delta H_r(x', \alpha', r') = -\frac{\partial\varphi}{\partial r'} = A^+J_2^+ + A^-J_2^-, \quad (15)$$

where

$$J_1^\pm = \int_{-\frac{l_0}{2}}^{\frac{l_0}{2}}dx\int_{r_1}^{r_2}dr\int d\alpha^\pm\frac{(x' - x)r^3\cos^2\alpha^\pm\sin\frac{2\pi x}{l_0}\exp\left(-\frac{\sigma_m^2}{100^2}\right)}{[(x' - x)^2 + r^2 + (r')^2 - 2rr'\cos(\alpha^\pm - \alpha)]^{\frac{3}{2}}}, \quad (16)$$

$$J_2^\pm = \iiint\frac{[r' - r\cos(\alpha^\pm - \alpha')]r^3\cos^2\alpha^\pm\sin\frac{2\pi x}{l_0}\exp\left(-\frac{\sigma_m^2}{100^2}\right)}{[(x' - x)^2 + r^2 + (r')^2 - 2rr'\cos(\alpha' - \alpha^\pm)]^{\frac{3}{2}}}dx\,dr\,d\alpha^\pm. \quad (17)$$

In (12)–(17), the limits of integration over α^+ and α^- are the same:

$$\int d\alpha^+ = \left(\int_0^{\frac{\pi}{2}} + \int_{\frac{3\pi}{2}}^{2\pi}\right)d\alpha^+ = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}}d\alpha^+;$$
$$\int d\alpha^- = \int_{\frac{\pi}{2}}^{\frac{3\pi}{2}}d\alpha^-. \quad (18)$$

Expressions (12)–(18) are exact solutions to magnetostatic problem (9) for a given distribution of magnetization magnetoelastic increments in an elastically bent tube under the assumption that the inner magnetic field is constant.

CALCULATION OF STRAY FIELD COMPONENTS FOR A SPECIFIC TUBE

Since we measured stray fields on the surface of an elastically bent tube made of St 20 steel, it seems reasonable to calculate the fields for this object. The length of the tube is 2 m, $r_1 = 10.5$ mm, $r_2 = 13.5$ mm, and $\sigma_t = 250$ MPa. Bending stresses were applied to the tube placed vertically, so that the vertical component of

the terrestrial magnetic field H_{0x} = 30 A/m was aligned with the tube axis. Since the magnetization magnetoelastic increments are linear functions of the field (see (9)), the effective values of the increments are those given in Fig. 1b times 0.74. The magnetization due to this field is $M(H_{0x}) \cong 8$ kA/m.

Let σ_b^m be 133 MPa. From (1)–(3), one easily finds that the sag in this case is y_0 = 20 mm. Using these data and the model parameters for St 3 steel, we calculated the tangential, $\Delta H_x(x)$, and normal, $\Delta H_1(x)$, components of the stray field, taking into account the contribution from the uniform magnetization $M(H_{0x})$. To do this, we supplemented expressions (14) and (15) for the magnetization magnetoelastic increments with the field $\delta H(M(H_{0x}))$ of a point dipole (a base 2 m long with magnetic charges $M(H_{0x})S$ = 0.00181 kA m at its ends). The components of the dipole field are

$$\delta H_x = [M(H_{0x})] = 18.1\left\{\frac{x+100}{[(x+100)^2+2.4]^{\frac{3}{2}}} - \frac{x-100}{[(x-100)^2+2.4]^{\frac{3}{2}}}\right\}, \quad (19)$$

$$\delta H_r = [M(H_{0x})] = 18.1\left\{\frac{1.85}{[(x+100)^2+3.42]^{\frac{3}{2}}} - \frac{1.85}{[(x-100)^2+3.42]^{\frac{3}{2}}}\right\}. \quad (20)$$

The integrals involved in (14) and (15) cannot be taken analytically. Therefore, we used numerical integration algorithms to obtain an approximate solution.

RESULTS AND DISCUSSION

Figure 2 shows the tangential component distribution along the tube on its five generatrices, namely, α = 0° (maximal tensile stresses in the bent tube), α = 45°, α = 90° (the neutral filament of the tube, where $\sigma_b = 0$), α = 135°, and α = 180° (the generatrix where compression stresses are maximal), in both loaded ($\omega t = \pi$, Fig. 2a) and unloaded ($\omega t = 2\pi$, Fig. 2b) states. It is seen that the curves in Figs. 2a and 2b differ only numerically: the extrema of the function $\Delta H_x(x)$ in the unloaded state are somewhat higher than in the loaded state, which is explained by the increased mean magnetization in the former case [4].

The function $\Delta H_x(x)$ varies to the greatest extent on the generatrix α = 0° and to the least extent on the generatrix α = 180°. This is because the value of ΔM_σ^- is smaller than ΔM_σ^+. The rest of the curves $\Delta H_x(x)$ (on the generatrices α = 45°, 90°, and 135°) lie between these two curves. The curves on the generatrices α = 135° and 180° virtually coincide with each other.

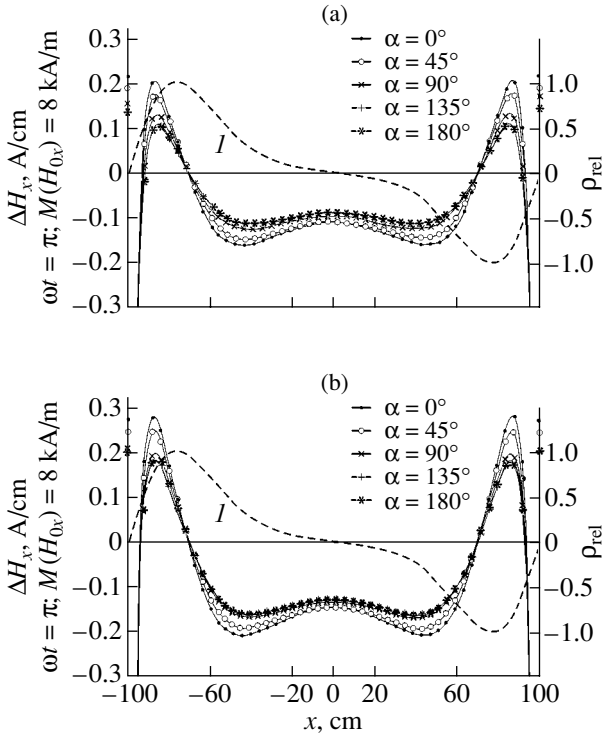


Fig. 2. Variation of the tangential component $\Delta H_x(x)$ of the stray field due to the magnetization magnetoelastic increment on the surface of the elastically bent tube along its generatrices α = 0°, 45°, 90°, 135°, and 180° (a) under maximal bending stresses and (b) under unloading. Dashed curve *1* shows the variation of the relative magnetic charge density $\rho_{rel} = \rho_m/\rho_m^{max}$ along the tube.

The functions $\Delta H_x(x)$ each have three maxima (at x = 0 and ±90 cm) and two minima (at x = 43 cm). Dashed line *1* shows the variation of the relative volume magnetic charge density $\rho_{rel}(x) = \rho_m^{\pm}/(\rho_m^{\pm})_{max}$. From Figs. 2a and 2b, it follows that the extrema of $\rho_{rel}(x)$, x = ±64 cm coincide with the zeros of $\Delta H_x(x)$. The curves $\Delta H_x(x)$ (Fig. 2) are almost completely associated with the magnetoelastic increment of the magnetization, since fields (19) and (20) have very narrow extrema at the end faces of the tube ($0.98 \le x \le 1$ m) and are close to zero in the remaining part.

The α dependences of ΔH_x are omitted, since their shape at any x = const remains the same: a broad peak at α = 0°, followed by a smooth decline as α tends to ±180°. All the curves $\Delta H_x(\alpha)$ are symmetric about α = 0°. Expressions (12), (13), (15), (17), and (20) for the normal component ΔH_r of the stray field are given here, since in [4] this field was calculated approximately. Figures 3a and 3b demonstrate the exact calculation of the curves $\Delta H_r(x)$ on five generatrices of the tube (α = 0°, 45°, 90°, 135°, and 180°).

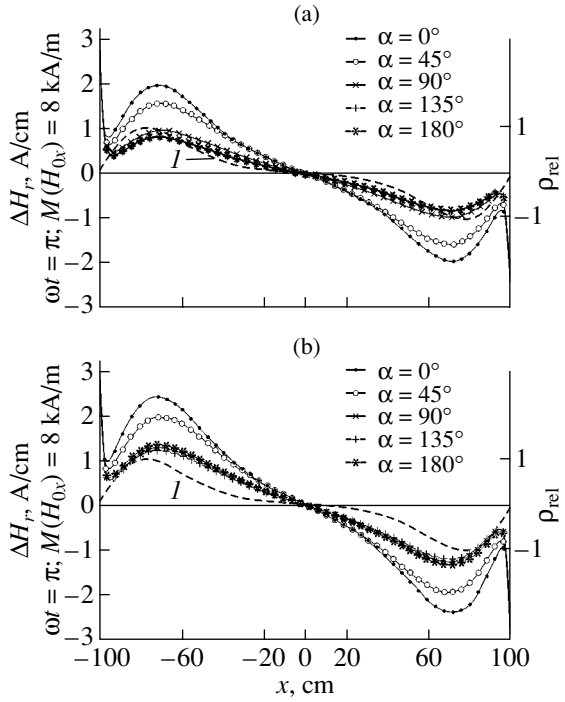


Fig. 3. Variation of the normal component $\Delta H_r(x)$ of the stray field due to the magnetization magnetoelastic increment along the tube generatrices $\alpha = 0°$, 45°, 90°, 135°, and 180° (a) under maximal bending stresses and (b) under unloading; (*1*) the function $\rho_{rel}(x)$.

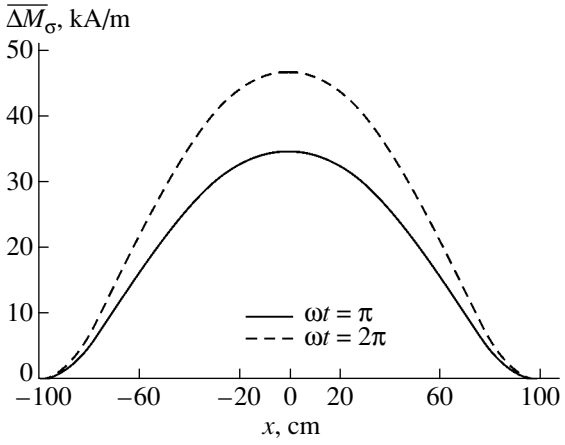


Fig. 4. Variation of the cross-section-averaged magnetization along the tube. $\omega t = \pi$, loading; $\omega t = 2\pi$, unloading.

As follows from Figs. 3a and 3b, the curves $\Delta H_r(x)$ taken at various α are similar to each other and to the curve $\rho_{rel}(x)$ (dashed line *1*): each of them has two extrema (at $x = \pm 70$ cm) and two inflection points. In the unloaded state ($\omega t = 2\pi$), the curves $\Delta H_r(x)$ are similar to those in the loaded state but run 20% higher. As with $\Delta H_x(x)$, this is because the mean magnetization somewhat increases upon unloading [4]. At the ends of the tube, the contribution of magnetic charges due to the uniform magnetization $M(H_{0x})$ prevails. As a result, the resultant normal component of the stray field tends to the value predicted by (31): $\Delta H_r \longrightarrow \pm 420$ A/m with $x \longrightarrow \pm 1$ m.

Finally, consider the distribution of the cross-section-averaged magnetization

$$\overline{\Delta M_\sigma}(x) = \overline{\Delta M_\sigma^+}(x) + \overline{\Delta M_\sigma^-} \qquad (21)$$

along the tube. The first term on the right of (21) is the average over the extended half-section of the tube; the latter is the average over the compressed half-section. Such averaging, which is easy to perform with expression (9) in view of (4), is of interest, since it is the cross-section-averaged magnetization that is measured by the ballistic method (using a feedthrough coil).

Figure 4 demonstrates the results of such a calculation. Once the bending load has been relieved ($\omega t = 2\pi$), the curve $\overline{\Delta M_\sigma}(x)$ runs somewhat higher than in the stressed state ($\omega t = \pi$). The bell-shaped form of this curve suggests that it, in general, reflects adequately the distribution of bending stresses in the tube.

Having calculated the tangential and normal components of the stray field (Figs. 2, 3), as well as the volume magnetic charge density (Figs. 2, 3, curves *1*) and the distribution of the cross-section-averaged magnetization (Fig. 4), we can trace the manner in which elastic stresses (4) appear in these curves.

The curve $\sigma_b(x)$ is a half-sinusoid and, accordingly, has a bell-shaped form with a peak at the center of the tube. Curves $\Delta M_\sigma^\pm(x)$ (9) also have a bell-shaped form: they reach a maximum at the center of the tube and fall to zero at its ends. However, the values of $\Delta M_0^\pm$ exponentially depend on the stress amplitude squared (see (9)). Therefore, one can assume that the dependences $\Delta M_\sigma(x)$ copy qualitatively the dependences $\sigma_b(x)$: as σ_b increases, so does ΔM_σ, and both quantities vanish simultaneously.

The derivative of $\Delta M_\sigma(x)$ with respect to x defines the volume magnetic charge density $\rho_{rel}(x)$ in the tube. Clearly, the derivative of the bell-shaped curve $\Delta M_\sigma(x)$ yields two extrema of the function $\rho_{rel}(x)$: a maximum and a minimum with two inflection points in between, which arise because of the exponential

$$\exp\left[-\left(\frac{\sigma_b^m}{\sigma_0 r_2}\right) r^2 \cos^2\alpha \cos^2\frac{\pi x}{l_0}\right]. \qquad (22)$$

It is easy to check that this exponential is minimal at the center of the tube ($x = 0$) and equals unity at its ends ($x = \pm l_0/2$). If the exponent in (22) were much smaller than unity at any point of the tube ($\sigma_b < \sigma_0$ in this case), we could expand expressions (7) into a series and leave only the first-order term, as was done in [4]. In this case, the extrema of the function $\rho_{rel}(x)$ were at the points $x = \pm 50$ cm. Exponential (22), appearing in (10),

shifts the extrema closer to the ends of the tube: $x = \pm 70$ cm (Figs. 2, 3, curves *1*).

Thus, measuring the normal component of the stray field, one can judge the distribution $\rho_{rel}(x)$ of volume magnetic charges in an elastically bent tube and find the region (near its end faces) where the effect of magnetic charges is the most significant. Comparing the values of the field $\Delta H_r(x)$ at the points $x = \pm l_0/2$ with its value given by (20) will make it possible to determine the uniform magnetization. Also, graphically integrating the curve $\Delta H_r(x)$, one can recover the distribution of the cross-section-averaged magnetization $\overline{\Delta M_\sigma}(x)$ (21) and, thereby, judge the distribution of bending stresses in the tube.

The tangential component of the stray field (Fig. 2) bears more information than the normal one. In moving from the end of the tube toward its center, the first zero of the function $\Delta H_x(x)$ defines the boundary of the region where the influence of magnetic charges is significant. The second zero is the position of the extrema of the functions $\rho_{rel}(x)$ and $\Delta H_r(x)$. The positions of the minima of the function $\Delta H_x(x)$ ($x = \pm 45$ cm in our case) specify the region where exponential (22) has a great effect on the stray field components. In this region, $|\sigma_b(x)| > \sigma_0$, where σ_0 is a characteristic stress outside this region. It should be noted that this region cannot be found from the positions of the inflection points. The minima of the curve $\Delta H_x(x)$ (the points $x = \pm 43$ cm) correspond to the characteristic point of variation of exponential (22): $\sigma_b^m = \sigma_0 = \overline{\sigma}_i$. Thus, the mean value of internal stresses of the first kind in the tube material may be found from the position of the minima.

CONCLUSIONS

Nondestructive inspection techniques based on measuring the tangential and normal components of a stray field on the surface of an elastically bent tube complement each other: although the tangential field is more sensitive to the features of the $\rho_{rel}(x)$ curve and, accordingly, provides more information, the normal field distribution allows one to recover the distribution of bending stresses in the tube.

Another advantage of the tangential field technique is the possibility of finding regions where the bending stress exceeds the stress σ_0 that is characteristic of given steel. This characteristic stress is related to the mean value of internal stresses in the steel.

A disadvantage of the tangential field technique is that this component is five to six times lower than the normal one. This imposes stringent requirements on the accuracy of positioning of the ferroprobe core.

REFERENCES

1. N. M. Belyaev, *Problems in the Strength of Materials* (Fizmatgiz, Moscow, 1956; Pergamon, Oxford, 1966).
2. V. G. Kulejew and M. B. Rigmant, Fiz. Met. Metalloved. **79**, 120 (1995).
3. V. G. Kulejew and M. B. Rigmant, Defektoskopiya, No. 9, 79 (1994).
4. V. G. Kulejew, A. A. Dubov, and V. V. Lopatin, Defektoskopiya, No. 1, 13 (2002).
5. K. M. Polivanov, *Fundamentals of Electrical Engineering* (Énergiya, Moscow, 1969), Part 3.
6. V. G. Kuleev, Defektoskopiya, No. 6, 65 (2002).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 12, 2003, pp. 1546–1553. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 45–53.

OPTICS, QUANTUM ELECTRONICS

X-ray Topographic Identification and Measurement of Plastic Strains and Elastic Stresses in Single Crystallites of Polycrystalline Diamond Layers

G. F. Kuznetsov

Institute of Radio Engineering and Electronics (Fryazino Branch), Russian Academy of Sciences, pl. Vvedenskogo 1, Fryazino, Moscow oblast, 141120 Russia

e-mail: gfk217@ire216.msk.su

Received December 24, 2002; in final form, May 22, 2003

Abstract—The well-known phenomenon of asterism is used as the basis for the development of an X-ray topographic method to identify and measure plastic strains and residual elastic stresses in single crystallites more than 3 μm in size in polycrystalline diamond layers. The amount of asterism is used as a quantitative measure of plastic strains in crystallites. The distribution of crystallites over the amount of asterism in 40- to 670-μm-thick microwave-plasma-deposited polycrystalline diamond layers is obtained. Shear plastic strains, which cause a misorientation from 0.4′ to 1.5° between different areas of a crystallite, are observed for the first time. The residual elastic stresses calculated in plastically strained crystallites vary between 2.7 kPa and 0.84 GPa.

INTRODUCTION

As applied to diamonds, X-ray topography (XRT) has been used mainly to examine the defect structure in natural diamond single crystals [1–3] and cracking in homoepitaxial diamond films [4]. Using many-crystal diffractometry, researchers have discovered and studied the anomalous transmission of X rays in natural diamond crystals for $\mu t < 1$ (μ is the linear X-ray absorption coefficient and t is the thickness of the material). X-ray diffractometry (XRD) was used at times for the phase analysis of artificial (synthesized) polycrystalline diamond layers (PDLs); however, the resolving power of these methods leaves much to be desired. More frequently, XRD methods were used to measure macroscopic elastic stresses in a PDL as a whole [5]. However, our recent results [6] indicate that the averaging of elastic stresses over all PDL crystallites is totally incorrect, since it is known today that residual elastic stresses in individual plastically strained crystallites of a layer may differ by three to five orders of magnitude, as demonstrated in this study. Therefore, we propose another approach: to detect plastically strained crystallites separately using the phenomenon of asterism, to measure the amount of asterism from diffraction reflections on the diffraction patterns, to calculate the misorientation angle between different areas of the crystallite lattice, to calculate residual elastic stresses in plastically strained crystallites, and to construct the crystallite distribution over the amount of asterism. In addition, such an approach allows one to distinguish between elastically and plastically strained crystallites.

In this work, we estimate mostly elastic strains localized within a crystallite. The calculation of macroscopic elastic strains in the PDL–substrate system as a whole will be the subject of further investigation. It is important to conceive the origin of local elastic stresses that cause plastic strains in several or most crystallites in the course of PDL growth on substrates [6], whereas neighboring crystallites in the same PDL remain elastically (but not plastically!) strained. The mathematical foundation of an XRT method developed by the author for the quantitative evaluation of plastic strains in individual crystallites is given below. In this method, the misorientation between individual crystallites is measured from the diffraction pattern (topogram in the context of this article), and then residual elastic stresses in these crystallites, albeit roughly (on the model level), are calculated.

The need for studying thick ($\geqslant 1$ μm) PDLs by nondestructive XRT methods has become topical. In [6], a brief review of methods for PDL study is given and an original XRT method to measure crystallite sizes and find the crystallite size distribution in 40- to 670-μm-thick PDLs is proposed. Because of the low X-ray absorption coefficient of diamond, the nondestructive XRT method can be used to examine PDLs throughout the range of practical thicknesses: from several micrometers to 1–5 mm.

EQUIPMENT AND METHODS FOR XRT STUDY OF POLYCRYSTALLINE DIAMOND LAYERS

Polycrystalline diamond layers 80 to 670 μm thick were grown on single-crystal silicon substrates in a microwave plasma-chemical reactor. Thinner layers (1–40 μm) were synthesized on silicon by the hot fila-

1063-7842/03/4812-1546$24.00

ment method. The layers start growing at many chaotically oriented nucleation centers (of size ≈5 nm) [6]. These centers grow up to crystallites and eventually give rise to the polycrystalline structure of the layers.

The real structure of PDLs was examined by a number of XRT and XRD methods. Based on the method of divergent beam from a quasi-point source (DBQPS) [7–10] (URS-0.1 X-ray equipment), researchers at the Institute of Radio Engineering and Electronics, Russian Academy of Sciences, have developed an original XRT technique [6] for the numerical evaluation of the crystallite size distribution in an irradiated PDL area. Its essence is that discontinuous Debye rings recorded on a standard-size (9 × 12 cm) photographic plate, which is used to take Laue patterns with an RKSO-1 X-ray camera, contain diffraction reflections only from those irradiated crystallites whose orientation meets the Bragg diffraction condition (for the most part from low-index crystallographic planes, such as {111}, {220}, or {311}). Note also that these reflections are the most intense in polycrystalline diamond. Characteristic radiations from Mo$K\alpha_{1,2}$, Cu$K\alpha_{1,2}$, Ni$K\alpha_{1,2}$, or Co$K\alpha_{1,2}$ anodes may be used.

The preferential orientations of crystallites were determined by a more rapid XRD method [7–10] using monochromatic Cu$K\alpha_1$ radiation. It was implemented with a double-crystal X-ray spectrometer–diffractometer built around a DRON-2 diffractometer. The new setup may accommodate (n; $-n$) geometry with the automatic recording of diffraction curves by a potentiometer.

In the configuration adopted, the DBQPS method [7–10] provides a linear geometrical resolution of about 7 μm when polychromatic (or white) X-ray radiation is used. For Co$K\alpha_1$ (λ = 1.78892 Å) and Mo$K\alpha_1$ (λ = 0.70929 Å) radiation, the resolution is 3 and about 1 μm, respectively [7–10]. The photographic resolution of the thick-film plates used is about 3 μm. This means that the XRT method makes it possible to record discrete reflections from individual crystallites of size ≥3 μm. With a smaller grain size, the Debye rings are expected to become continuous when Mo$K\alpha_1$ or Mo$K\alpha_2$ radiation, as well as higher wavelength Cu$K\alpha_{1,2}$ or Co$K\alpha_{1,2}$ radiation, is used. Indeed, in the case of fine-grained samples of thickness from 1 to 40 μm, the rings observed in the topograms are, as a rule, continuous. However, under specific growth conditions (low rate of primary and secondary nucleation), some of the crystallites in layers of thickness $h \geq 40$ μm were large enough for the intensity distribution to be discontinuous not only in the Debye rings due to the characteristic radiation but also in reflections due to polychromatic X-ray radiation.

PLASTIC STRAIN IN CRYSTALLITES OF POLYCRYSTALLINE DIAMOND LAYERS: DETECTION AND MEASUREMENT

Topograms taken of relatively thick (40–670 μm) PDL samples demonstrated [6] the well-known phenomenon of asterism of diffraction spots [11] from a set of crystallites. Asterism shows up as a considerable elongation of diffraction spots from crystallites in the Bragg direction compared with the size of the same spots in the azimuthal direction; i.e., normal diffraction spots transform into diffraction "tails" along the radial or Bragg direction. In the topogram from a PDL sample 80 μm thick, the Bragg size of individual reflections is almost equal to, or even somewhat smaller than, their azimuthal size (Fig. 1a). This means that the corresponding crystallites are not prone to plastic deforma-

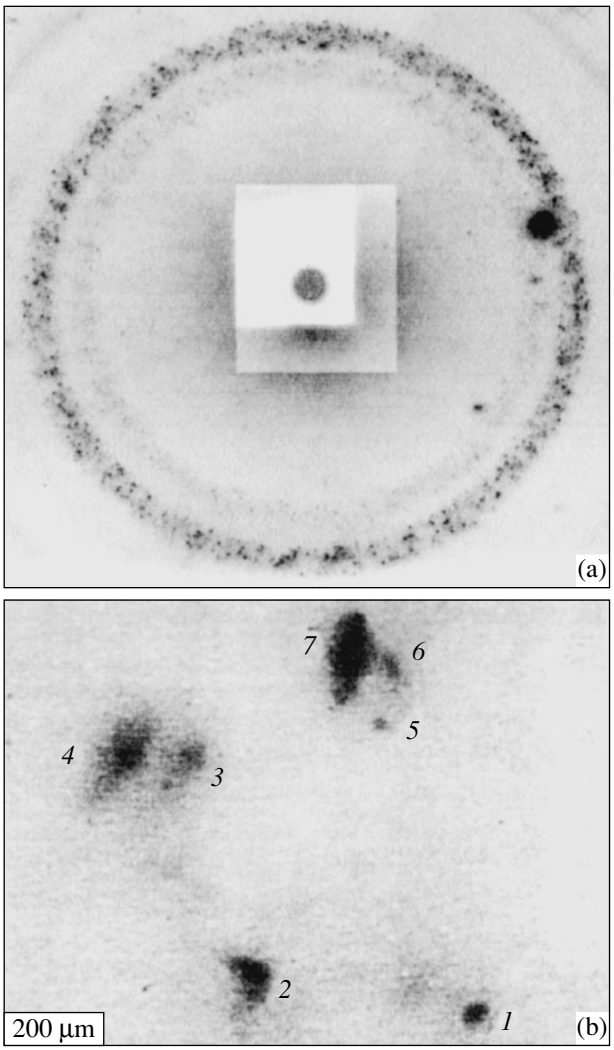


Fig. 1. (a) General view of the X-ray topogram from crystallites in a polycrystalline diamond layer 80 μm thick; (b) enlarged part of the topogram in Fig. 1a. Laue diffraction of characteristic Co$K\alpha_1$ (outer ring) and Co$K\alpha_2$ (inner ring) radiation. Spots *1* and *5* without asterism; spots *2–4*, *6*, and *7* with asterism.

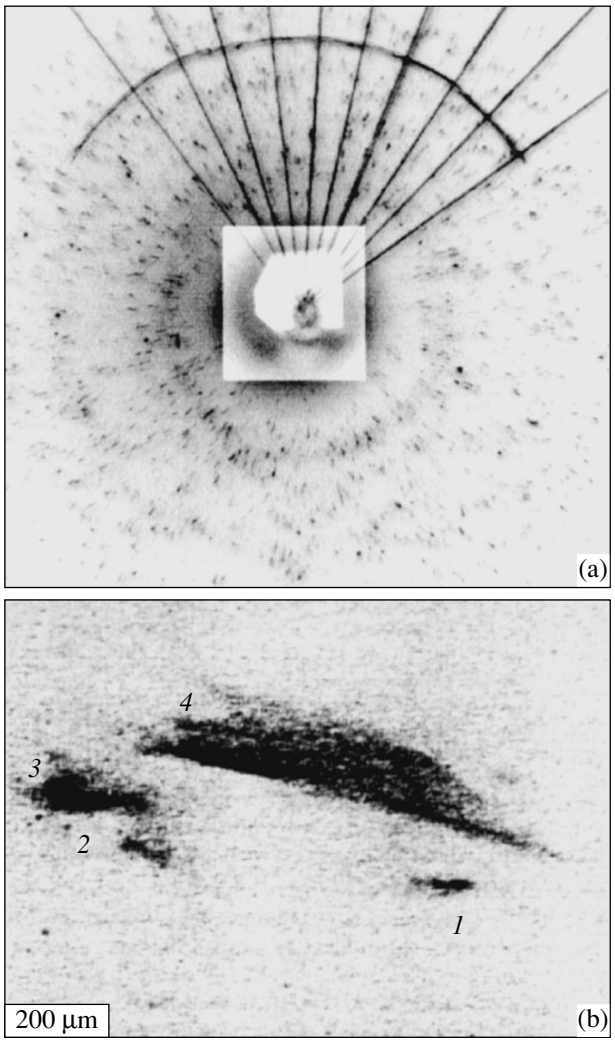


Fig. 2. (a) General view of the X-ray topogram from crystallites in the PDL 200 μm thick (Mo$K\alpha_{1,2}$ radiations). (b) Part of the topogram in Fig. 2a. Spots *1*, *3*, and *4* show asterism in ascending order.

tion. In such crystallites (marked *1* and *2* in Fig. 1b), only purely elastic strains and their associated elastic stresses may be observed. However, the topogram from the same sample exhibits a number of reflections where the Bragg size of spots exceeds their azimuthal size by a factor of 2–3 or more, i.e., demonstrates asterism (crystallites *3*, *4*, *6*, and *7* in Fig. 1b). Consequently, the crystallites responsible for "asterism tails" on the topograms have experienced plastic deformation. Thus, the XRT method for the first time visualized the fact that the same area of a relatively thin (h = 40–80 μm) PDL sample may simultaneously contain elastically and plastically strained crystallites. Such a result appeared to be unexpected for both the author and other researchers. It raised a fundamental question: How do local elastic stresses that are high enough to overcome the high-temperature ultimate strength of polycrystalline diamond and cause plastic deformation in some PDL crystallites, while being insufficient to plastically deform other (neighboring) crystallites, arise? Below, we will try to answer this question.

In PDL samples with thickness $h \geq 100$ μm, the Bragg sizes of many or even the vast majority of reflections are several times larger than their azimuthal sizes (see Fig. 2a and also enlarged reflections *1–4* in Fig. 2b). The pronounced asterism of the diffraction spots suggests that many, though not all, diamond grains are plastically strained [6, 11]. This serves as direct evidence of the fact that the pressure exerted by surrounding crystallites exceeded the high-temperature ultimate strength of diamond. These local elastic stresses cause plastic strains in some of the crystallites in samples of thickness 40–80 μm or in the vast majority of crystallites in PDLs of thickness $h \geq 100$ μm during growth. Since the topogram also contains fairly symmetric and even circular (according to the shape of the X-ray tube focus) reflections, one may conclude that the corresponding crystallites in the same polycrystalline diamond layer experience only some purely elastic deformation. Thus, even rather thick PDL samples may exhibit both plastically strained and plastically unstrained crystallites.

Subject to a number of constraints, the length of the diffraction tails in the Bragg direction may serve as an approximate measure of plastic strains in crystallites. On the topograms, the length of asterism spots in the Bragg direction varied in the 10–170 μm range, reaching 1.0–2.5 mm in exceptional cases.

The size of one diffraction point in the Bragg direction is about 3 μm for monochromatic Co$K\alpha_1$ radiation (≈1 μm for Mo$K\alpha_1$) and ≈7 μm for polychromatic radiation. Then, in units of wavelength, a reflected point in a diffraction spot measures (in the Bragg direction) $\delta\lambda_{K\alpha_1} \approx 5.4 \times 10^{-5}$ Å for monochromatic Mo$K\alpha_1$ and Co$K\alpha_1$ radiation and $\delta\lambda \approx 2.0 \times 10^{-4}$ Å for white or polychromatic radiation. For asterism tails on reflections formed by polychromatic radiation, this interval would be $\delta\lambda = (5.0–7.6) \times 10^{-4}$ Å or, in exceptional cases, $\delta\lambda_{max} \approx 0.11$ Å.

In the Laue transmission geometry for PDLs, the position of a diffraction reflection from an individual crystallite is determined by the following relationships (in the case of characteristic $K\alpha_1$ or $K\alpha_2$ radiation):

$$l_1 = B\tan 2\theta, \quad (1)$$

$$l_2 = B\tan(2\theta + 2\Delta\theta), \quad (2)$$

$$\Delta l = l_2 - l_1 = B\tan(2\theta + 2\Delta\theta) - B\tan 2\theta. \quad (3)$$

Here, l_1 and l_2 are the distances from the center of the incident beam to the end points of an asterism spot (the length of an asterism tail is $\Delta l = l_2 - l_1$); B is the distance between the PDL sample and photographic plate (usually B = 40 mm); and θ is the Bragg angle for diffraction from low-index {111} planes, which have a high

repetition factor in the cubic lattice of diamond. Diffraction reflections from these planes are one of the most intense for individual crystallites in PDLs. Since the length Δl of asterism tails is easy to measure from topograms, the amount of asterism may be found from Eq. (3) in angular units $2\Delta\theta$:

$$2\Delta\theta = \arctan\{[\Delta l/[B + B\tan^2 2\theta + \Delta l \tan 2\theta]]\}. \quad (4)$$

The amount of asterism is then used to calculate elastic stresses, which are related to plastic strains, in individual crystallites.

Obviously, formula (4) is applicable for $\theta < 45°$. Having measured the linear asterism Δl from the diffraction reflection for any of the crystallites in the topogram and having converted this value to its angular equivalent, one may numerically evaluate residual elastic stresses responsible for plastic strains in the given crystallite. Since the generation of dislocations is an elementary mechanism of shear plastic deformation, dislocation boundaries will divide a crystallite (like a tiny single crystal) into several mutually misoriented areas. Taking the actual amount of asterism measured in the topogram as the angle of maximal misorientation between all areas of a crystallite, we may assume in model calculations that an incoherent dislocation subboundary divides the crystallite into only two mutually misoriented areas. We will apply the approximate theoretical model formula for an incoherent dislocation boundary [12, 13]

$$\sigma \approx E b_{\parallel}/h, \quad (5)$$

where E = 1143 GPa is the Young's modulus of diamond [15, 16], $b_{\parallel}$ is the projection of the Burgers vector of dislocations onto the plastic shear boundary, and h is the crystallite thickness. For low misorientations between incoherent dislocation boundaries, $b_{\parallel} \approx b\Delta\theta$ and $h \approx b/\Delta\theta$. Then, (5) transforms into the convenient computational formula

$$\sigma \approx E(\Delta\theta)^2. \quad (6)$$

Thus, knowing the angular amount of asterism, one can approximately evaluate residual elastic strains, which are related to plastic strains, in each of the PDL crystallites on the topogram.

As follows from Figs. 1a and 1b (topograms taken of 40- to 80-μm-thick PDLs), relatively small (≤5 μm) elastically (but as yet not plastically) strained crystallites appear as small spots, and their sizes in the Bragg and azimuthal directions are practically the same. Diffraction spots elongate in the Bragg direction (with the formation of diffraction tails) when adjacent areas in the crystallite lattice irradiated by monochromatic Co$K\alpha_{1,2}$ or Mo$K\alpha_{1,2}$ radiation turn out to be misoriented relative to each other. It is precisely this situation that occurs in single crystals under shear plastic deformation [14]. Each grain in a polycrystal may be viewed as a tiny single crystal. Thus, the asterism observed in our paper seems to be associated primarily with misorientation between separate areas of a crystallite that experienced shear plastic deformation. This is consistent with the concept of asterism put forward in [11].

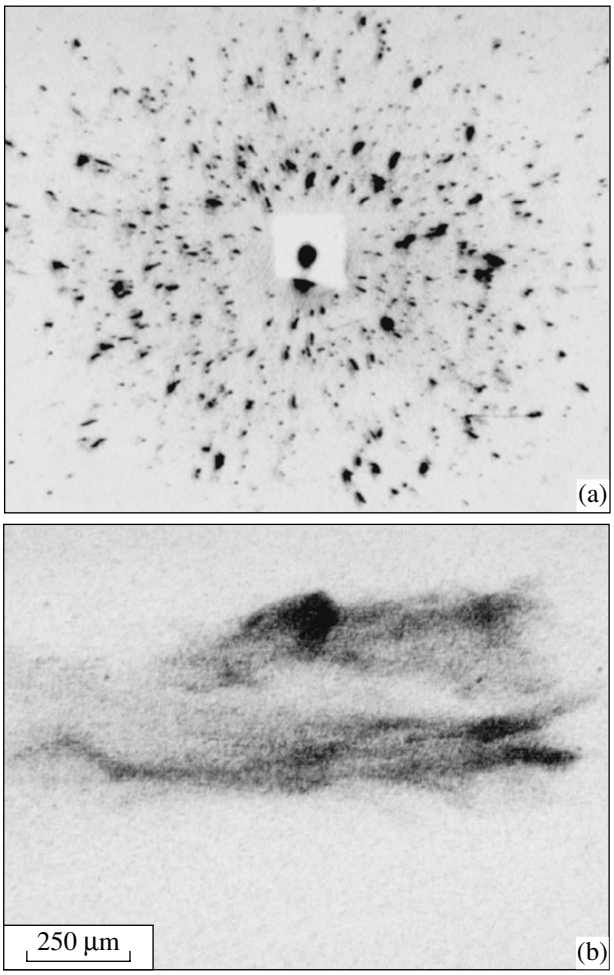


Fig. 3. (a) General view of the X-ray topogram from crystallites in the PDL 670 μm thick (Mo$K\alpha_{1,2}$ radiation). (b) Part of the topogram in Fig. 3a that shows the crystallite with a maximal asterism.

As follows from the above measurements, the length of asterism spots in the Bragg direction lies between 10 and 170 μm but may reach 1.0–2.5 mm (see the topograms taken from the 200- and 670-μm-thick PDLs; Figs. 2 and 3, respectively). Then, in the case of Mo$K\alpha_1$ radiation, the minimal asterism for plastically strained diamond crystallites that is estimated by (4) is $\Delta\theta_1 \approx 0.007069°$ or $\Delta\theta_1 \approx 0.42'$; the strong asterism for a set of crystallites ranges up to $\Delta\theta_2 \approx 0.107597°$ or 6.6′; and, finally, the maximal asterism observed in some of the crystallites is as high as $\Delta\theta_3 \approx 0.315622°$ (19′) or even $\Delta\theta_4 \approx 1.551969°$, as follows from the measurements and calculations.

Using approximate formula (6), one can calculate residual elastic stresses in the plastically strained crystallites for which misorientations caused by shear plastic deformation are known. From the misorientations

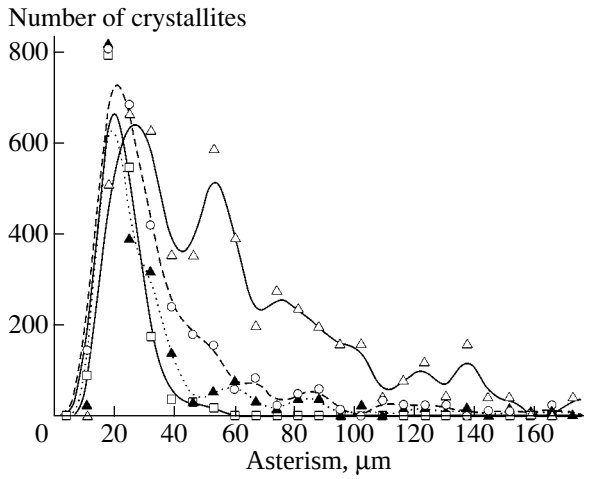


Fig. 4. Crystallite asterism distribution for four PDLs of thickness (□) 40, (○) 80, (△) 200, and (▲) 670 μm.

given above, we find that the elastic stress is $\sigma_1 \approx$ 17.4 kPa for the minimal plastic strain; $\sigma_2 \approx 4.0$ MPa or $\sigma_3 \approx 0.21$ GPa for large strains; and as high as $\sigma_4 \approx$ 0.84 GPa for the maximal plastic strain in a number of crystallites. Such are the residual elastic stresses in plastically strained crystallites of relatively thick ($h \geq$ 100 μm) PDL samples.

In relatively thin (h = 40–80 μm) PDLs, the asterism of the spots was in the range 10–300 μm, as measured from the topograms taken with Co$K\alpha_{1,2}$ radiation. According to (4), the minimal angular asterism of crystallites in 40- to 80-μm-thick PDLs is thus equal to $\Delta\theta_m \approx 0.002778°$, and the corresponding residual elastic stresses in these plastically strained crystallites were found to be $\sigma_m \approx 2.7$ kPa. Thus, during the growth of relatively thin (<100 μm) samples, a significant plastic strain arises only in a number of crystallites. As is evident from the topographic measurements and calculations, the elastic stresses vary from 2.7 to 17.4 kPa for crystallites with a low asterism, from 2.4 to 4.0 MPa for those with a high asterism, and from 0.7 to 0.84 GPa in crystallites where the asterism is maximal. The lower and upper limits of these ranges refer to the thin (40–80 μm) and thick (100–670 μm) layers, respectively.

The DBQPS data for the asterism of the images of crystallites in different PDL samples can be treated mathematically in the same manner as data for crystallite sizes [6]. The crystallite distributions with the amount of asterism (expressed in micrometers) are presented in Fig. 4 for four PDL samples. Most of the crystallites fall into the first (the highest) maximum, the positions of which are virtually the same for three of the four samples. The amount of asterism corresponding to this basic maximum is almost 20 μm. The data for the thin (40 and 80 μm) and thick (≥200 μm) PDL samples can be conveniently compared with the table, which lists all the measurements and calculations. In each of these plastically strained crystallites, the residual elastic stresses corresponding to the basic maximum are $\sigma_{m1} \approx 10.75$ kPa and $\sigma_{m2} \approx 17.4$ kPa for the thin and thick PDLs, respectively. In addition, three weaker maxima with an asterism of 60, 85, and 125 μm are present in Fig. 4. However, they appear only in the case of relatively thick ($h \geq 100$ μm) PDLs. The number of crystallites covered by these additional maxima is considerably less (1 to 10%) than in the basic maximum (Fig. 4).

The sharpest and highest maxima are observed for the fourth sample, where the position of the basic maximum in the case of Mo$K\alpha_{1,2}$ radiation shifts to 28 μm. The second asterism maximum in this sample is only slightly lower (80% of the basic maximum). In addition, the second maximum (Δl_2 = 55 μm), as well as the rather broad third (from Δl_{3bg} = 75 μm to Δl_{3en} = 105 μm) and fourth ones, are several times higher than the corresponding maxima for the rest of the samples. It should be added that the third maximum is so broad that its upper part appears as an inclined straight line extending from 75 to 105 μm of linear asterism with the height decreasing from 42 to 25% of the basic maximum. The fourth maximum is split into two submaxima at Δl_{4a} = 125 μm and Δl_{4b} = 135 μm. The last maximum lies at Δl_5 = 180 μm. It should be emphasized that the fourth and third samples have the same thickness, 200 μm. Moreover, both were grown at the Institute of General Physics, Russian Academy of Sciences, by the same method and with the same setup. However, according to our XRT data, the crystallite distributions over the amount of asterism in these samples differ dramatically. Therefore, one can argue that, during growth, the conditions for initiation and evolution of plastic deformation in crystallites of these equally thick PDLs were totally different. Conversely, for samples 1, 3, and 5, which were grown by two different methods and at different institutes (sample 1 to 40 μm thick was grown at the Institute of Physical Chemistry, Russian Academy of Sciences, by the hot filament method, while samples 3 (2 to 200 μm thick) and 5 (670 μm thick) were grown in a microwave plasma at the Institute of General Physics Institute, Russian Academy of Sciences), striking agreement is observed between the crystallite asterism distributions. The basic maxima in these samples are almost coincident. The distinctive feature of these PDLs is that the width of the basic maximum increases slightly with the layer thickness. It seems quite logical to expect this feature. However, the shift of the basic asterism maximum to 28 μm and the formation of four much higher extra maxima in the fourth sample 200 μm thick with increasing asterism lead us to think that the growth conditions for this sample were so changed that shear plastic deformation was initiated in a much greater number of crystallites. In the fifth sample, which was grown in the same setup but is much thicker (670 vs. 200 μm!), the same tendency toward the formation of extra maxima with a higher asterism is observed, but these maxima are one order of

Asterism of individual crystallites, as well as of groups of crystallites, in the maxima of the crystallite asterism distribution curves. The crystallites were plastically deformed during the growth of polycrystalline diamond layers. The asterism was measured by the diffraction reflections in the topograms. Also given are the calculated values of the residual elastic stresses σ_n in plastically strained crystallites

Sample no.	PDL thickness, μm	X-ray radiation	Amount of asterism Δl		Residual elastic stresses in separate groups of crystallites		
			μm	deg	kPa	MPa	GPa
1	40	Co$K\alpha_{1,2}$	10	0.002778	2.7	–	–
			20	0.005555	10.75	–	–
2	80	Co$K\alpha_{1,2}$	10	0.002778	2.7	–	–
			20	0.005555	10.75	–	–
			100	0.043573	661.0	–	–
			300	0.129810	–	5.87	–
3	200	Mo$K\alpha_{1,2}$	10	0.007069	17.4	–	–
			20	0.012673	56.0	–	–
			60	0.038008	–	0.50	–
			85	0.053835	–	1.0	–
			125	0.079145	–	2.2	–
			170	0.107597	–	4.03	–
4	200	Mo$K\alpha_{1,2}$	$\Delta l_{\text{lim}} = 10$	0.007069	$\sigma_{\min} = 17.4$	–	–
			$\Delta l_m = 28$	0.017742	$\sigma_m = 110$	–	–
			$\Delta l_2 = 55$	0.034843	$\sigma_2 = 422.7$	–	–
			$\Delta l_{3bg} = 75$	0.047505	$\sigma_{3bg} = 785.8$	–	–
			$\Delta l_{3en} = 105$	0.066492	–	$\sigma_{3en} = 1.5$	–
			$\Delta l_{4a} = 125$	0.079156	–	$\sigma_{4a} = 2.18$	–
			$\Delta l_{4b} = 135$	0.085469	–	$\sigma_{4b} = 2.54$	–
			$\Delta l_5 = 180$	0.113917	–	$\sigma_5 = 4.52$	–
5	670	Mo$K\alpha_{1,2}$	10 μm	0.007069	17.4	–	–
			20 μm	0.012674	56.0	–	–
			170 μm	0.107597	–	4.0	–
			1 mm	0.315622	–	–	0.21
			2.5 mm	1.551969	–	–	0.84

magnitude lower than those in the fourth sample (200 μm thick).

Thus, with the same PDL growth technology, the variation of the growth conditions even for not very thick layers results in different distributions of plastically strained crystallites over the amount of asterism and in different residual elastic stresses in them.

POSSIBLE REASONS FOR HIGH ELASTIC STRESSES IN GROWING POLYCRYSTALLINE DIAMOND LAYERS

The room-temperature ultimate strength of natural single-crystalline diamond varies from 9 to 190 GPa [15, 16] depending on the type of deformation (compression, tension, or shear), whereas for commercial diamonds, it is considerably lower, ranging from 0.23 to 0.48 GPa [16]. Calculations and measurements give approximately the same values of shear strength: 121 and 132 GPa, respectively [16].

The cubic lattice parameters of silicon, $a_1 = 5.4282$ Å, and diamond, $a_2 = 3.5676$ Å, differ substantially. The lattice mismatch for the (111)Si–diamond system is about 42%. When diamond layers grow on a single-crystalline silicon substrate, an intermediate layer of hexagonal silicon carbide (6H-SiC) is known to form, as confirmed by our XRD examination of the phase composition of the PDL/(001)Si system and by optical measurements [6]. In this system, the planar {111}C network of diamond is brought into contact with the (0001)SiC hexagonal plane. In this case, the mismatch decreases to 41.4%, i.e., becomes slightly less than for a diamond layer grown directly on the (111)Si substrate without the intermediate layer.

For our system, representations and formulas (1.5–1.9) from [17] have the form

$$\begin{aligned}\sigma_{xx} &= \sigma_{yy} = [E_2/(1-\nu_2)]\varepsilon_{xx} \\ &= [E_2/(1-\nu_2)]\{6f_{12}(h_1+h_2)/[(E_2h_2^3/E_1h_1) \\ &\quad + (E_1h_1^3/h_2) + 6h_1h_2 + 4h_1^2 + 4h_2^2]\}h_2 \\ &\quad - [f_{12}h_1 + (A/2)(E_2h_2^2/E_1 - h_1^2)]/(E_2h_2/E_1 + h_1),\end{aligned} \tag{7}$$

where

$$\begin{aligned}A &= [E_2/(1-\nu_2)]\{6f_{12}(h_1+h_2)/[(E_2h_2^3/E_1h_1) \\ &\quad + (E_1h_1^3/h_2) + 6h_1h_2 + 4h_1^2 + 4h_2^2]\}h_2.\end{aligned} \tag{8}$$

Using these relationships, we can calculate tensile stresses in the PDL under conditions of isostructural heteroepitaxy. The calculation also takes into account the elastic relaxation of the system due to bending stresses and the great difference in the Young's moduli for silicon, E_1 = 176.58 GPa, and polycrystalline diamond clusters, E_2 = 1143 GPa [15, 16], as well as in the Poisson's ratios: ν_1 = 0.215 and ν_2 = 0.0691, respectively. Then, for a substrate thickness h_1 = 3 mm and a PDL thickness h_2 = 600 μm, we find that the tensile stress in the polycrystalline diamond layer is $\sigma_{(xx)2} \approx \sigma_{(yy)2}$ = 274.7 GPa, which exceeds the tensile strength σ = 190 GPa of natural single-crystalline diamond at room temperature.

Thus, in the case of isostructural heteroepitaxy, stresses caused by the lattice mismatch would exceed the ultimate strength of natural single-crystalline diamond, to say nothing of synthetic PDLs, for which the ultimate strength is much lower. The internal stresses arising in thick PDLs may hardly be related to only the large diamond–substrate lattice mismatch. At the initial stage of growth, when diamond nuclei grow on the silicon substrate, a kind of isostructural heteroepitaxy through the 6H-SiC intermediate layer may take place; however, the above calculations give the result for the heteroepitaxial layer–substrate system as a whole.

Our XRT data clearly show that, during growth, local elastic stresses in some of the crystallites start to exceed the high-temperature ultimate strength of diamond, causing these crystallites to experience plastic (specifically, shear) deformation.

The basic physical reason for the plastic deformation of individual crystallites in PDLs is likely to be intercrystallite pressure in the PDL, which appears when the nuclei, reaching the size of the crystallites, come into contact with each other. Since the crystallites are randomly oriented, they may apply pressure to each other through both crystallographic planes and sharp edges. In other words, a growing PDL plastically deforms itself. Intergranular spacers are also a factor in this process. As is well known, they may trap lattice imperfections, such as point and linear defects, as well as nondiamond phases of carbon.

The elementary act of plastic deformation in single crystals is the generation of a dislocation loop or semiloop that eventually emerges on the free surface of the single crystal. In our case, the semiloop emerges on the surface of a crystallite; i.e., it comes out into the intergranular space. In particular, dislocations generated in the single-crystalline body of a crystallite and emerging on its surface may form a grain boundary. However, neither ordered planar dislocation clusters nor segments of curvilinear dislocations have been detected in diamond single crystals and PDLs. Otherwise, they might be used to evaluate local elastic stresses acting in real single crystals or individual PDL crystallites during the generation and pinning of dislocations in the crystal. Similar XRT measurements and calculations were carried out by the author for III–V semiconductors [18].

Texture. The relative intensities of diffraction reflections obtained with the double-crystal spectrometer–diffractometer by the pseudorocking curve technique [6, 8, 19] show that the (111) reflections from the lower part of the polycrystalline diamond, where the texture only starts to evolve, have a maximal intensity (as is the case for diamond powders listed in the ASTM index). In the upper part of each of the layers (even in the 80-μm-thick samples), the most intense reflections come from the (331)-oriented crystallites. In the 670-μm-thick samples, the reflections from the (110)-oriented crystallites, which run parallel to the growing surface, are the most intense; i.e., the texture forms during the growth process. The (110) texture is characteristic of optical-grade films synthesized in a microwave discharge [6, 20].

CONCLUSIONS

(1) Based on the DBQPS technique and the well-known phenomenon of asterism, we developed an original XRT method for the identification and measurement of plastic strains in individual PDL grains exceeding 3 μm in size. It enables one to measure and calculate residual elastic stresses in plastically strained crystallites, as well as to separate out reflections from purely elastically strained crystallites.

(2) The amount of asterism is used as a quantitative measure of plastic strains in specific crystallites of polycrystals. A mathematical substantiation of the quantitative XRT analysis is given.

(3) From XRT data, the crystallite asterism distribution is constructed for PDLs of thickness ranging from 40 to 670 μm. Based on the effect of asterism in X-ray diffraction, shear plastic strains in crystallites of diamond layers ≥40 μm thick are found for the first time. The strains cause a misorientation between the crystallites in the range from 0.4′ to 1.5°. The corresponding residual elastic stresses in plastically strained crystallites were calculated to be in the range from 2.7 kPa to 0.84 GPa. The latter value exceeds the ultimate strength

of artificial polycrystalline diamonds (crystallites rather than intergranular spacers).

(4) This XRT method may be used to examine crystallites in thin layers of any polycrystalline metals or other materials.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (Sukhov's grant no. 03-03-32396).

REFERENCES

1. A. R. Lang, in *Collection of Articles*, Ed. by A. M. Elistratov (Mir, Moscow, 1965).
2. Krishan Lal, S. N. Goswami, and A. R. Verma, Solid State Commun. **81**, 461 (1992).
3. Krishan Lal, Proc. Ind. Nat. Sci. Acad. A **64**, 609 (1998).
4. C. K. Suzuki, A. H. Shinihara, P. H. Godoy, *et al.*, Diamond Relat. Mater. **7**, 289 (1998).
5. D. Rats, L. Binbault, L. Vandenbulcke, *et al.*, J. Appl. Phys. **78**, 4994 (1995).
6. G. F. Kuznetsov, V. G. Ral'chenko, V. P. Varnin, *et al.*, Kristallografiya **47**, 333 (2002) [Crystallogr. Rep. **47**, 298 (2002)].
7. G. F. Kuznetsov, Doctoral Dissertation (Moscow Inst. of Radio Engineering and Electronics, USSR Acad. Sci., 1989).
8. G. F. Kuznetsov, Kristallografiya **21**, 847 (1976) [Sov. Phys. Crystallogr. **21**, 485 (1976)].
9. G. F. Kuznetsov and S. A. Semiletov, Nondestructive Testing of Epitaxial and Polycrystalline Films in *Microelectronics by X-ray Diffraction Methods* (TsNII Élektronika, Moscow, 1975), Ser. Microelectronics, No. 1.
10. G. F. Kuznetsov, in *Proceedings of the 9th International Workshop on Physics of Semiconductor Devices, New Delhi, 1997*.
11. A. Guinier, *Theorie et Technique de la Radiocristallographie* (Dunod, Paris, 1956; Fizmatgiz, Moscow, 1961); see also *X-ray Diffraction in Crystals, Imperfect Crystals, and Amorphous Bodies* (Freeman, San Francisco, 1963; reprinted by Dover, New York, 1994).
12. B. K. Vaĭnshteĭn, V. M. Fridkin, and V. L. Indenbom, *Modern Crystallography* (Nauka, Moscow, 1979; Springer-Verlag, Heidelberg, 1982), Vol. 2.
13. I. Vlasov, V. Ralchenko, D. Zakharov, *et al.*, Phys. Status Solidi A **174**, 11 (1999).
14. G. S. Zhdanov, *Physics of Solid State* (Mosk. Gos. Univ., Moscow, 1961).
15. M. Werner, S. Klose, F. Szucs, *et al.*, Diamond Relat. Mater. **6**, 344 (1997).
16. *Natural Diamonds of Russia* (Polyaron, Moscow, 1997).
17. Yu. A. Tkhorik and L. S. Khazan, *Plastic Strain and Misfit Dislocations in Heteroepitaxial Systems* (Naukova Dumka, Kiev, 1983).
18. G. F. Kuznetsov, Kristallografiya **46**, 320 (2001) [Crystallogr. Rep. **46**, 276 (2001)].
19. G. F. Kuznetsov, Kristallografiya **45**, 326 (2000) [Crystallogr. Rep. **45**, 294 (2000)].
20. I. I. Vlasov, V. G. Ralchenko, E. D. Obraztsova, *et al.*, Appl. Phys. Lett. **71**, 1789 (1997).

Translated by M. Lebedev

Technical Physics, Vol. 48, No. 12, 2003, pp. 1554–1564. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 54–65.

RADIOPHYSICS

Effect of Diffraction on the Electrodynamic Characteristics of Two-Dimensional Coaxial Bragg Resonators

N. S. Ginzburg, N. Yu. Peskov, and A. S. Sergeev

Institute of Applied Physics, Russian Academy of Sciences, ul. Ul'yanova 46, Nizhni Novgorod, 603950 Russia

Received May 20, 2003

Abstract—The electrodynamic properties of coaxial two-dimensional Bragg resonators with two-dimensional distributed feedback are analyzed. These resonators are made of coaxial waveguide sections with doubly periodic corrugation, which provides coupling and mutual scattering of four partial waves. Two of them propagate along the waveguide, while the other two propagate in the transverse (azimuthal) direction. It is shown that the high azimuthal index selectivity of two-dimensional Bragg resonators may be related to a qualitative difference in topology of the dispersion characteristics of azimuth-symmetric and asymmetric normal waves propagating in infinite waveguides of such a geometry. For the finite-length systems used as two-dimensional Bragg resonators, the eigenmode spectrum is found for two types of boundary conditions that correspond to the limiting cases of perfectly matched (open) systems and, conversely, of systems closed for the extraction of transverse electromagnetic fluxes. Perimeter-to-length ratios of the resonator at which the Q factor of the fundamental azimuth-symmetric mode is greater than those of the other modes are determined. The applicability domain of the geometrical approach, which was earlier applied to two-dimensional Bragg resonators, is discussed. © *2003 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

Distributed two-dimensional feedback as a means for providing the spatial coherence of radiation from hollow and ribbon-shaped electron beams with a diameter much exceeding the wavelength was suggested in [1–3]. This mechanism can be implemented with planar or coaxial two-dimensional Bragg resonators. Millimeter-wave free-electron masers (FEMs) using the new feedback mechanism are being experimentally studied today [4, 5]. Therefore, a detailed examination of the electrodynamic characteristics of two-dimensional Bragg resonators seems to be topical. In previous works (except for [6]), the geometrical optics approximation, ignoring diffraction effects, was used. This approximation made it possible to find the eigenmode spectrum of two-dimensional Bragg resonators and demonstrated a high transverse (azimuthal) index selectivity of the devices. Also, a nonlinear nonstationary theory of FEMs that was developed within the framework of geometrical optics [3, 6–8] for the most part corroborated the conclusion of the linear theory that two-dimensional Bragg structures may provide the spatial coherence of radiation from fully developed relativistic electron beams. At the same time, it is clear that a more thorough quasi-optical analysis may not only determine the applicability domain of the geometrical approach but also reveal additional features of two-dimensional Bragg structures.

This paper extends the analysis to coaxial Bragg resonators. Section 1 describes a basic model, which relies on the method of coupled waves for four partial electromagnetic energy fluxes that propagate in the longitudinal and transverse (azimuthal) directions and experience mutual scattering on the two-dimensional Bragg structure. Basic equations that allow for diffraction effects are also presented in Sect. 1. In Sect. 2, we consider the dispersion characteristics of normal waves in infinitely long two-dimensional coaxial Bragg structures. The behavior of the dispersion curves for a symmetric normal wave near the Bragg frequency is shown to be qualitatively distinct from the behavior for asymmetric waves (hereafter, the waves are classified in terms of the azimuthal structure of partial waves that propagate in the longitudinal direction). This distinction is, in fact, the reason for the selectivity of two-dimensional Bragg resonators. In Section 3, we consider finite-length systems. The eigenmode spectrum is found for two limiting cases: perfectly matched (open) systems and systems closed for the extraction of transverse electromagnetic fluxes (subsequently referred to as closed systems for brevity). It is shown that even closed systems offer a high azimuthal index selectivity when the resonator's perimeter is one order of magnitude greater than the wavelength. Precise matching between the output channels for transverse energy fluxes is expected to improve the selectivity and provide the possibility of further increasing the resonator perimeter.

1. MODEL AND BASIC EQUATIONS

Let an electrodynamic system (Fig. 1a) represent a coaxial corrugated waveguide with a mean diameter r_0. The corrugation is made as a superposition of two

1063-7842/03/4812-1554$24.00

oppositely wound helices (doubly periodic corrugation):

$$a = a_1 \cos(\bar{h}_z z)\cos(\overline{M}\varphi) = \frac{a_1}{2}[\cos(\bar{h}_z z - \overline{M}\varphi) + \cos(\bar{h}_z z + \overline{M}\varphi)], \quad (1)$$

where $2a_1$ is the corrugation depth; $\bar{h}_z = 2\pi/d_z$; d_z is the corrugation period along the z axis; $\overline{M}$ is the number of corrugation starts along the circumference; and z and φ are the longitudinal and azimuthal coordinates, respectively.

We also assume that the curvature of the waveguide is small; i.e., its radius far exceeds the wavelength λ and spacing (gap) a_0 between the conductors:

$$r_0 \gg \lambda, \quad r_0 \gg a_0. \quad (2)$$

Under these assumptions, the dispersion relation for the eigenmodes of the coaxial waveguide can be reduced to the form [9]

$$\frac{\omega^2}{c^2} = h_z^2 + h_x^2 + \kappa_p^2, \quad (3)$$

which is similar to the dispersion relation for a planar waveguide. Here, h_z is the longitudinal wavenumber; $h_x = M/r_0$ is the azimuthal wavenumber; $\kappa_p = p\pi/a_0$ is the radial wavenumber; and M and p are the numbers of wave variations in the azimuthal and radial directions, respectively.

If the corrugation depth is small, $\bar{h}_{z,x} a_1 \ll 1$, the field in the system can be represented as a superposition of four coupled wave flows, two of which, $A_\pm$, propagate in the longitudinal $\pm z$ direction and the other two, $B_\pm$, propagate azimuthally:

$$\mathbf{E} = \mathrm{Re}[(A_+\mathbf{E}_1^0 e^{ih_1 z} + A_-\mathbf{E}_1^0 e^{-ih_1 z} + B_+\mathbf{E}_2^0 e^{ih_2 x} + B_-\mathbf{E}_2^0 e^{-ih_2 x})e^{i\omega t}], \quad (4)$$

where $x = r_0\varphi$ is the azimuthal position; $A_\pm(x, z)$ and $B_\pm(x, z)$ are the wave flow amplitudes, which are slowly varying functions of the longitudinal and azimuthal coordinates; and $\mathbf{E}_{1,2}^0(r)$ are functions describing the radial structure of the wave flows, which coincides with the mode structure of a coaxial waveguide.

On Bragg lattice (1), partial wave flows (4) couple and mutually scatter if the geometrical parameters $\bar{h}_{x,z}$ of the lattice are such that the propagation constants $h_{1,2}$ satisfy the Bragg resonance condition (Fig. 1b)

$$h_1 \approx \bar{h}_z, \quad (5a)$$

$$h_2 \approx \bar{h}_x, \quad (5b)$$

where $\bar{h}_x = \overline{M}/r_0$.

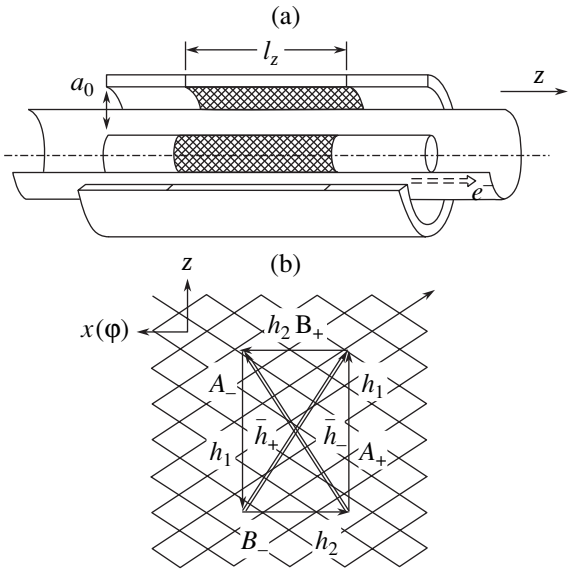


Fig. 1. (a) Schematic of a coaxial two-dimensional Bragg waveguide with a hollow electron beam moving in the +z direction and (b) the diagram illustrating the scattering of the partial waves by the two-dimensional Bragg lattice, where $\bar{\mathbf{h}}_\pm = \bar{\mathbf{h}}_x\mathbf{x}^0 \pm \bar{\mathbf{h}}_z\mathbf{z}^0$ are the lattice vectors.

We will assume that

$$\bar{h}_x = \bar{h}_z = \bar{h}, \quad (6)$$

which corresponds to the scattering of wave flows that have the same number p of variations along the azimuth. In this case, the geometrical parameters of the system are related as $\overline{M}/r_0 = 2\pi/d_z$. For simplicity, we will restrict our consideration to the scattering of lower order waves whose field does not vary along the radius; i.e., we set $p = 0$. In this case, $TE_{M,0}$ waves $A_\pm$ with small azimuthal indices (including the lowest order TEM wave with $M = 0$), which propagate along the waveguide, couple with $TE_{M,0}$ waves $B_\pm$ with large azimuthal indices $M \gg 1$, which propagate in the transverse (azimuthal) direction. Note that, when conditions (2) are satisfied, the structure of the $TE_{M,0}$ waves is similar to that of TEM waves (i.e., dispersion relation (3) applies).

Under conditions (2) and (6), the mutual scattering of four electromagnetic flows can be described by the equations of coupled waves (cf. [3, 6])

$$\pm\frac{\partial A_\pm}{\partial z} + i\delta A_\pm + i\alpha(B_+ + B_-) = 0,$$
$$\frac{i}{2\bar{h}}\frac{\partial^2 B_\pm}{\partial z^2} \pm \frac{\partial B_\pm}{\partial x} + i\delta B_\pm + i\alpha(A_+ + A_-) = 0. \quad (7)$$

Here, $\delta = (\omega - \bar{\omega})/c$ is the frequency offset from the exact Bragg resonance frequency $\bar{\omega} = \bar{h}c$ and α is the

wave coupling coefficient, which for $TE_{M,0}$ waves equals

$$\alpha = \frac{a_1 \bar{h}}{2a_0}, \tag{8}$$

when the corrugations on both conductors are in phase.

Note that the equation for partial flows $B_\pm$, which propagate in the transverse $\pm x$ ($\pm\varphi$) direction, is parabolic and allows for diffraction effects (it is clear that these waves are trapped in a finite-length system if diffraction spreading is disregarded). At the same time, these effects are neglected for the $A_\pm$ waves, which propagate in the longitudinal $\pm z$ direction, because they do not qualitatively change the characteristics of the system.

Since the system is coaxial, all the partial wave flows must meet the cyclicity condition

$$\begin{aligned} B_\pm(x + l_x, z) &= B_\pm(x, z), \\ A_\pm(x + l_x, z) &= A_\pm(x, z), \end{aligned} \tag{9}$$

where $l_x = 2\pi r_0$ is the perimeter of the system.

Due to the cyclicity condition, we can expand the fields into the Fourier series

$$\begin{aligned} A_\pm(x, z) &= \sum_{m=-\infty}^{\infty} A_\pm^m(z) e^{2\pi i m x / l_x}, \\ B_\pm(x, z) &= \sum_{m=-\infty}^{\infty} B_\pm^m(z) e^{2\pi i m x / l_x} \end{aligned} \tag{10}$$

and treat each of the harmonics as a normal wave characterized by the index m. Substituting expansions (10) into Eqs. (7) for the amplitudes of harmonics, we obtain the system of ordinary differential equations

$$\begin{aligned} &\pm \frac{dA_\pm^m}{dz} + i\delta A_\pm^m + i\alpha(B_+^m + B_-^m) = 0, \\ &\frac{1}{2\bar{h}} \frac{d^2 B_\pm^m}{dz^2} + \left(\delta \pm \frac{2\pi m}{l_x}\right) B_\pm^m + \alpha(A_+^m + A_-^m) = 0. \end{aligned} \tag{11}$$

According to relationships (4) and (5) for the partial $B_\pm$ waves, which propagate in the azimuthal direction, the index m is a complement to the azimuthal index $\overline{M}$. Thus, considering the usual set of waves in a coaxial waveguide where the highest-Q mode has $m = 0$ (which subsequently will be referred to as the symmetric mode), we may say that only the $A_\pm$ waves are actually symmetric, whereas the partial $B_\pm$ waves have a nonzero azimuthal index $\overline{M}$. In the general case, for normal waves in an infinitely long system (Sect. 2) or for the eigenmodes of a finite-length resonator (Sect. 3), a normal wave (eigenmode) with an index m is a superposition of coupled partial $A_\pm$ waves (with an index m) and $B_\pm$ waves (with an index $\overline{M} \pm m$). The applicability condition for Eq. (10) is $\overline{M} \gg m$.

2. DISPERSION CHARACTERISTICS OF NORMAL WAVES IN DOUBLY PERIODIC CORRUGATED COAXIAL WAVEGUIDES

Assuming that a coaxial structure with doubly periodic corrugations on its walls is infinitely long and representing a solution to Eqs. (11) as $A_\pm^m = a_\pm^m e^{i\Gamma z}$ and $B_\pm^m = b_\pm^m e^{i\Gamma z}$, we arrive at the dispersion relation for the normal waves

$$\begin{aligned} &\left(2\bar{h}\delta - 2\bar{h}\frac{2\pi m}{l_x} - \Gamma^2\right)\left(2\bar{h}\delta + 2\bar{h}\frac{2\pi m}{l_x} - \Gamma^2\right) \\ &\times (\delta^2 - \Gamma^2) = 8\alpha^2 \bar{h}\delta(2\bar{h}\delta - \Gamma^2). \end{aligned} \tag{12}$$

It is clear that, in the general case, dispersion relation (12) describes coupling between four partial waves. When the coupling coefficient is $\alpha = 0$, the relation splits into four equations that describe two longitudinal waves $A_\pm$ whose dispersion characteristics are determined by the relationships

$$\delta = \pm\Gamma \tag{13a}$$

and two quasi-critical modes $B_\pm$ that propagate in the azimuthal direction with the dispersion law

$$2\bar{h}\delta = \pm 2\bar{h}\frac{2\pi m}{l_x} + \Gamma^2. \tag{13b}$$

Note that, under the above assumptions, dispersion characteristics (13) for the partial waves can be obtained directly from initial dispersion relation (3). Near the Bragg frequency, $\omega = \bar{\omega} + c\delta = c(\bar{h} + \delta)$, where $\delta \ll \bar{h}$. Then, in view of the condition $\overline{M} \gg m$ and Bragg condition (5b), Eq. (3) for the partial waves $B_\pm$, propagating in the transverse direction, yields

$$\frac{\bar{\omega}^2}{c^2} + 2\frac{\bar{\omega}}{c}\delta = h_z^2 + \frac{(\overline{M} \pm m)^2}{r_0^2} \approx h_z^2 + \bar{h}^2 \pm \frac{2\bar{h}m}{r_0}.$$

Assuming that $h_z = \Gamma$, we arrive at dispersion relation (13b). Similarly, for the partial waves $A_\pm$, which propagate in the longitudinal direction, we substitute the wavenumber $h_z = \pm h + \Gamma$ into dispersion relation (3) and take into account (5a). Neglecting the terms of the higher order of smallness, we obtain

$$\frac{\bar{\omega}^2}{c^2} + 2\frac{\bar{\omega}}{c}\delta = (\pm h + \Gamma)^2 + \frac{m^2}{r_0^2} \approx \bar{h}^2 \pm 2\bar{h}\Gamma,$$

which is equivalent to dispersion relation (13a).

Figure 2 shows dispersion diagrams for the normal waves for $m = 0$ and $m \neq 0$. Comparing these curves with the asymptotes given by Eqs. (13) and shown by thin lines, we can conclude that branches *1* and *2* refer

to the partial $A_\pm$ waves, while branches *3* and *4* refer to the $B_\pm$ waves. On the whole, the basic difference between these dispersion diagrams and the corresponding characteristics of one-dimensional Bragg structures is the presence of a transmission band near the exact Bragg frequency $\bar{\omega}$ (i.e., in the vicinity of $\delta = 0$). This feature is due to the fact that, under Bragg resonance conditions (5), the scattering process involves quasi-critical $B_\pm$ waves whose unperturbed dispersion characteristics (13b) pass near the Bragg frequency.

Of still greater importance is the qualitative difference in the behavior of the dispersion characteristics of the normal waves with zero and nonzero azimuthal indices near the Bragg frequency (i.e., in the vicinity of $\delta = 0$). At $m = 0$, unperturbed dispersion characteristics (13b) of the partial waves merge together. When coupling is taken into account ($\alpha \neq 0$), dispersion relation (12) reduces to

$$\begin{aligned}(2\bar{h}\delta - \Gamma^2)(\delta^2 - \Gamma^2) &= 8\alpha^2\bar{h}\delta,\\ 2\bar{h}\delta &= \Gamma^2.\end{aligned} \qquad (14)$$

Equation (14) implies that one of the branches does not obey the conventional parabolic dependence of the frequency on the wavenumber squared near the Bragg frequency (such behavior is typical near the stop band boundaries, for example, in one-dimensional Bragg structures [10–12]). Instead, it exhibits a fourth-degree dependence (cf. [6]):

$$\Gamma^4 = 8\alpha^2\bar{h}\delta. \qquad (15)$$

Thus, near the Bragg frequency ($\delta = 0$), both the wave's group velocity and its derivative vanish. The significant difference of the dispersion curves for the symmetric normal waves from those for the partial waves near the Bragg frequency is illustrated in Fig. 3. Here, the dispersion characteristics are shown on an enlarged scale compared with Fig. 2a, i.e., on a scale comparable to the mode separation (eigenmodes with a different number of longitudinal variations are shown by closed circles). As is seen from Fig. 3, the fundamental mode has an extremely low group velocity of the normal wave ($m = 0$, $n = 1$). As is shown in Sect. 2, this is the specific feature of the given mode, because of which its Q factor is much higher than those of other symmetric and asymmetric modes.

Let us pursue our analysis in terms of geometrical optics. Formally, this means passage to the limit $\lambda \longrightarrow 0$ (i.e., $\bar{h} \longrightarrow \infty$ in view of (5)) in Eqs. (11). In this case, the term $\sim d^2B_\pm^m/dz^2$, which describes the diffraction spreading of the $B_\pm$ waves, may be neglected in Eqs. (11). Accordingly, dispersion relation (12) reduces to

$$\left(\delta - \frac{2\pi m}{l_x}\right)\left(\delta + \frac{2\pi m}{l_x}\right)(\delta^2 - \Gamma^2) = 4\alpha^2\delta^2. \qquad (16)$$

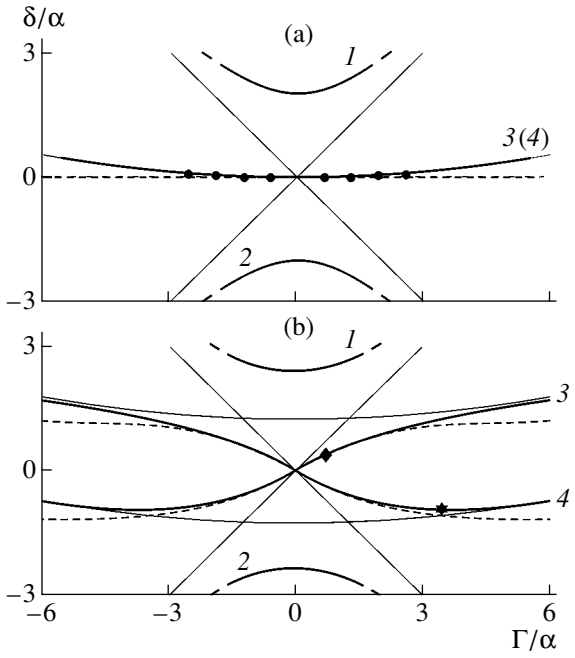


Fig. 2. Dispersion diagrams for normal waves in an infinitely long coaxial Bragg structure at $\bar{\mathbf{h}}/\alpha = 35$ for (a) symmetric waves $m = 0$ and (b) asymmetric waves with $m = \pm 1$ (classification in terms of the structure of longitudinal partial waves). Thin lines, dispersion curves for the partial waves; dashed lines, dispersion curves for the normal waves within the geometrical optics approximation. The circles refer to the positions of (a) symmetric modes near the Bragg frequency and (b) asymmetric modes shown in Fig. 9 ($\alpha l_x = \alpha l_z = 5$).

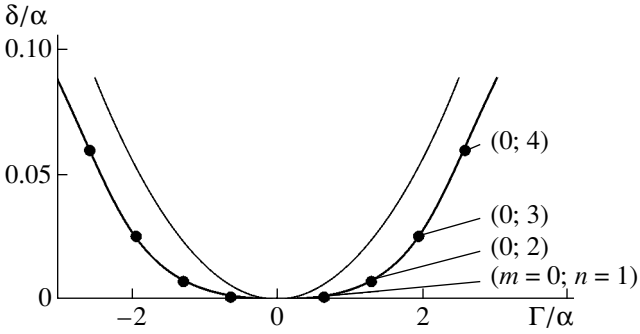


Fig. 3. Difference in dispersion laws for the symmetric normal wave (thick line) and for the partial quasi-critical wave (thin line) near the Bragg frequency at $\bar{\mathbf{h}}/\alpha = 35$. The circles indicate the position of the symmetric modes of the resonator at $\alpha l_z = 5$.

Dispersion characteristics described by Eq. (16) are shown in Fig. 2 by dashed lines. For small wavenumbers $\Gamma \leq \alpha$, these characteristics are fitted well by the dispersion curves obtained within the quasi-optical approximation, i.e., by Eqs. (12). Even the geometrical approximation reveals the qualitative difference in the behavior of the dispersion curves for the normal sym-

metric ($m = 0$) and asymmetric ($m \neq 0$) waves near the Bragg frequency. Ultimately, this difference is responsible for the azimuthal index selectivity of the system. At the same time, curves *3* and *4* at $m = 0$ transform into the straight line $\delta = 0$. As a result, the geometrical approximation encounters problems in describing the spectrum of the resonator's symmetric modes with a different longitudinal index, because these modes are degenerate in eigenfrequency (the eigenfrequencies of all the modes equal to the Bragg frequency). In addition, these modes have no diffraction losses; i.e., formally they have an infinitely high Q factor. However, as was first shown in [6], allowance for diffraction effects removes this degeneracy, transforming the corresponding dispersion curve into that described by Eq. (15).

3. MODE SELECTION IN TWO-DIMENSIONAL COAXIAL BRAGG RESONATORS

To calculate the eigenmode spectrum of a two-dimensional Bragg resonator in the form of a section of a coaxial waveguide with doubly periodic corrugation over a region of finite length l_z, it is necessary to set edge conditions for the system, i.e., at $z = 0$ and $z = l_z$. Consider two limiting cases: (a) a perfectly matched (open) system and (b) a closed system. In case (a), we assume that the partial modes propagating both longitudinally ($A_\pm$) and transversely ($B_\pm$) are perfectly matched at the ends of the system. In a desired frequency range, this can, in general, be achieved by specially profiling the waveguide's cross section and selecting the corrugation parameters. We also assume that external electromagnetic energy flows associated with either the partial $A_\pm$ waves or the partial $B_\pm$ waves are absent. Under these conditions, for the partial $A_\pm$ waves, we have, as in the case of one-dimensional Bragg resonators (cf. [10–12]),

$$A_+(x, z = 0) = 0, \quad A_-(x, z = l_z) = 0. \tag{17}$$

Essentially, conditions (17) mean that the longitudinal electromagnetic energy flows leave the system without reflecting from its ends $z = 0$ and l_z (however, the distributed scattering of these waves into $B_\pm$ waves, which is described by Eqs. (10), still exists inside the resonator). Assuming that a similar situation also occurs for the transversely propagating $B_\pm$ waves, one should set the so-called no-reflection condition at the corrugation boundaries (cf. [13])

$$\left[\frac{dB_\pm^m}{dz} - i\sqrt{2\bar{h}\left(\pm\frac{2\pi m}{l_x} + \delta\right)}B_\pm^m\right]\Bigg|_{z=0} = 0,$$
$$\left[\frac{dB_\pm^m}{dz} + i\sqrt{2\bar{h}\left(\pm\frac{2\pi m}{l_x} + \delta\right)}B_\pm^m\right]\Bigg|_{z=l_z} = 0. \tag{18}$$

Provided that this condition is satisfied and the waves are uncoupled ($\alpha = 0$), the fields $B_\pm$ freely diffract through the boundaries, as in the case of a regular waveguide.

However, the electrodynamic system at the ends of the resonator (at the corrugation boundaries) is irregular, which may cause reflections primarily for the quasi-critical $B_\pm$ waves. Therefore, it is of interest to study a completely closed resonator as the opposite limiting case (case (b)). This case can be realized by placing below-cutoff (for the quasi-critical $B_\pm$ waves) tapers at the ends of the corrugated region. If these waves are totally reflected from the resonator's ends, the boundary conditions are

$$B_\pm(x, z = 0) = 0, \quad B_\pm(x, z = l_z) = 0. \tag{19}$$

We also assume that the $A_\pm$ waves, which propagate along the structure, do not "sense" the irregularity of the waveguide's cross section and so do not reflect from the corrugation boundary. Thus, boundary conditions (17) for these waves remain valid.

(i) Matched (open) system. Applying boundary conditions (17) and (18) to dispersion relation (14) for the azimuth-symmetric modes ($m = 0$), we arrive at the characteristic equation given in the Appendix. For strongly coupled ($\alpha l_{x,z} \gg 1$) highest-Q azimuth-symmetric modes at a frequency close to the Bragg frequency ($\delta \ll \alpha$), the complex eigenfrequencies in view of specific dispersion law (15) are given by [6]

$$\delta_n \frac{\pi^4 n^4}{8\bar{h}\alpha^2 l_z^4} + i\frac{\pi^4 n^4}{4\bar{h}\alpha^3 l_z^5}, \tag{20}$$

where $n = 1, 2, 3, \ldots$ is the longitudinal mode index.

These modes are indicated by closed circles in the dispersion curve shown in Fig. 3. Their eigenfrequencies $\omega_{m,n}$ and quality factors $Q_{m,n}$ are given by

$$\omega_{m,n} = \bar{\omega} + c\,\mathrm{Re}(\delta_{m,n}),$$
$$Q_{m,n} = \frac{\bar{\omega}}{2c\,\mathrm{Im}(\delta_{m,n})}. \tag{21}$$

As follows from relationships (20) and (21), the mode with one longitudinal field variation ($n = 1$) has the highest Q factor (i.e., the lowest diffraction losses). As the longitudinal mode number increases, the diffraction losses grows as $\sim n^4$. Note that, with the same geometry and coupling coefficient, the loss factor for a conventional Bragg resonator in the form of a section of a corrugated coaxial waveguide is expressed as [10–12]

$$\mathrm{Im}(\delta_n) \approx \frac{\pi^2 n^2}{\alpha^2 l_z^3}. \tag{22}$$

Comparing relationships (20) and (22) shows that, with the same coupling coefficient α, a two-dimensional Bragg resonator has a much higher Q factor and a much higher selectivity (including the longitudinal index (n) selectivity). This is a consequence of dispersion law (15) mentioned above. For the two-dimensional resonator, the longitudinal partial-wave distribution at the fundamental mode $n = 1$ for $\alpha l_x = \alpha l_z = 5$ is shown in Fig. 4a.

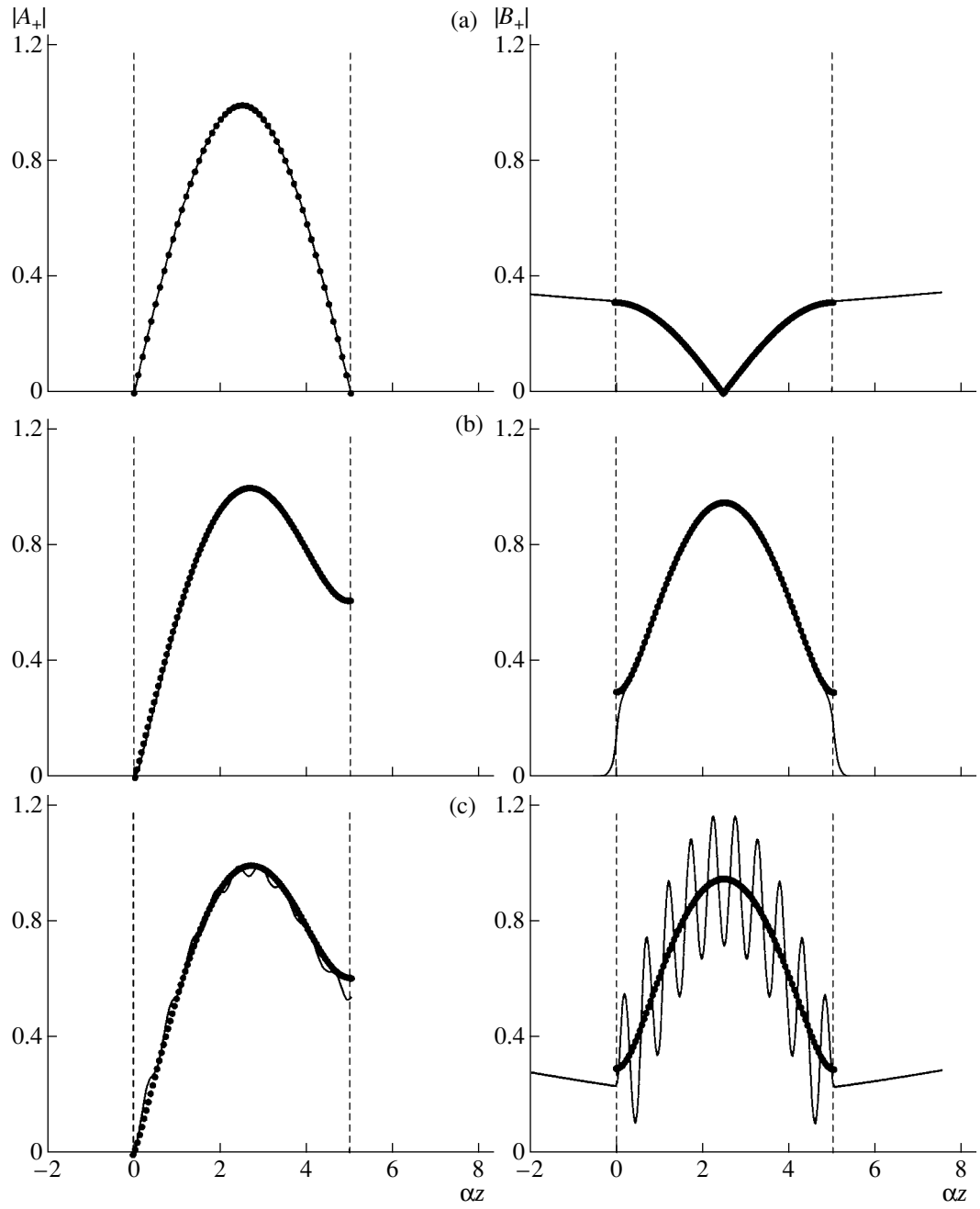


Fig. 4. Longitudinal structure of the partial A_+ and B_+ waves for modes ($m = 0$, $n = 1$) with (a) $\delta \approx 0$, (b) $\delta \approx -2\alpha$, and (c) $\delta \approx +2\alpha$ for the matched (open) resonator at $\bar{\mathbf{h}}/\alpha = 35$ and $\alpha l_x = \alpha l_z = 5$. The circles show the field structures obtained within the quasi-optical approximation. The dashed lines bound the corrugated region.

The analytically found Q factors (loss factors) for the symmetric modes are in good agreement with a numerical solution to the complete characteristic equation that follows from system of linear equations (11) subject to boundary conditions (18) and represents a sixth-order determinant. The complex eigenfrequencies were calculated as zeroes of an analytical function using the principle of argument [14].

It should be noted that, along with the family of modes near the Bragg frequency ($\delta \approx 0$), there exists a family of high-Q symmetric modes with frequencies lying near $\delta \approx \pm 2a$ (Figs. 4b, 4c), i.e., near extra extrema of the dispersion characteristics. Obviously, the group velocity of normal waves at these extrema also tends to zero. However, unlike the dispersion branch that passes near the Bragg frequency, here the second derivative

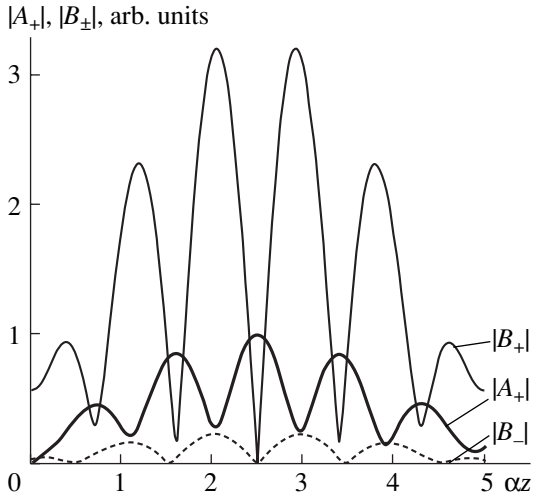


Fig. 5. Longitudinal structure of the partial A_+ and $B_\pm$ waves for the highest-Q asymmetric mode $m = 1$ in the case of the matched resonator at $\bar{\mathbf{h}}/\alpha = 35$ and $\alpha l_x = \alpha l_z = 5$. The frequency of this mode is shown in Fig. 2b by the asterisk.

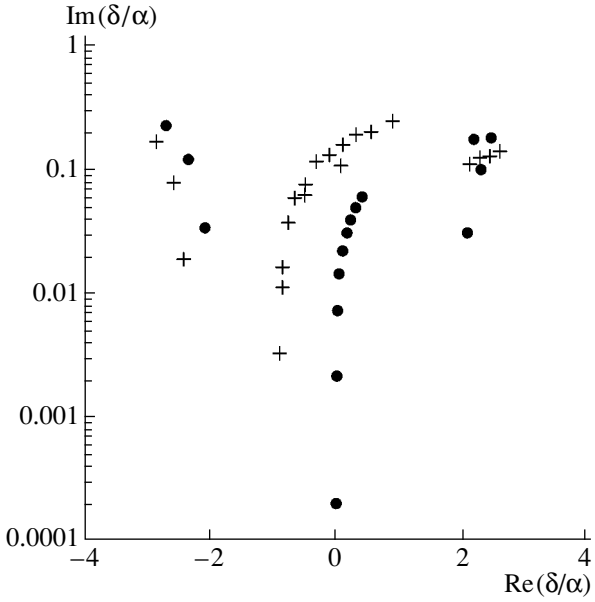


Fig. 6. Mode spectrum of the matched resonator at $\bar{\mathbf{h}}/\alpha = 35$ and $\alpha l_x = \alpha l_z = 5$. Circles, symmetric modes $m = 0$; crosses, asymmetric modes $m = 1$.

remains nonzero. As a result, the Q factor of these modes is considerably lower than those of the modes near $\delta \approx 0$.

For asymmetric modes ($m \neq 0$), complex eigenfrequencies were also found by numerical simulation. It turned out that, of the asymmetric modes differing in longitudinal index n, the mode whose frequency lies near the minimum of the dispersion curve (i.e., near the zero of the group velocity) has the lowest losses (highest Q factor) (cf. [9, 15]). In the dispersion diagram (Fig. 2b), the position of this mode at $m = 1$ is indicated by the asterisk. Figure 5 illustrates the spatial structures of the partial waves of this mode. However, it should be noted that the second derivatives of the asymmetric modes also remain nonzero. Therefore, the Q factor of the asymmetric modes is significantly lower than that of the fundamental symmetric mode, whose frequency is near the Bragg frequency. In addition, the highest-Q asymmetric modes feature a large number of longitudinal variations of partial waves and a much smaller intensity of longitudinal partial waves compared with the transverse partial waves (cf. Figs. 4, 5). Thus, these modes can be additionally selected by an electron beam because of the significant difference in the synchronism conditions or the relatively small amplitude of the operating wave (when describing interaction with an electron beam [3, 6–8], we assume that the synchronous wave is the A_+ wave, which propagates in the positive z direction).

Thus, our analysis of the resonator's eigenmodes shows that, when the perimeter of the system equals its length, the losses of asymmetrical modes are more than an order of magnitude higher than those of the fundamental symmetric mode (Fig. 6). This provides a high electrodynamic selectivity of the resonator.

At the same time, as the l_x of the system increases with its length l_z remaining unchanged, the selectivity of a two-dimensional Bragg resonator somewhat degrades. This may be explained by the fact that, as the perimeter l_x increases, dispersion curves *3* and *4* for asymmetric waves ($m \neq 0$) flatten and, in the limit $l_x \longrightarrow \infty$ (see Eq. (12)), approach the perimeter-independent dispersion characteristic of the symmetric wave. Yet this circumstance places very weak constraints on the system's transverse dimension. Figure 7a plots the minimum loss factor Im(δ) versus the perimeter of the system for the asymmetric modes ($m = 1$) at $\alpha l_z = 5$. The dashed line here shows the diffraction losses of the fundamental (highest-Q) symmetric mode. The losses of these modes become equal at a perimeter $\alpha l_x \geq 80$. The irregularity of the curves in Fig. 7 stems from the fact that the variation of the perimeter shifts the position of the dispersion curve so that its extremum covers modes with different numbers n of longitudinal variations. Thus, as the perimeter varies, the number of longitudinal variations of the highest-Q mode changes. An increase in the azimuthal index m is equivalent to a decrease in the effective perimeter of the system to $l_x^{\text{eff}} = l_x/m$, as follows from dispersion relation (12). Therefore, as follows from Fig. 7, an increase in the azimuthal index, in general, does not increase the Q factor of spurious modes.

It should also be noted that, increasing the length l_z in proportion to the perimeter l_z, one can retain the selectivity of the resonator at large perimeters via an increase in the Q factor of the fundamental symmetric

mode (according to (20), the loss factor of this mode decreases as l_z^5). For comparison, Fig. 7b shows the minimum losses versus perimeter at $\alpha l_z = 10$ in the family of asymmetric modes $m = 1$. As the length increases twofold, the maximum width of the system at which the resonator retains its selectivity (i.e., the Q factor of the fundamental symmetric mode remains higher than the Q factor of the asymmetric modes) grows to $\alpha l_x \le 300$. Also note that all the relationships presented above involve the normalized length and perimeter of the resonator (αl_x and αl_z). Thus, there is an additional way of maintaining the selectivity: to increase the geometrical (absolute) value of the perimeter and simultaneously diminish the coupling coefficient α (for example, by decreasing the corrugation depth), thus keeping the normalized perimeter value constant. Clearly, the length of the resonator l_z should be increased in the same proportion.

Since the nonlinear dynamics of FEMs with two-dimensional Bragg resonators was analyzed in terms of geometrical optics [3, 6–8], it is important to compare our results (the quasi-optical approach) with earlier results for coaxial two-dimensional Bragg resonators (the geometrical optics approximation). In the latter case, the partial wave field distribution over the corrugated region for the fundamental symmetric mode ($\delta \approx 0$) is shown by closed circles in Fig. 4a. The structure of the fundamental mode is virtually the same except for a small region near the corrugation boundary. Thus, in simulating electron beam–wave interaction, the operating wave (A_+) field structure calculated within the geometrical optics approximation adequately describes beam bunching and energy extraction from the beam. At the same time, it is clear that, while the field structure of the $B_\pm$ waves has a discontinuity at the corrugation boundary in terms of geometrical optics, allowance for diffraction through the boundary makes the solutions continuous. Figures 4b and 4c compare the partial field structures obtained within these two approximations for modes with $\delta \approx \pm 2\alpha$. For these modes, the geometrical optics approach also approximates well their spatial structure. The complex eigenfrequencies obtained by the two approaches are also in good agreement. In particular, for the mode illustrated in Fig. 4c, the quasi-optical approximation gives $\delta = 2.09 + i0.032$, while the geometrical optics approximation gives $\delta = 2.09 + i0.035$.

The problem of degeneracy in the longitudinal index and the problem of the infinite Q factor for azimuth-symmetric modes, which arise when the nonlinear dynamics of FEMs is simulated in terms of geometrical optics, were solved as follows. It was assumed that the diffraction Q factor of the symmetric modes far exceeds the ohmic Q factor. This means that the Q factor of these modes is limited by ohmic losses, while the Q factor of the azimuth-asymmetric modes is limited by diffraction losses. Within such a model, degeneracy in the longitudinal index is removed when interaction with the electron beam is taken into account (the admittance due to the electron beam varies significantly with the number of longitudinal variations of the mode excited in the resonator).

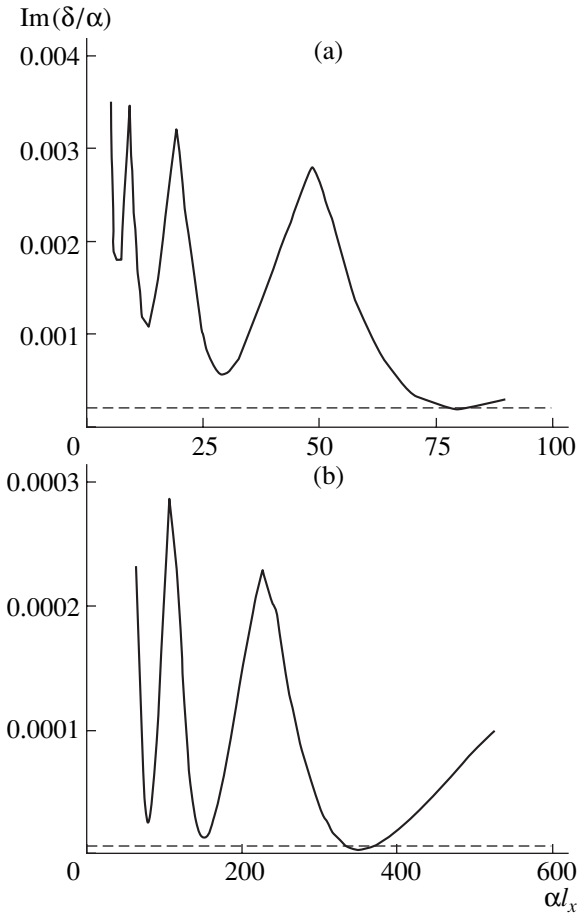


Fig. 7. Diffraction loss factor Im(δ) for the highest-Q mode $m = 1$ versus the system's perimeter αl_x for the matched resonator at $\bar{\mathbf{h}}/\alpha = 35$ and αl_z = (a) 5 and (b) 10. Dashed line, loss factor of the symmetric mode $m = 0$ with the maximum Q factor.

(ii) Closed system. For a resonator that totally reflects transverse electromagnetic energy flows from the corrugation boundaries, the field structure of partial waves that is obtained for the highest-Q azimuth-symmetric mode $m = 0$ by numerically solving Eqs. (10) with boundary conditions (17) and (19) is given in Fig. 8. This mode has one longitudinal field variation for the $A_\pm$ modes and two longitudinal variations for the $B_\pm$ modes, and its frequency is close to the exact Bragg resonance frequency as before. It is clear that energy extraction from the closed resonator is released only by means of $A_\pm$ wave radiation (in Fig. 8, the amplitude of the A_+ wave at the resonator end $z = l_z$ is nonzero though very small).

For asymmetric waves, numerical simulation reveals several families of high-Q modes. As for the

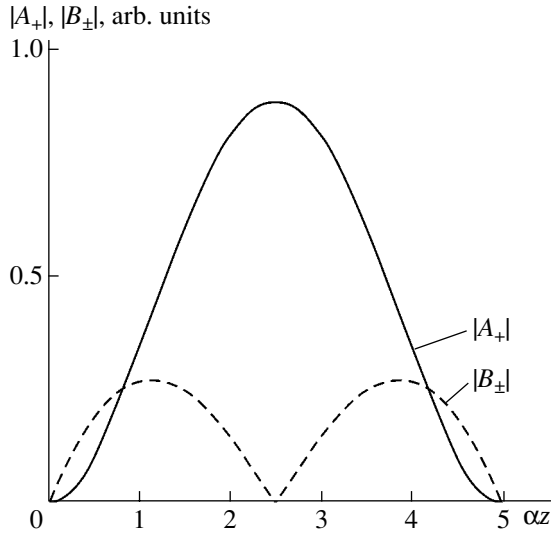


Fig. 8. Longitudinal structure of the partial A_+ and $B_\pm$ waves for the fundamental high-Q mode $m = 0$ of the closed resonator at $\bar{\mathbf{h}}/\alpha = 35$ and $\alpha l_x = \alpha l_z = 5$.

open resonator, modes of the first family are near the minimum of the group velocity and their eigenfrequency is thus markedly different from the Bragg frequency. The structure of the partial waves for the asymmetric mode $m = 1$, which has the highest Q factor in this family, is shown in Fig. 9a; its frequency is indicated by the asterisk in Fig. 2b.

The second family of high-Q asymmetric modes lies near the Bragg frequency ($\delta \approx 0$). This is directly associated with boundary condition (19) and with the mode spectrum generated by this condition (in the absence of corrugation ($\alpha = 0$), these modes have an infinite Q factor). At $\alpha = 0$, these modes consist only of the partial $B_\pm$ waves, which circulate in the azimuthal direction. In general, corrugation decreases the Q factor of these modes because of reradiation into the longitudinal $A_\pm$ modes. However, in the case of doubly periodic corrugation, interference between scattered A_+ and A_- waves may result in a situation where the energy flux from the resonator is absent and the Q factor of some modes tends to infinity even at a nonzero coupling coefficient. Indeed, if the condition

$$\frac{\pi^2 n^2}{2\bar{h} l_z^2} = \frac{2\pi m}{l_x} \tag{23}$$

is met, one can easily obtain the solution

$$B_- \equiv 0, \quad B_+ = C\sin\left(\frac{n\pi}{l_z}z\right),$$
$$A_+ = -A_- = -\frac{i\alpha l_x}{n\pi}C\left(1 - \cos\left(\frac{n\pi}{l_z}z\right)\right) \tag{24}$$

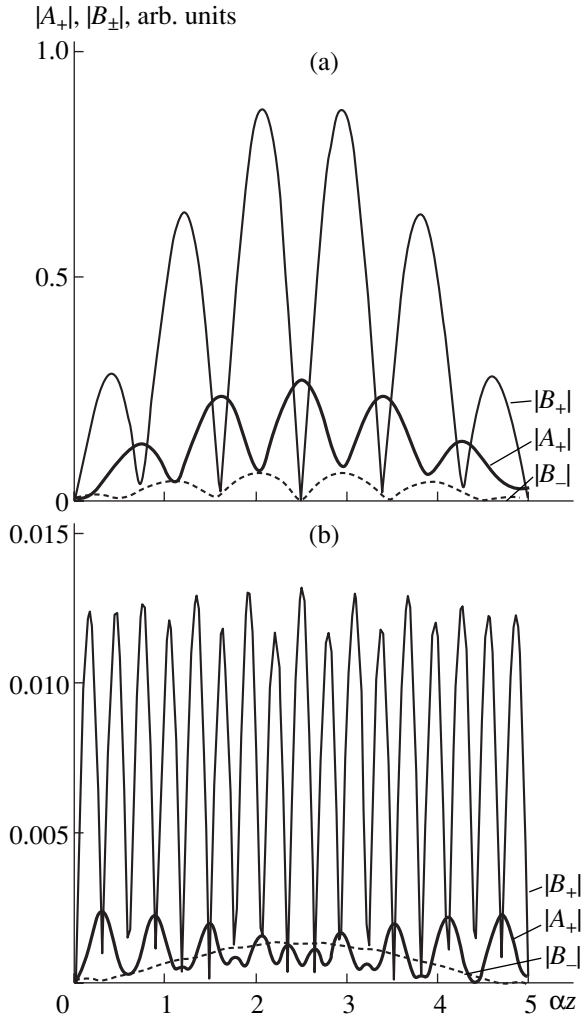


Fig. 9. Longitudinal structure of the partial A_+ and $B_\pm$ waves for asymmetric modes $m = 1$, which have a maximum Q factor among modes from two different families, in the case of the closed resonator at $\bar{\mathbf{h}}/\alpha = 35$, $\alpha l_x = \alpha l_z = 5$, and $\mathrm{Re}(\delta/\alpha) = $ (a) –0.91 and (b) 0.37. The frequencies of these modes are indicated by the asterisk and diamond, respectively, in Fig. 2.

to Eqs. (10) at the exact Bragg frequency ($\delta = 0$) in view of boundary conditions (19).

Since electromagnetic energy fluxes through the resonator boundary is zero ($A_+ = A_- = 0$), the Q factor of this mode tends to infinity. Note, however, that this mode exists only if the partial waves $B_\pm$ are almost totally reflected and the geometrical dimensions of the system are related by (23). Even a small deviation of the dimensions from those defined by formula (23) adversely affects the Q factor of this mode family. In particular, the simulation of the two-dimensional Bragg resonator with $\alpha l_x = \alpha l_z = 5$ show (Fig. 9b) that, when condition (23) is violated, the partial waves of these modes are no longer symmetric, $A_+(x, z) \neq -A_-(x, z)$, and, accordingly, $B_-(x, z) \neq 0$. The eigenfrequencies of

these modes can be roughly evaluated from the relationship $\mathrm{Re}(\delta_n) = \pi^2 n^2/2\bar{h}l_z^2 - 2\pi m/l_x$. Eventually, for the parameters chosen, the Q factors of the spurious modes considered become much lower than the Q factor of the fundamental azimuth-symmetric mode. It is also noteworthy that, as for the open (matched) resonator, high-Q asymmetric modes have a considerably greater number of longitudinal variations, and the A_+ wave synchronous with the electron beam has a lower amplitude compared with these parameters for the symmetric fundamental mode (Fig. 9). This circumstance extends the possibility of electronically selecting these waves when they interact with the electron beam, because resonance excitation conditions for the spurious modes are bound to differ from the synchronism condition for the fundamental mode. In addition, since the structural factor is $A_+^{\max}/B_+^{\max} \ll 1$, these waves are coupled to the beam more weakly.

In general, a closed resonator seems to have the more complex design from the standpoint of provision of selectivity. Nevertheless, our analysis shows that the Q factor of the fundamental azimuth-symmetric mode may far exceed the Q factors of the other modes even in a substantially overmoded resonator (Fig. 10).

To conclude, the parameters $\alpha l_x = \alpha l_z = 5$ of the two-dimensional Bragg resonator that were used in the simulation meet the experimental conditions in which a coaxial FEM with two-dimensional distributed feedback was investigated with a high-current accelerator at Strathclyde University [5]. In those experiments, the two-dimensional Bragg resonator designed for coupling the TEM and $TE_{25,0}$ waves (with coupling coefficient $\alpha \approx 0.2$ cm^{-1}) at an operating wavelength of 8 mm had a characteristic length l_z and perimeter l_x of about 25 cm. Thus, the given resonator can be used for mode selection by azimuthal and longitudinal indices under any boundary conditions (specified by the shape of the resonator and corrugation) imposed on the transverse $B_\pm$ flows.

ACKNOWLEDGMENTS

We are grateful to N.F. Kovalev and M.I. Petelin for encouraging discussions.

This work was supported in part by the Russian Foundation for Basic Research and INTAS.

APPENDIX

Let us derive formulas for the complex frequencies of symmetric modes lying near the Bragg frequency, i.e., of those belonging to branch *3* of the dispersion diagram in Fig. 2a. For azimuth-symmetric waves, the set of equations (11) can be reduced to

$$\frac{dA_+}{dz} + i\delta A_+ + 2i\alpha B = 0, \tag{A1}$$

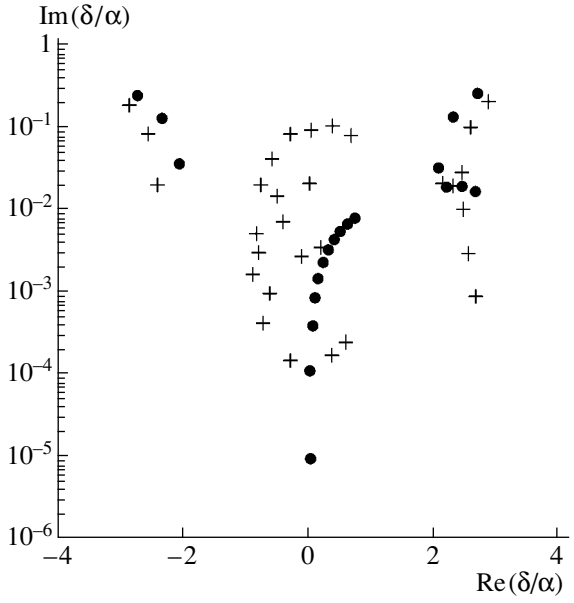


Fig. 10. Mode spectrum in the closed resonator at $\bar{\mathbf{h}}/\alpha = 35$ and $\alpha l_x = \alpha l_z = 5$. Circles, symmetric modes ($m = 0$); crosses, asymmetric modes ($m = 1$).

$$\frac{dA_-}{dz} - i\delta A_- - 2i\alpha B = 0, \tag{A2}$$

$$\frac{1}{2\bar{h}}\frac{d^2B}{dz^2} + \delta B + \alpha(A_+ + A_-) = 0 \tag{A3}$$

with the boundary conditions

$$\begin{aligned} \frac{dB}{dz} - i\sqrt{2\bar{h}\delta}B|_{z=0} &= 0, \\ \frac{dB}{dz} + i\sqrt{2\bar{h}\delta}B|_{z=l_z} &= 0. \end{aligned} \tag{A4}$$

Representing the partial waves as $A_\pm = a_\pm e^{i\lambda z}$ and $B_\pm = be^{i\lambda z}$, we obtain dispersion relation (14), which is a fourth-degree algebraic equation in λ and, consequently, has four roots λ_k ($k = 1$–4).

Next, we calculate the eigenvectors of the matrix T of coefficients for the system of equations (A1)–(A3):

$$T\begin{pmatrix} a_k^+ \\ a_k^- \\ b_k \\ b_k' \end{pmatrix} = i\lambda_k \begin{pmatrix} a_k^+ \\ a_k^- \\ b_k \\ b_k' \end{pmatrix}$$

(here, b_k' is the derivative of B); represent the solution as

$$A_\pm(z) = \sum_{k=1}^{4} c_k a_k^\pm e^{i\lambda_k z}, \quad B(z) = \sum_{k=1}^{4} c_k b_k e^{i\lambda_k z}; \tag{A5}$$

and substitute (A5) into boundary conditions (A4) to obtain a characteristic equation for the complex eigenfrequencies δ:

$$\Delta = 0. \tag{A6}$$

Here, Δ is the determinant of the matrix of coefficients for arbitrary constants c_k. In the case $|\delta| \ll 1$ under study, dispersion relation (14) can be approximately written as

$$\lambda^4 = 8\bar{h}\alpha^2\delta.$$

The roots λ_k of this equation are four values of $\sqrt[4]{8\bar{h}\alpha^2\delta}$. Let $\tilde{\lambda}$ be that branch of the root corresponding to a positive real number at a positive real δ. The complex plane δ is cut along the imaginary semiaxis $\mathrm{Im}\,\delta \geq 0$ (the same refers to the term $\sqrt{2\bar{h}\delta}$ in boundary conditions (14)). Thus, we can write $\lambda_{1,2} \approx \pm\tilde{\lambda}$ and $\lambda_{3,4} \approx \pm i\tilde{\lambda}$. Now we drop the terms of higher order of smallness in $\sqrt[4]{\delta}$ to obtain

$$\Delta \approx \begin{vmatrix} 1 & -1 & i & -i \\ e^{i\tilde{\lambda}l_z} & -e^{i\tilde{\lambda}l_z} & ie^{-\tilde{\lambda}l_z} & -ie^{\tilde{\lambda}l_z} \\ i & -i & 1 & -1 \\ ie^{i\tilde{\lambda}l_z} & -ie^{-i\tilde{\lambda}l_z} & e^{-\tilde{\lambda}l_z} & -e^{\tilde{\lambda}l_z} \end{vmatrix}$$

$$+\left\{\begin{vmatrix} 1 & -1 & i & -i \\ e^{i\tilde{\lambda}l_z} & -e^{i\tilde{\lambda}l_z} & ie^{-\tilde{\lambda}l_z} & -ie^{\tilde{\lambda}l_z} \\ i & -i & i & i \\ ie^{i\tilde{\lambda}l_z} & -ie^{-i\tilde{\lambda}l_z} & e^{-\tilde{\lambda}l_z} & -e^{\tilde{\lambda}l_z} \end{vmatrix}\right. \tag{A7}$$

$$\left.+\begin{vmatrix} 1 & -1 & i & -i \\ e^{i\tilde{\lambda}l_z} & -e^{i\tilde{\lambda}l_z} & ie^{-\tilde{\lambda}l_z} & -ie^{\tilde{\lambda}l_z} \\ i & -i & 1 & -1 \\ ie^{i\tilde{\lambda}l_z} & ie^{-i\tilde{\lambda}l_z} & -ie^{-\tilde{\lambda}l_z} & -ie^{\tilde{\lambda}l_z} \end{vmatrix}\right\},$$

where

$$\varepsilon = \sqrt[4]{\frac{\delta\bar{h}}{2\alpha^2}}.$$

With (A7), characteristic equation (A6) reduces to

$$\begin{aligned}(e^{\tilde{\lambda}l_z} - e^{-\tilde{\lambda}l_z})(e^{-i\tilde{\lambda}l_z} - e^{i\tilde{\lambda}l_z}) &= \varepsilon\{(e^{\tilde{\lambda}l_z} - e^{-\tilde{\lambda}l_z}) \\ \times (e^{-i\tilde{\lambda}l_z} + e^{i\tilde{\lambda}l_z}) - i(e^{\tilde{\lambda}l_z} + e^{-\tilde{\lambda}l_z})(e^{-i\tilde{\lambda}l_z} - e^{i\tilde{\lambda}l_z})\}.\end{aligned} \tag{A8}$$

It is easy to check that, at $\alpha l_z \gg 1$, the roots of Eq. (A8) are given by the approximate expressions

$$\tilde{\lambda}l_z \approx \pi n + i\varepsilon,$$

which yield formula (20).

REFERENCES

1. A. V. Arzhannikov, N. S. Ginzburg, V. S. Nikolaev, N. Yu. Peskov, A. S. Sergeev, S. L. Sinitsky, R. P. Zotkin, and M. V. Yushkov, in *Abstracts of the 14th International Conference on Free Electron Lasers (FEL), Kobe, Japan, 1992*, p. 214.
2. N. S. Ginzburg, N. Yu. Peskov, and A. S. Sergeev, Pis'ma Zh. Tekh. Fiz. **18** (9), 23 (1992) [Sov. Tech. Phys. Lett. **18**, 285 (1992)].
3. N. S. Ginzburg, N. Yu. Peskov, and A. S. Sergeev, Opt. Comm. **112**, 151 (1994).
4. A. V. Arzhannikov, N. V. Agarin, V. B. Bobylev, *et al.*, Nucl. Instrum. Methods Phys. Res. A **445**, 222 (2000).
5. A. W. Cross, I. V. Konoplev, K. Ronald, *et al.*, Appl. Phys. Lett. **80**, 1517 (2002).
6. N. S. Ginzburg, I. V. Konoplev, and A. S. Sergeev, Zh. Tekh. Fiz. **66** (5), 108 (1996) [Tech. Phys. **41**, 465 (1996)].
7. N. S. Ginzburg, N. Yu. Peskov, A. D. R. Phelps, *et al.*, IEEE Trans. Plasma Sci. **24**, 770 (1996).
8. N. S. Ginzburg, N. Yu. Peskov, A. S. Sergeev, *et al.*, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **44**, 533 (2001).
9. L. A. Vainshtein, *Electromagnetic Waves* (Sov. Radio, Moscow, 1957).
10. N. F. Kovalev, M. I. Petelin, and M. G. Reznikov, USSR Inventor's Certificate No. 720592, Byull. Izobret., No. 9 (1980).
11. V. L. Bratman, N. S. Ginzburg, and G. G. Denisov, Pis'ma Zh. Tekh. Fiz. **7**, 1320 (1981) [Sov. Tech. Phys. Lett. **7**, 565 (1981)].
12. G. G. Denisov and M. G. Reznikov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **25**, 562 (1982).
13. V. L. Bratman, M. A. Moiseev, and M. I. Petelin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **16**, 622 (1973).
14. M. A. Lavrent'ev and B. V. Shabat, *Methods of the Theory of Functions of Complex Variable* (Nauka, Moscow, 1973).
15. G. G. Denisov, Candidate's Dissertation (Inst. Appl. Phys., Acad. Sci. USSR, Gor'kii, 1985).

Translated by A. Khzmalyan

Technical Physics, Vol. 48, No. 12, 2003, pp. 1565–1569. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 66–70.

RADIOPHYSICS

Quality Evaluation of Tunable Microwave Filters on Ferroelectric Capacitors

V. V. Pleskachev and I. B. Vendik

St. Petersburg State Electrotechnical University (LÉTI), ul. Prof. Popova 5, St. Petersburg, 197376 Russia

e-mail: mwlab@eltech.ru

Received March 18, 2003

Abstract—The tuning efficiency of microstrip filters that use ferroelectric capacitors as control elements depends on the properties of the capacitors and the filter's resonators. The properties of these components are included if the quality criterion of a tunable filter is defined as the ratio of the center frequency tuning bandwidth to the passband of the filter. A quality criterion suggested in this work allows one to estimate the limiting characteristics of tunable filters. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Planar tunable filters may find wide application in modern telecommunication systems. These filters are fabricated by several competing technologies, which differ in the way in which the center frequency of the filter's resonators is tuned. This parameter may be tuned mechanically [1] or with ferromagnetic media [2], semiconductor varactors [3], microelectromechanical capacitors [4], or ferroelectric (FE) capacitors [5]. The last three methods use, in essence, electrically controlled variable capacitors. Tunable filters with FE capacitors seem to be the most promising, since they offer rapid center frequency tuning and are easy to fabricate (and, accordingly, cheap).

The basic parameters of tunable filters are the same as those of filters with a fixed center frequency (passband, insertion loss, wavefront steepness, etc.). However, in the former case, these parameters must be defined at the upper and lower limits of the center frequency tuning bandwidth. Usually, the quality of a tunable filter is described by the ratio of the center frequency tuning bandwidth to the geometrical mean of its passbands at the lower and upper center frequencies [6]. Such an estimate suffers from a serious disadvantage: it does not include insertion losses. The quality criterion suggested in this work is related to the Q factor of the microstrip lines and losses in FE capacitors. It also depends on the coefficient of coupling between the capacitors and microstrip lines of the resonators.

TUNABLE MICROSTRIP FILTERS

(i) Structure of the resonators. Consider two types of microstrip resonators with FE capacitors (Fig. 1). The electrical lengths of the microstrip lines of the resonators are designated by Θ_0 and Θ_g; C is the variable capacitance of the FE resonator, which depends on the applied voltage. The choice of these resonators stems from the fact that they are readily compatible with the planar technology of microstrip tunable filters.

(ii) Tuning ability of the resonators. Resonance conditions for short- and open-circuited resonators are written as

$$\frac{y_0}{\omega_0 C} - \tan(\Theta_0) + \cot(\Theta_g) = 0, \quad (1a)$$

$$\frac{y_0}{\omega_0 C} + \cot(\Theta_0) + \cot(\Theta_g) = 0, \quad (1b)$$

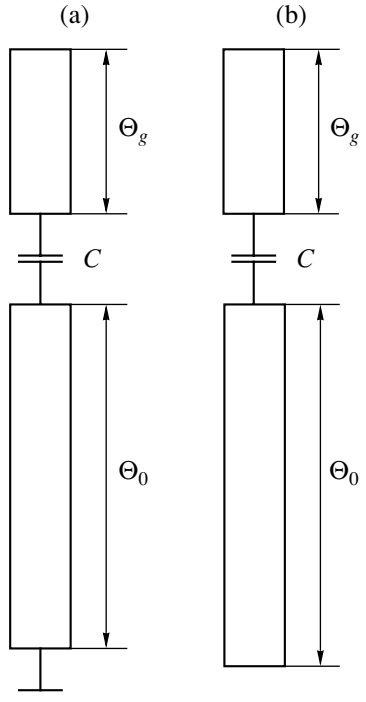


Fig. 1. Tunable microstrip resonators with FE capacitors: (a) short-circuited resonator and (b) open-circuited resonator.

1063-7842/03/4812-1565$24.00 © 2003 MAIK "Nauka/Interperiodica"

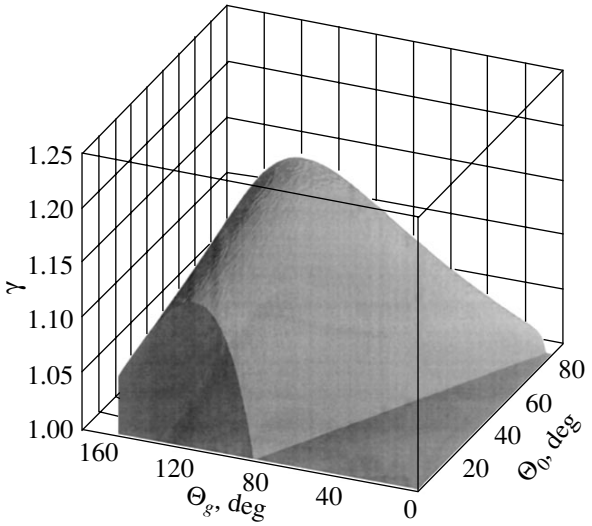


Fig. 2. Shift of the relative center frequency of the short-circuited resonator vs. electrical lengths of the microstrip sections.

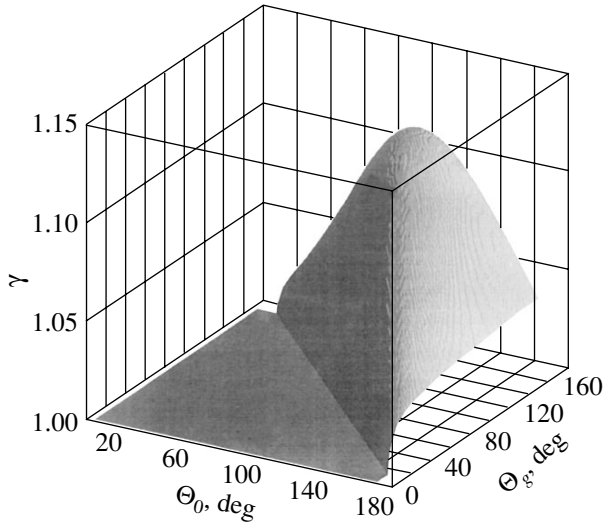


Fig. 3. Shift of the relative center frequency of the open-circuited resonator vs. electrical lengths of the microstrip sections.

respectively, where Θ_0 and Θ_g are the electrical lengths of microstrips at a resonance frequency ω_0.

Clearly, for resonance to occur, the conditions

$$\tan(\Theta_0) - \cot(\Theta_g) > 0, \tag{2a}$$

$$\cot(\Theta_0) + \cot(\Theta_g) < 0 \tag{2b}$$

should be satisfied.

Let us introduce a parameter that relates the tuning bandwidth of the resonator to the relative shift of the center frequency:

$$\gamma = \omega_0^{up}/\omega_0^{low}, \tag{3}$$

where ω_0^{low} and ω_0^{up} are the lower and upper center frequencies and $\gamma > 1$.

Hereafter, the electrical lengths Θ_0 and Θ_g will be calculated at the lower frequency ω_0^{low}. At the upper center frequency, the electrical lengths will be equal to $\gamma\Theta_0$ and $\gamma\Theta_g$, respectively. With (3), resonance conditions (1a) and (1b) can be transformed into

$$\frac{n}{\gamma} = \frac{\tan(\gamma\Theta_0) - \cot(\gamma\Theta_g)}{\tan(\Theta_0) - \cot(\Theta_g)}, \tag{4a}$$

$$\frac{n}{\gamma} = \frac{\cot(\gamma\Theta_0) + \cot(\gamma\Theta_g)}{\cot(\Theta_0) + \cot(\Theta_g)}, \tag{4b}$$

where $n = C_1/C_2$ is the tuning ability of an FE capacitor. Equations (4a) and (4b) are solved numerically for γ. Figures 2 and 3 show the tuning coefficient γ vs. the electrical lengths Θ_0 and Θ_g for resonators of the two types with a tuning ability $n = 2$.

For the short-circuited resonator, the maximal relative shift of the center frequency is 1.2 for $\Theta_0 = 45°$ and $\Theta_g = 90°$. For the open-circuited resonator, the shift equals 1.13 for $\Theta_0 = 120°$ and $\Theta_g = 120°$.

(iii) Q factor of the resonators. The Q factor of microstrip resonators with tunable FE capacitors is calculated with the method described in [5]. The Q factor of a short-circuited resonator is given by

$$Q = Q_0 \times \frac{\dfrac{\Theta_0}{\cos^2(\Theta_0)} + \dfrac{\Theta_g}{\sin^2(\Theta_g)} + \tan(\Theta_0) - \cot(\Theta_g)}{\dfrac{\Theta_0}{\cos^2(\Theta_0)} + \dfrac{\Theta_g}{\sin^2(\Theta_g)} + (\tan(\Theta_0) - \cot(\Theta_g))2Q_0\tan\delta}, \tag{5a}$$

where Q_0 is the Q factor of the microstrip lines, which is defined as the ratio of the propagation constant of the line to the double damping factor, and $\tan\delta$ is the loss tangent of the FE capacitor.

The Q factor of an open-circuit resonator is given by

$$Q = Q_0 \times \frac{\dfrac{\Theta_0}{\cos^2(\Theta_0)} + \dfrac{\Theta_g}{\sin^2(\Theta_g)} - (\cot(\Theta_0) + \cot(\Theta_g))}{\dfrac{\Theta_0}{\sin(\Theta_0)^2} + \dfrac{\Theta_g}{\sin^2(\Theta_g)} - (\cot(\Theta_0) + \cot(\Theta_g))2Q_0\tan\delta}. \tag{5b}$$

As with the resonance conditions, only a definite set of pairs of electrical lengths that satisfy inequalities (2a) and (2b) may be used to calculate Q.

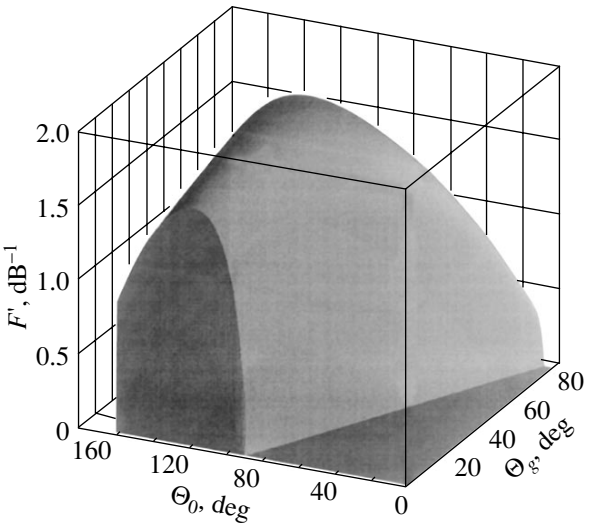


Fig. 4. Modified quality criterion of the tunable filter with short-circuited resonators.

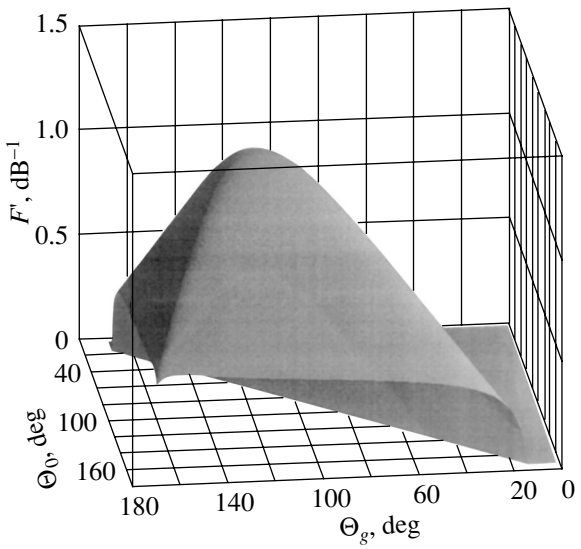


Fig. 5. Modified quality criterion of the tunable filter with open-circuited resonators.

QUALITY CRITERION FOR A TUNABLE FILTER

There is an empirical relationship between the insertion losses (in dB) and the intrinsic Q factor of a filter [7]:

$$L = \frac{4.34N}{(\Delta\omega_f/\omega_{0f})Q}, \quad (6)$$

where N is the filter order, $\Delta\omega_f$ is the filter passband, ω_{0f} is the center frequency of the filter, and Q is the intrinsic Q factor.

Since $Q = \omega_{0f}/\Delta\omega_r$, where $\Delta\omega_r$ is the 3-dB width of the resonance curve, expression (6) can be recast as

$$\Delta\omega_f = \Delta\omega_r\frac{4.34N}{L}. \quad (7)$$

The quality criterion of a tunable filter that was introduced in [6] is

$$F = \frac{\omega_0^{up} - \omega_0^{low}}{\sqrt{\Delta\omega_f^{up}\Delta\omega_f^{low}}}. \quad (8)$$

With (7), we rearrange (8) into the form

$$F = \frac{\sqrt{L^{up}L^{low}}}{4.34N}\frac{\omega_0^{up} - \omega_0^{low}}{\sqrt{\Delta\omega_r^{up}\Delta\omega_r^{low}}}. \quad (9)$$

Expression (9) can be written as

$$F = \frac{\sqrt{L^{up}L^{low}}}{4.34N}\frac{\sqrt{\gamma} - 1/\sqrt{\gamma}}{\sqrt{(Q^{up})^{-1}(Q^{low})^{-1}}}, \quad (10)$$

where Q^{low} and Q^{up} are the intrinsic Q factors of the filter at the lower and upper center frequencies and $\sqrt{L^{up}L^{low}}$ is the geometrical mean of the losses at both frequencies.

One may use a modified definition of the quality criterion of a tunable filter. In this case, the quality criterion depends not only on the number of tuning bandwidths (order) but also on insertion losses:

$$F' = F\frac{1}{\sqrt{L^{up}L^{low}}} = \frac{1}{4.34N}\frac{\sqrt{\gamma} - 1/\sqrt{\gamma}}{\sqrt{(Q^{up})^{-1}(Q^{low})^{-1}}}. \quad (11)$$

Figures 4 and 5 show the modified quality criterion for a third-order tunable microstrip filter with short- and open-circuited resonators, respectively. To calculate the modified quality criterion, we used the parameters of copper microstrip resonators made on a two-layer insulating substrate (a 0.5-mm-thick Polikor layer and a 1-μm-thick BSTO layer) with a permittivity of 1000 and a loss tangent of 0.01. Integrated FE capacitors represent 5- to 10-μm-wide gaps between two segments of the microstrip line of the resonator. The width of the capacitors was selected such that it provided a desired capacitance value. The loss tangents of the FE capacitors ($n = 2$) were $\tan\delta_{low} = 0.01$ (for the zero control voltage) and $\tan\delta_{up} = 0.005$ (for the maximal control voltage). The layout of the tunable microstrip filter is presented in Fig. 6. Each of the resonators is equipped with a special bias circuit to control the capacitance of the FE capacitor. Measured and calculated characteristics of the filter are demonstrated in Fig. 7. The center frequency shifts from 4.4 to 4.65 GHz. Concurrently, the insertion losses decrease from 15 to 8 dB. The geometrical mean value of the passband equals 80 MHz. Table 1 lists the parameters of the FE capacitor at various control voltages. These

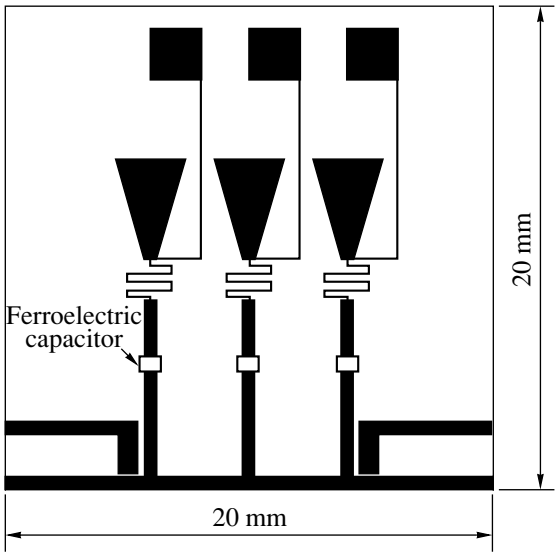


Fig. 6. Layout of a tunable filter prototype with short-circuited resonators.

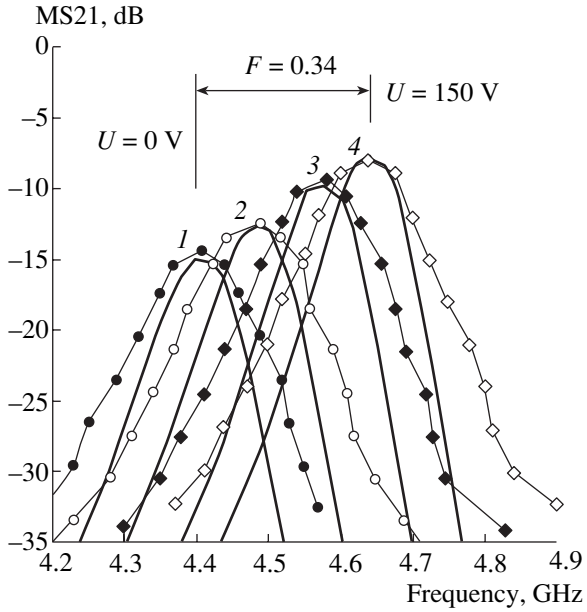


Fig. 7. Experimental and calculated characteristics of the tunable filter shown in Fig. 6.

parameters were derived from the experimental characteristics of the filter. The quality criterion F' was found to be $F'_{exp} = 0.34$ dB^{-1}, which agrees well with the calculated value $F'_{calc} = 0.38$ dB^{-1}.

If the Q factor of the microstrip lines far exceeds the reciprocal value of $\tan\delta$ of the FE capacitor, the modified quality criterion does not depend on the electrical lengths of the microstrips of the resonator. As follows from calculations, the modified quality criterion of a tunable microstrip filter depends only on the order N of the filter and on the switching parameter K of the FE capacitor provided that the condition $Q_0 \gg \tan\delta^{-1}$ is satisfied (as in the case of superconducting microstrips lines). The switching parameter is given by [6]

$$K = \frac{(n-1)^2}{n\tan\delta_{low}\tan\delta_{up}}. \quad (12)$$

In this case, the modified quality criterion of a tunable filter is found by the formula

$$F' = \frac{1}{8.68N}\sqrt{K}. \quad (13)$$

The modified quality criterion of a lossless tunable filter incorporated into microstrip lines is a limiting quantity achievable with tunable capacitors of given quality. The estimates of this limit are given in Table 2 for filters of different order that use capacitors of a given quality. At a frequency of 10 GHz, the switching parameter of the best FE capacitors reaches 5000.

Table 1. Parameters of the planar BSTO capacitor that correspond to the characteristics in Fig. 7

U, V	C, pF	$\tan\delta$
0	0.35	0.044
50	0.29	0.034
100	0.23	0.023
150	0.19	0.017

Table 2. Modified quality criterion of a tunable filter with perfect microstrip lines

K	N	F', dB^{-1}
5000	2	4.1
	3	2.7
	4	2.0
15 000	2	7.1
	3	4.7
	4	3.5
	5	2.8

CONCLUSIONS

The limiting characteristics of tunable microstrip filters are presented. A modified quality criterion of tunable filters is introduced. This parameter allows one to estimate the performance of a tunable filter vs. the parameters of its microstrip lines and the quality of the control capacitors.

ACKNOWLEDGMENTS

The authors thank O.G. Vendik for the fruitful discussion.

This work was supported by the Ministry of Science, Industry, and Technology of the Russian Federation (project no. 239 SP/MLP-10, 2002).

REFERENCES

1. H. Xu, E. Gao, and Q. Y. Ma, IEEE Trans. Appl. Supercond. **11**, 353 (2001).
2. D. E. Oates and G. F. Dionne, IEEE Trans. Appl. Supercond. **9**, 4170 (1999).
3. A. R. Brown and G. M. Rebeiz, IEEE Trans. Microwave Theory Tech. **48**, 1157 (2000).
4. D. Peroulis, S. Pacheco, K. Sarabandi, and L. P. B. Katehi, IEEE MTT-S Int. Microwave Symp. Dig. **1**, 341 (2001).
5. I. Vendik, O. Vendik, V. Pleskachev, and M. Nikol'sky, Integr. Ferroelectr. **49**, 83 (2002).
6. I. B. Vendik, O. G. Vendik, and E. L. Kollberg, in *Proceedings of the 29th Europe Microwave Conference, Munich, 1999*, Vol. 3, pp. 187–190.
7. G. Matthaei, L. Young, and E. M. T. Jones, *Microwave Filters, Impedance Matching Networks, and Coupling Structures* (Artech, Boston, 1980).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 12, 2003, pp. 1570–1574. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 71–75.

ELECTRON AND ION BEAMS, ACCELERATORS

Effect of the Periodic Ion Density Channel on the Behavior of the Ion Hose Instability of a Relativistic Electron Beam

A. G. Zelensky and E. K. Kolesnikov

Smirnov Research Institute of Mathematics and Mechanics,
St. Petersburg State University, St. Petersburg, 198904 Russia
e-mail: zel_alex@-mails.ru

Received July 24, 2002; in final form, May 16, 2003

Abstract—The dynamics of a relativistic electron beam propagating in an ion channel with a periodically varying density is considered. The behavior of the ion hose instability at different parameters of the beam–ion channel system is studied using the spread mass model. Conditions are determined under which the ion hose instability does not hinder the beam propagation over distances on the order of 100 betatron lengths of the beam.

One of the new promising methods for guiding relativistic electron beams (REBs) in gas–plasma media is based on the idea of transporting an electron beam through a preformed plasma channel produced artificially by ionizing the neutral component of the background gas with auxiliary UV laser radiation. The role of the plasma channel so produced is twofold. On the one hand, it neutralizes the perturbing effect of various external forces on the beam, so that an REB propagates along a nearly straight path. On the other hand, the channel ensures equilibrium conditions for guiding beams with radii of 1–10 cm and with currents significantly higher than the current of a transversely stabilized REB propagating in a spatially uniform plasma.

The distance over which an REB can be guided with an artificially preformed plasma channel is determined by a variety of dynamic processes that gradually destroy the beam.

The evolution of a beam–channel system such that the beam duration is comparable to the characteristic bounce period of the plasma channel ions in the potential well of the beam is governed by the common dynamics of the REB electrons and channel ions. Both theory and experiment show that, in such an interaction, conditions in the beam–channel system may become favorable for the onset of various instabilities, the most dangerous of which is ion hose instability (IHI) [1–9].

In [10], it was proposed that, in an undulator of a free electron laser, an ion channel with a periodically varying density be used to guide electron beams over distances on the order of several betatron lengths of the beam and to excite transverse oscillations of the beam electrons over the same distances. Studies on this subject were continued in [11, 12] on the basis of the model of a rigid beam. In this model, the growth of IHI is absolute. That is why, in [11, 12], the effect of phase mixing of the beam electron trajectories was taken into account by introducing the dissipation coefficient. The results obtained in [10–12] showed that, in the absence of external focusing fields, a beam in which the outer electrons execute cylindrically symmetric oscillations can be stably guided over distances of several betatron lengths by means of an ion channel in which the spatial period of density variation is equal to the betatron length of the beam.

In this paper, we investigate the onset and behavior of the IHI of an REB propagating over a large distance along an ion channel with a periodically varying density and analyze how the instability dynamics depends on different parameters of the beam–channel system. In contrast to [11, 12], we do not assume that the amplitudes of transverse oscillations of the beam electrons and channel ions are small.

We consider a paraxial axisymmetric REB propagating in the z direction along a preformed plasma channel with a periodically varying density. We assume that the plasma electrons are instantaneously pushed away from the beam path in the radial direction by the strong electric field of the beam front, so that the beam is guided against the background of positively charged channel ions, which partially neutralize the beam space charge. The spatial period of density variation in the ion channel, L_{per}, varies from 0 to $2\lambda_{\beta e}$, where $\lambda_{\beta e}$ is the wavelength of the betatron oscillations of the beam electrons. The density variation in the ion channel is such that the degree of charge neutralization of the beam, f, has the form $f(z) = f_0[1 + f_1\cos(2\pi z/L_{per})]$. Here, f_0 is the degree of charge neutralization of a beam in an ion channel with uniform (nonvarying) density (the case in which an electron beam is guided along a uniform ion channel in the ion-focused regime), the coefficient f_1 varies from 0 to f_0, and z is the ion channel length. We consider both narrow and wide ion chan-

nels, i.e., those whose characteristic radii are, respectively, smaller and larger than the beam radius. The length of the electron beam, X, measured from the beam front, is set equal to $2\lambda_{\beta i}$, where $\lambda_{\beta i}$ is the wavelength of oscillations of the channel ions about the beam. The distance z over which an REB is transported in an ion channel is set equal to 100 (in units of the betatron length $\lambda_{\beta e}$ of the beam).

Unfortunately, the behavior of IHI can be studied analytically only in the linear stage and only in a few cases. The nonlinear instability stage, in which the displacements of the beam with respect to the plasma channel are comparable to the transverse size of the beam or the channel, can only be studied by numerical methods.

In order to investigate unstable excited states of the beam–channel system, we apply the spread mass model, which was developed in [13, 14] and was used in [1] to study the guiding of a beam with a uniform ion channel whose radius is larger than the beam radius.

In this model, the beam (channel) is divided into segments, each of thickness $\Delta\tau$ in the longitudinal direction, in the form of a sequence of rigid disks having the same density profile as the beam (channel). The mass of the segment changes from disk to disk and ranges between an infinitely large value (which corresponds to a particle far from the system axis) and a nonzero minimum value (which corresponds to a particle near the axis $r = 0$).

The common dynamics of the disks is described by the set of partial differential equations formulated in [1, 13]. For Gaussian radial profiles of the beam electron density and the ion density in the channel, the equations have the form

$$\begin{cases} \dfrac{\partial^2 Y_\eta}{\partial z^2} = -2\eta k_{\beta e}^2 \dfrac{a^2}{Y_\eta - D}\left\{1 - \exp\left[-\dfrac{(Y_\eta - D)^2}{2a^2}\right]\right\}, \\ \dfrac{\partial^2 D_\nu}{\partial x^2} = -2\nu k_{\beta i}^2 \dfrac{a^2}{D_\nu - Y}\left\{1 - \exp\left[-\dfrac{(D_\nu - Y)^2}{2a^2}\right]\right\}, \end{cases}$$

where Y_η and D_ν are the radial displacements of the ηth disk of the beam and the νth disk of the channel, the variable $x = ct - z$ is the distance from the beam front, z plays the role of a time variable, and c is the speed of light.

The maximum displacements Y and D of the centers of mass of the beam and channel segments are related to the displacements of the disks by the expressions

$$Y = \int_0^1 Y_\eta w_e(\eta)\, d\eta,$$

$$D = \int_0^1 D_\nu w_i(\nu)\, d\nu.$$

For Gaussian density profiles in the beam and in the channel, the weighting functions $w_{e,i}$ have the form [1] $w_e(\eta) = w_i(\eta) = 12\eta^2(1 - \eta)$. The rest of the notation in the above equations is as follows: $k_{\beta e}^2$ is the squared betatron wavenumber of the beam electrons; $k_{\beta i}^2$ is the squared wavenumber corresponding to the frequency of the radial oscillations of the ions in the channel; and the quantity in the denominators is expressed as $a^2 = (R_b^2 + R_c^2)/2$, where R_b and R_c are the beam and channel radii, respectively.

We consider a test beam with a current of 5 kA and a radius of 1 cm, the electron energy being $E = 4.5$ MeV (in which case the relativistic factor $\gamma = 10$). The degree of charge neutralization f_0 is equal to 0.1. The radius of the ion channel is set equal to 0.1 cm for a narrow channel and 1 or 5 cm for wide channels. To be specific, we assume that the radial profiles of the beam electron density and the density of the channel ions are both Gaussian.

First, we investigate the case of a narrow channel such that the characteristic radius of the electron beam is larger than the characteristic channel radius, $R_b > R_c$.

In an ion channel in which $f_1 = 0.1f_0$ and the ion density varies periodically on the spatial scale $L_{\text{per}} = 0.25\lambda_{\beta e}$, the maximum amplitude $Y_{\max}$ of oscillations of the centers of mass of the beam segments behaves in essentially the same manner as in the case of IHI in a uniform ion channel (with $f_1 = 0$ and $L_{\text{per}} = 0$). The IHI saturates at the level $Y_{\max} = 3.1$ for $X = 2\lambda_{\beta i}$ (Fig. 1, curve *1*). Here and below, the quantities $Y_{\max}$ and Y are normalized to the initial beam radius R_b. The behavior of individual beam segments is also analogous to that in a uniform ion channel. For $f_1 = 0.4f_0$, the maximum amplitude $Y_{\max}$ at the end of the beam pulse exhibits oscillatory behavior and the period of oscillations of the beam segments is longer than that in a uniform channel. The instability saturates and becomes stabilized. Simulations carried out for a larger value of the coefficient f_1, namely, $0.8f_0$, showed that, from approximately the beginning of the second half of the beam pulse ($X \geq \lambda_{\beta i}$), the maximum amplitude $Y_{\max}$ starts to oscillate between the values 3.2 and 7, in which case the instability, as before, saturates.

When the spatial period L_{per} of density variation in the ion channel is equal to $0.25\lambda_{\beta e}$ and $f_1 = 0.1f_0$, the IHI reaches saturation and the maximum amplitude $Y_{\max}$ executes small oscillations. In this case, the plot of the function is gentler than that for $R_b = R_c$ and the oscillation amplitude is smaller. A representative behavior of an individual segment is illustrated in Fig. 2b. For $f_1 = 0.4f_0$, the instability is stabilized and, at distances

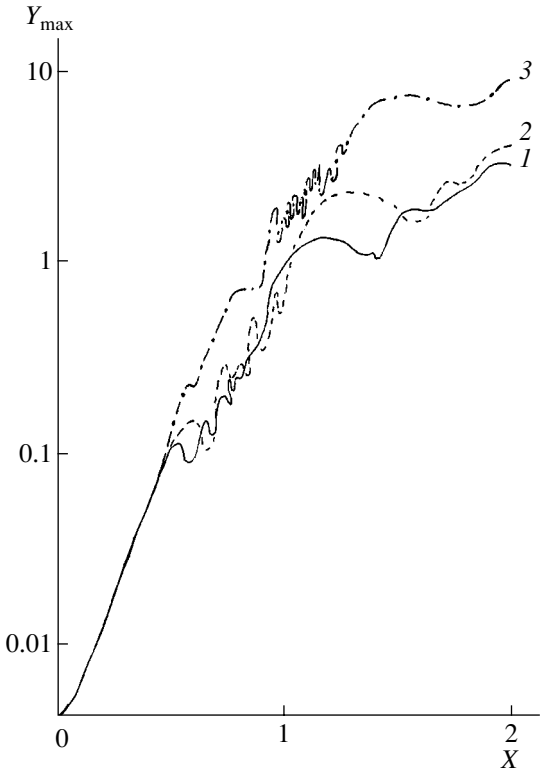


Fig. 1. Dependence of the normalized maximum beam displacement Y_{max} on the normalized beam length X for different ratios of the beam radius to the channel radius: R_b/R_c = (*1*) 10, (*2*) 1, and (*3*) 0.2.

longer than $X = 0.025\lambda_{\beta i}$, the maximum amplitude in the instability saturation stage oscillates between the values 1 and 6. A further increase in the coefficient f_1 to $0.8f_0$ leads to a detuning between the oscillations of the electron beam and the ion channel. With this value of f_1, even the oscillating segments in the leading portion of the beam deviate substantially from the axis of the beam–channel system; nevertheless, an REB can be successfully transported over the distance $z = 12.5\lambda_{\beta e}$, which is longer than the corresponding distance $z = 10\lambda_{\beta e}$ for $R_b = R_c$ and for a beam pulse length of about $0.1\lambda_{\beta i}$.

In an ion channel in which $f_1 = 0.1f_0$ and the spatial period of density variation L_{per} is equal to the betatron length $\lambda_{\beta e}$ of the beam, the behavior of individual beam segments differs from that in a uniform ion channel. That is why, during the first half of the beam pulse, the maximum amplitude Y_{max} exhibits a somewhat different behavior. Nevertheless, the overall behavior of Y_{max} is qualitatively the same as that in a uniform ion channel. The IHI is stabilized and, at the end of the beam pulse, the maximum amplitude Y_{max} saturates at a level of 3.5. For a coefficient f_1 equal to $0.4f_0$, the IHI also saturates, but the plot of the function Y_{max} during the second half of the beam pulse is of oscillatory nature. The maximum amplitude Y_{max} is equal to 1 at $X = 0.66\lambda_{\beta i}$ and reaches a value of 10 at the end of the pulse. For $f_1 = 0.8f_0$, the IHI grows in an uncontrolled manner. However, with such ion channels, REBs with pulse lengths of about $0.1\lambda_{\beta e}$ can be successfully guided over distances on the order of $z = 25\lambda_{\beta i}$.

In an ion channel in which $f_1 = 0.1f_0$ and the spatial period of density variation L_{per} is equal to $1.25\lambda_{\beta e}$, the maximum amplitude Y_{max} behaves in essentially the same manner as in a uniform ion channel. The IHI is suppressed and the maximum amplitude Y_{max} saturates at a level of 3.5 at the end of the beam pulse. In an ion channel with $f_1 = 0.4f_0$, the behavior of Y_{max} is analogous to that in a channel with $L_{per} = \lambda_{\beta e}$. The instability, as before, saturates, but during the second half of the beam pulse the maximum amplitude Y_{max} exhibits oscillatory behavior and becomes as large as 10 at $X = 2\lambda_{\beta i}$. Note that the behavior of individual beam segments is similar to that in the case of $R_b = R_c$, but the amplitude Y of oscillations of the center of mass of a segment is smaller. For example, at $X = 0.06\lambda_{\beta i}$, the amplitude Y is 0.02, while for $R_b = R_c$ it is 0.03. An increase in the coefficient f_1 to $0.8f_0$ leads to a detuning between the oscillations of the beam and ion channel, in which case, however, an REB with a pulse length of about $\sim 0.1\lambda_{\beta i}$ can be transported over a distance on the order of $26\lambda_{\beta e}$.

In an ion channel with $f_1 = 0.1f_0$ and with a longer period L_{per} of density variation ($1.5\lambda_{\beta e}$), the behavior of Y_{max} is again essentially the same as in the case of IHI in a uniform ion channel, and the value of Y in the beam segments in the front part of the beam oscillate at a higher frequency in comparison with that for $L_{per} = 1.25\lambda_{\beta e}$. For $f_1 = 0.4f_0$, the IHI reaches a saturation stage such that the function Y_{max} becomes as large as 15 at the end of the beam pulse and exhibits a pronounced oscillatory behavior during the second half of the pulse. In an ion channel in which the coefficient f_1 is increased to $0.8f_0$, ion hose oscillations grow in an uncontrolled manner; however, such ion channels also provide the possibility of transporting REBs with pulse lengths of about $0.1\lambda_{\beta i}$ over long distances (on the order of $27\lambda_{\beta e}$).

In the case of IHI in an ion channel with $f_1 = 0.1f_0$ and $L_{per} = 2\lambda_{\beta e}$, the plot of Y_{max} is again analogous to that for a uniform channel: the maximum amplitude Y_{max} reaches a value of 5 and the amplitude Y of oscillations of the beam segments between the oscillation bursts is smaller than that in a channel with $L_{per} = 1.5\lambda_{\beta e}$. In ion channels with $f_1 = 0.4f_0$–$0.8f_0$, the maximum amplitude Y_{max} behaves in a manner similar to the behavior of Y_{max} in a channel with $L_{per} = 1.5\lambda_{\beta e}$.

Calculations show that, in systems in which the characteristic radius of the electron beam is smaller than or equal to that of the ion channel ($R_b \le R_c$), the IHI possesses all of the regular features revealed above.

Figure 1 shows the dependence of the maximum beam displacement Y_{max} (normalized to the beam radius R_b) on the beam pulse length X (expressed in units of $\lambda_{\beta i}$) for $L_{per} = 0.25\lambda_{\beta e}$ and $f_1 = 0.1f_0$ and for different ratios of the beam radius to the channel radius: R_b/R_c =

10 (curve *1*), 1 (curve *2*), and 0.2 (curve *3*). We can see that, in an ion channel with such parameters, the IHI saturates and the maximum amplitude Y_{max} in a wide channel is larger than in a narrow channel. Note that this conclusion is also valid for an REB propagating in a uniform ion channel.

The effect of the dimensions of an ion channel with $L_{per} = 0.5\lambda_{\beta e}$ and $f_1 = 0.1f_0$ on the behavior of the normalized (to the beam radius R_b) amplitude Y of oscillations of the center of mass of the beam segment at $X = 0.06\lambda_{\beta i}$ is illustrated in Fig. 2, from which we can see how the instability is excited. In such a channel, the center of mass of the segment starts to oscillate earlier than in a wide channel and the maximum amplitude Y of the radial oscillations of the segment is six times smaller. The plots show clearly that the wider the channel, the longer the distance the beam segment travels between two successive pronounced oscillation bursts. It can also be seen that the spatial period of density variation in the ion channel influences the duration of oscillation bursts: the bursts become five to six times longer.

Figure 3 illustrates the behavior of the normalized (to the beam radius R_b) amplitude Y of oscillations of the center of mass of the beam segment at $X = 0.31\lambda_{\beta i}$ in ion channels whose radius is equal to the beam radius ($R_b = R_c$) and in which the spatial period of density variation L_{per} is $0.5\lambda_{\beta e}$ and the coefficient f_1 is varied between $0.1f_0$ and $0.8f_0$. We can see that the larger the coefficient, the larger the amplitude of oscillations of the centers of mass of the beam segments. Moreover, as the coefficient is increased, the maximum oscillation amplitude is seen to increase and to occur closer to the entrance to the channel.

An analysis of the results of numerical simulations allows us to draw the following conclusions:

(i) The main parameter that determines the development of IHI is the coefficient f_1. In ion channels in which the density varies periodically on a spatial scale of 0 to $2\lambda_{\beta e}$ and the coefficient f_1 is smaller than $0.4f_0$, transverse oscillations of the beam and channel segments saturate and the IHI is stabilized, in which case an REB can be guided over distances on the order of 100 betatron lengths of the beam, regardless of the value of the ratio of the beam radius to the channel radius. For larger values of the coefficient f_1, the IHI is not suppressed. However, the calculations show that here, too, a beam with a length on the order of $0.1\lambda_{\beta i}$ can be transported over a distance of up to $50\lambda_{\beta e}$ in an ion channel with the appropriate spatial period of density variation and the appropriate ratio between the transverse dimensions of the beam and the channel.

(ii) Short-period density variations ($L_{per} \le 0.3\lambda_{\beta e}$) in an ion channel have practically no effect on the dynamics of IHI development. In an ion channel in which the ion density varies periodically on a spatial scale $L_{per} \sim 0.5\lambda_{\beta e}$ or longer, the maximum amplitude of the transverse oscillations of the beam segments is markedly larger than that in the case of an REB transported in a

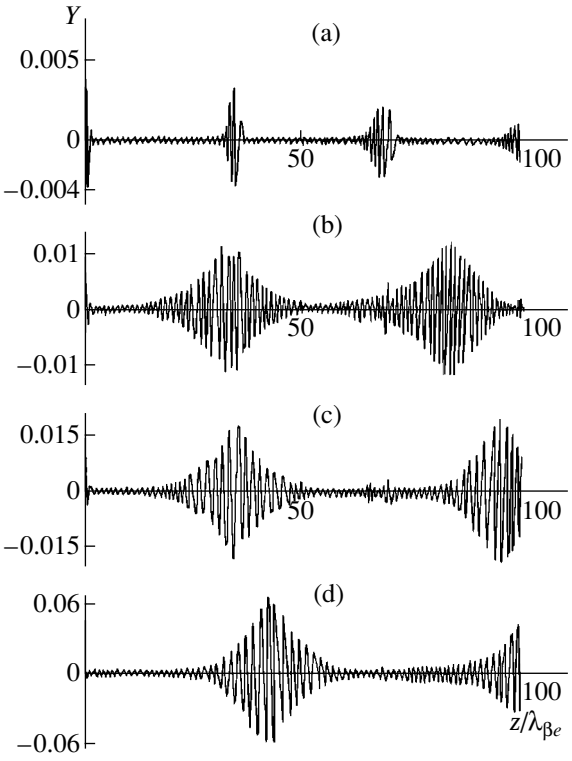


Fig. 2. Effect of the dimensions of an ion channel on the amplitude Y of oscillations of the center of mass of the beam segment at $X = 0.06\lambda_{\beta i}$ for (a) $f_1 = 0$, $L_{per} = 0$, and $R_b = R_c$; (b) $f_1 = 0.1f_0$, $L_{per} = 0.5\lambda_{\beta e}$, and $R_b = 10R_c$; (c) $f_1 = 0.1f_0$, $L_{per} = 0.5\lambda_{\beta e}$, and $R_b = R_c$; and (d) $f_1 = 0.1f_0$, $L_{per} = 0.5\lambda_{\beta e}$, and $R_b = 0.2R_c$.

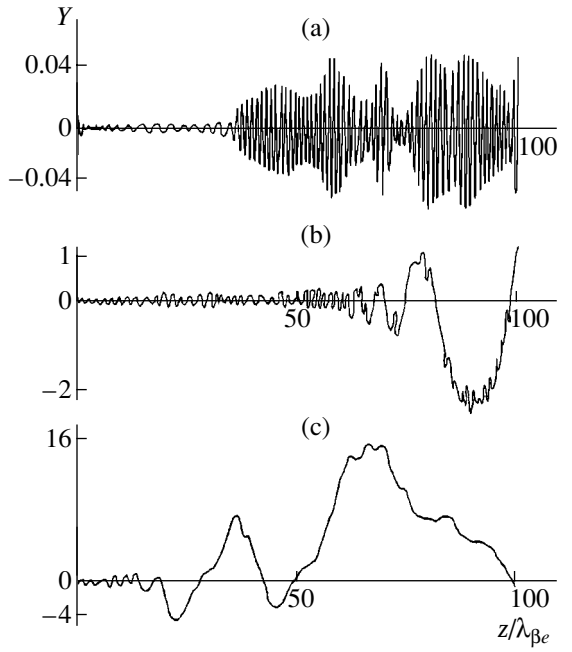


Fig. 3. Behavior of the amplitude Y of oscillations of the center of mass of the beam segment at $X = 0.31\lambda_{\beta i}$ for $L_{per} = 0.5\lambda_{\beta e}$ and $R_b = R_c$ and for three different values of the coefficient f_1: (a) $0.1f_0$, (b) $0.4f_0$, and (c) $0.8f_0$.

uniform channel. The pattern of these transverse oscillations also differs qualitatively from that in a uniform ion channel: there are no pronounced oscillation bursts.

(iii) The maximum amplitude Y_{max} of the transverse oscillations of the beam segments depends on the ratio R_b/R_c. Specifically, the smaller the ratio, the larger the amplitude Y_{max} and the longer the distance the beam segments travel between two successive pronounced oscillation bursts.

REFERENCES

1. H. L. Buchanan, Phys. Fluids **30**, 211 (1987).
2. G. J. Caporaso, A. G. Cole, and K. W. Struve, IEEE Trans. Nucl. Sci. **NS-30**, 2507 (1983).
3. R. E. Lipinsky, J. R. Smith, and I. R. Shokair, Phys. Fluids B **2**, 2764 (1990).
4. K. J. O'Brien, J. Appl. Phys. **65**, 9 (1989).
5. R. F. Fernsler, S. P. Slinker, M. Lampe, *et al.*, Phys. Plasmas **2**, 4338 (1995).
6. E. K. Kolesnikov and A. S. Manuĭlov, Radiotekh. Élektron. (Moscow) **37**, 694 (1992).
7. E. K. Kolesnikov and A. S. Manuĭlov, Zh. Tekh. Fiz. **65** (1), 165 (1995) [Tech. Phys. **40**, 90 (1995)].
8. V. B. Vladyko and Yu. V. Rudyak, Fiz. Plazmy **17**, 623 (1991) [Sov. J. Plasma Phys. **17**, 366 (1991)].
9. E. K. Kolesnikov and A. S. Manuĭlov, Zh. Tekh. Fiz. **70**, 68 (2000) [Tech. Phys. **45**, 591 (2000)].
10. J. D. Miller and R. M. Gilgenbach, Phys. Fluids **30**, 3165 (1987).
11. R. A. Bosch and R. M. Gilgenbach, Phys. Fluids **31**, 634 (1988).
12. R. A. Bosch and R. M. Gilgenbach, Phys. Fluids **31**, 2006 (1988).
13. E. P. Lee, Phys. Fluids **21**, 1327 (1978).
14. E. J. Lauer, R. J. Briggs, T. J. Fessenden, *et al.*, Phys. Fluids **21**, 1344 (1978).

Translated by O. Khadin

Technical Physics, Vol. 48, No. 12, 2003, pp. 1575–1578. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 76–80.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Twist of Growing Bacterial Colonies

E. G. Rapis

Laboratory of Applied Physics, Tel Aviv University, Ramat Aviv, Tel Aviv, 69978 Israel

Received May 20, 2003

Abstract—Twist observed in growing bacterial colonies at the macrolevel is explained in terms of the self-assembly (self-organization) of film-forming protein clusters, since the *in vitro* and *in vivo* behavior and symmetry properties of protein in an open thermodynamically nonequilibrium system are identical. The self-assembly of elastic protein films in the course of condensation in the protein–water system obeys the laws of the elasticity theory. As the viscosity of the system grows, the transition of the protein from the liquid-crystal to the solid phase occurs. This transition has a nonlinear dynamics, which also shows up at the macrolevel. Opposite vorticities (twist) appear in the system. Such a modification of protein has been named protos. It is hypothesized that the formation of an elastic nonequilibrium protos film is consistent with the behavior and orientation of elastic forces and magnetic fields in the presence of unlike electric charges. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

It has been found recently that, when the nutrient medium dries out, bacteria of growing colonies unexpectedly start spiraling in opposite directions [1, 2]. This amazing phenomenon, which is observed under a microscope, has not yet found a sound scientific description in terms of an adequate mathematical model. The questions to be tackled are as follows. What is the reason for the helical "dance" of bacteria? What forces make individual bacteria rotate simultaneously? Do bacterial cells themselves or their derivatives rotate? Earlier, Mendelson observed a similar effect, twisting motions during the formation of fibers from growing bacterial colonies, under a microscope (i.e., macroscopically) [3–5]. Later, a team of American researchers supposed that this effect is of general character, i.e., reflects the elastic properties of condensing matter, and, based on experimental data for the behavior of the fibers, worked out a dynamic mathematical model in terms of the elasticity theory. The prognostic value of this model, e.g., as applied to the appearance of Sun bursts, turned out to be higher than that of the rubber model available at that time. Yet a number of important issues remained unclear: What components of the living matter cause the processes mentioned above? What are their mechanisms? What is the reason for opposite helical motions when fibers grow in biological objects?

It is well known that growing fibers of any biological cells are the product of synthesis. In our opinion, clusters of protein molecules, rather than protein individual molecules, are synthesized; in other words, we are dealing with the self-assembly of protein films. Many such films produce twisted fibers observed at the macrolevel in experiments.

It has been recently reported that cellular organelle catalysts (Golgi apparatus, endoplasmic network, etc.) may be responsible for protein synthesis and transport. The basic components of these catalysts are also self-organized stacked membranes (films). In addition, experimental data obtained by Noji [6] suggest that gamma fibers in protein may rotate about the vertical axis. Noji observed the rotation at the microlevel and conjectured that this effect is driven by phosphorescence.

From the above, it seemingly follows that the dance of bacterial colonies is due the helical motion of protein gamma fibers during its microscopic self-organization. However, a question now arises: Why does the process, observed only at the microlevel (the rotation of gamma fibers of protein about the vertical axis), show up at the macrolevel as the helical motion of the entire system in opposite directions? To clarify the role of protein, we carried out experiments (described below) on the self-assembly of protein at the micro-, meso-, and macroscales (the classification of the scales is given elsewhere [7–12]), eliminating other components of the living matter from the reaction.

EXPERIMENTAL

We studied the condensation dynamics in water–protein colloids under equilibrium and nonequilibrium conditions. Different amounts of the samples (in bulk or as droplets) were applied *in vitro* on the same substrates (microscope slide). On some of the substrates, the system was open; on the others, closed (covered by a special glass cover). In the open system, the water evaporates much faster and the process kinetics changes compared with the closed system. Thus, at room temperature and normal atmospheric pressure,

1063-7842/03/4812-1575$24.00

the protein condensed under different kinetic and thermodynamic conditions.

The phenomenology and symmetry of the condensation product and arising flows were visualized with optical, MIN-8 polarizing, JEOL scanning electron, and confocal laser scanning (CLS) microscopes. Five series (a total of more than 25000) of experiments were carried out with 15 water-soluble proteins (egg protein, bull serum albumin, human and rabbit globulin, human hemoglobin, human crystallin, lysozyme, fibrin, etc.).

RESULTS

The basic results of the experiments are as follows. In the drying open protein–water colloid placed on a transparent hard wettable substrate, autowave processes are observed and defects forming ordered regular structures in the form of blocks or cells appear [7–12] (Fig. 1). During the dehydration, the front of three-dimensional different-colored multiple-scale fluctuations is in continuous motion, causing alternating density zones. These zones have a spiral shape and superpose on one another like sustained standing autowaves.

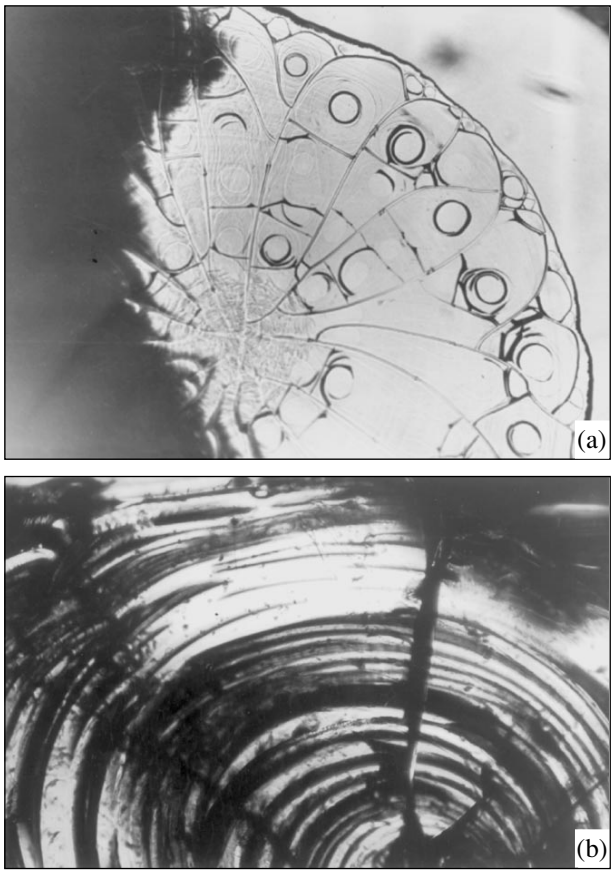


Fig. 1. (a) Division of a protein nonequilibrium film by large-scale straight and helical defects into cells with mussel-type nuclei and (b) three-dimensional opposite vorticities observed during nucleation. Optical microscope, ×200.

Then, during the *in vitro* condensation of the open water–protein colloid, liquid crystal films with various large-scale defects always appear. The amount of these defects grows with the formation of a discontinuous helical structure of mirror and chiral symmetry, followed by the spontaneous formation of clusters of cells or domains with nuclei. On both the nano- and microscale, the denser phase appears as differently colored stacked thin films. In the experiments we often observed highly anisotropic pairs (cascades) of opposite vortices producing three-dimensional conic self-complementary dendritic film structures (Figs. 1, 2) [10–12]. These films have a network of fine defects that cut their surface and produce discontinuous "bird's wing" or "porcupine" space symmetry [13, 14]. This is consistent with earlier observations [15]. The defects move in an avalanche-like manner, and bright fluorescent bending lines and films are observed in the CLS microscope (Fig. 3). As a rule, such films consist of two oppositely rotating half-moon-like branches off a mutual frame. The branches end up with a thin line (tail). The tails connect several pairs of such films, as if describing parallel structured ellipsoidal orbits with defects. The orbits resemble lines of stellar magnetic force. Branches of opposite rotation may connect to form patterns with discontinuous helical and chiral symmetry. Such behavior is reminiscent of that of magnetic forces (see, e.g., [16]) in high-temperature superconductors of the second kind that exhibit antiferromagnetic properties [17].

The results of our experiments suggest that the behavior of protos protein films is similar to the growth of bacterial colonies. In both cases, we see opposite flows of the material at the macrolevel (twist). This means that the behavior of bacterial colonies follows the unified laws of protein self-organization *in vitro* and *in vivo*. That is why, when the nutrient medium dries (i.e., the water concentration in growing bacterial colonies decreases), the sudden cooperative transition of the protein (as on the glass) from the liquid to denser phase occurs (as soon as the density of the medium reaches a critical value) and the self-assembly of the films with nonlinear random dynamics is observed.

From the above, one may assume that the process of protein self-assembling becomes more extensive, more energy-consuming, and synchronous under changing conditions. Like a typhoon in the terrestrial atmosphere, this process involves not only forming fibers but also bacterial cells, causing the dance with vorticity and mirror symmetry on the macrolevel. This is a vivid and extremely rare macrodynamic demonstration of helical mirror and chiral symmetry in the course of living protein self-organization on glass (*in vitro* and *in vivo*).

Thus, one may argue that twist in the culture of growing bacterial colonies and their cellular fibers *in vivo*, as well as in the consolidating protein film *in vitro*, is nothing more than material (protein) particles that are visualized at the meso- and macrolevels. It may be con-

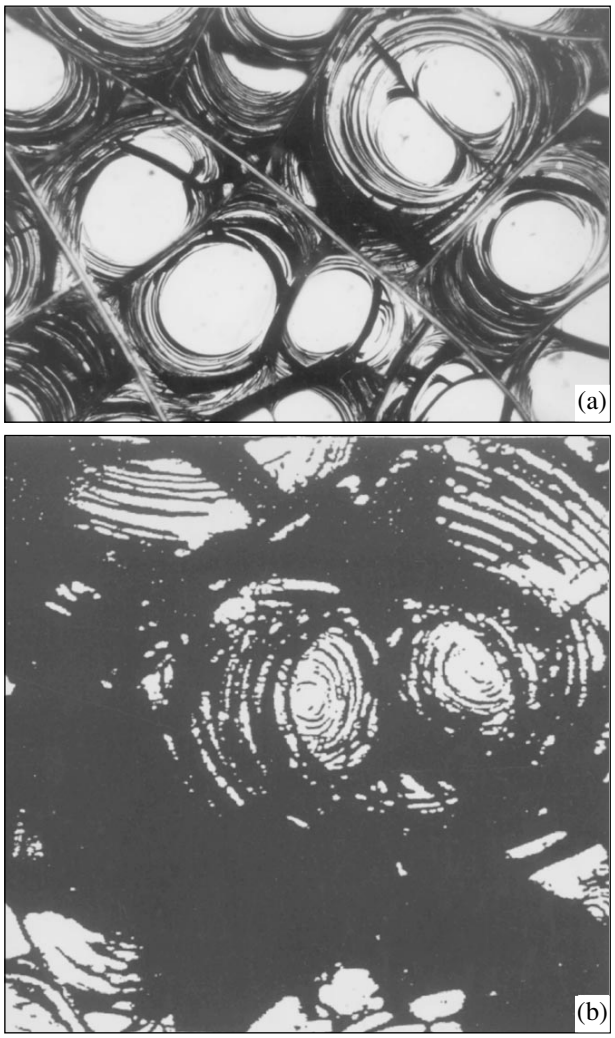


Fig. 2. (a) Further stabilization of the process: the appearance of straight defects in the cell nucleus. (b) Division of the central field with the formation of two coil-like daughter vortices *in vitro*. Optical microscope, ×140.

jectured that they move along helical lines of force that form a specularly symmetric pattern. It is known that only lines of magnetic force may be optically visualized as moving in an electric field [18]. Therefore, there is reason to suppose that the above-described visualization of material flows along elliptic lines of force is due to the presence of a magnetic field in the system provided that the polarization of the protein film is found.

Based on available experimental data for the polarization of biological systems (the presence of unlike poles in cells, protein microtubules, the protein spindle during mitosis, embryonic structures, etc.) [19–23], we can state with a great degree of certainty that the protos protein film in our experiments is also polarized. This statement is supported by autowave processes, semitransparent hemispheres moving toward each other with different velocities, left-hand and right-hand rotations, etc., occurring in the film [9–11].

Fig. 3. Fluorescence of a solid unstable protein (alternating zones of different brightness), nucleation, and conical nanostructures *in vitro*. CLS microscope, ×1 000 000.

Thus, the presence of unlike poles (hence, the possibility of polarization) may be considered to be proved. However, this fact cannot yet be substantiated quantitatively. In addition, this phenomenon (if it does exist) is hard to explain by the presence of electrolytes and magnetic dipoles in the protein colloidal system [24]. In particular, it remains unclear how dipole–dipole interactions may produce two poles in a biological object. Of interest in this respect are new data on the properties of colloidal suspensions such as a protein solution. It has been found that suspension particles in this open system behave in a nontrivial manner upon condensation. They form a long chain of like-charge attractions [25]. It is likely that the properties of a protein colloid that are observed upon condensation cause attractive unlike poles to appear. However, a general theory of formation of unlike electric poles in biological objects that is adequate for experimental conditions is still lacking.

Today, we may only hypothesize that the formation of a protein elastic film correlates with the behavior and orientation of magnetic fields due to material polarization. Here are several examples in favor of this hypothesis. First, the prognostic value of the mathematical model (in terms of the elasticity theory) that was worked out based on the behavior of twisting bacterial fibers has been proved [3–5]. Also, it has been experimentally found that protein films show a number of phenomenological properties that are characteristic of magnetic fields [9–11] and living biological objects. These are discontinuous mirror and chiral symmetry, bird's wing symmetry (also called magnetic symmetry), high anisotropy, the halving of biological cells

(their nuclei and cytoplasm) [10, 11], the reconnection of magnetic lines of force [26–29], the presence of dendritic conic structures with fractal properties; the connection of oppositely rotating branches (akin to antiferromagnetic connection), the formation of stacked films with alternating colors (Fig. 3), optical activity, and sensitivity to magnetic field [10, 11]. These space–time features are typical of systems with random nonlinear behavior (see, e.g., [29–31]).

CONCLUSIONS

Our experimental data, showing the complex space–time pattern of the self-assembly process in growing bacterial colonies at the macrolevel, may be described in terms of the behavior of nonlinear dynamic systems. The interpretation given in this work is based on the known polarization-related effects in protein: the presence of helical mirror and chiral symmetry at the meso- and macrolevel (the fact established in our previous works), sensitivity to magnetic field, and all the other analogies listed above. This suggests that the self-assembly of protein nonequilibrium films follows the general physical laws of condensation and self-organization. No doubt, our hypothesis needs quantitative verification (in particular, the relevant parameters of the system, such as elasticity, electrical conductivity, etc., should be measured). Yet even today our experiments, which shed light on the on-glass behavior of protos protein, basically clarify the reason for twist in bacterial colonies *in vitro*. However, of most importance is the fact that our experiments have made it possible to work out a simple and realistic model of dynamics of nonequilibrium protein film that provides further insight into the twist of bacteria in opposite directions upon assembling protein films not only in bacterial colonies but also in other biological systems. A quantitative mathematical model is now under development and will be described in subsequent publications.

ACKNOWLEDGMENTS

I sincerely thank Profs. M. Amus'i, E. Braudo, V. Buravtsev, V. Volkov, A. Zaikin, M. Klinger, L. Manevich, Yu. Neéman, I. Prigogine, and M. Safro for the fruitful discussions, valuable comments, and encouragement.

REFERENCES

1. E. Ben Jacob and P. Garik, Nature **343**, 23 (1990).
2. E. Ben Jacob, H. Shmueli, O. Shochet, and A. Tenenbaum, Physica A **187**, 387 (1992).
3. N. Mendelson and S. L. Keener, J. Bacteriol. **151**, 455 (1982).
4. N. Mendelson, Science **3**, 32 (1992).
5. N. Mendelson, Proc. Natl. Acad. Sci. USA **75**, 2478 (1978).
6. H. Noji, Science **28**, 1844 (1998).
7. E. Rapis and G. Gassanova, Zh. Tekh. Fiz. **61** (4), 62 (1991) [Sov. Phys. Tech. Phys. **36**, 406 (1991)].
8. E. Rapis, Pis'ma Zh. Tekh. Fiz. **21** (5), 13 (1995) [Tech. Phys. Lett. **21**, 321 (1995)].
9. E. Rapis, Pis'ma Zh. Tekh. Fiz. **23** (7), 28 (1997) [Tech. Phys. Lett. **23**, 263 (1997)].
10. E. Rapis, Zh. Tekh. Fiz. **70** (1), 122 (2000) [Tech. Phys. **45**, 121 (2000)].
11. E. Rapis, Zh. Tekh. Fiz. **71** (10), 104 (2001) [Tech. Phys. **46**, 1307 (2001)].
12. E. Rapis, in *Abstracts of the 4th Interdisciplinary Symmetry Congress ISIS, Technion, Haifa, 1998*, pp. 119–120.
13. P. Poulin, H. Stark, T. C. Lubensky, and D. A. Weitz, Science **275**, 1770 (1997).
14. B. Jorome and J. Commandeur, Nature **386**, 589 (1997).
15. J. Langmuir, J. Chem. Phys. **1**, 756 (1933).
16. J. Heagy, T. L. Carroll, and L. M. Pecora, Phys. Rev. Lett. **73**, 3528 (1994).
17. Z. Y. Zhang and M. G. Lagally, Science **276**, 377 (1997).
18. G. Boebinger, A. Passner, and J. Bevk, Sci. Am. 59 (June, 1996).
19. A. Gonzalez-Reyes, H. Elliot, and D. St Johnston, Nature **375**, 654 (1995).
20. R. Krant, W. Chia, L. Y. Jan, *et al.*, Nature **383**, 50 (1996).
21. J. Hyams, Nature **387**, 397 (1996).
22. Y. Zheng, M. L. Wong, B. Alberts, and T. J. Mitchison, Nature **378**, 578 (1995).
23. R. Heald, R. Tournebize, T. Blank, *et al.*, Nature **382**, 420 (1996).
24. G. Gilat, Chem. Phys. Lett. **121**, 9 (1985).
25. Ch. Murray, Nature **385**, 203 (1997).
26. N. H. Brummell, F. Cattaneo, and J. Toomre, Science **269**, 1370 (1993).
27. P. Allen, Nature **375**, 729 (1995).
28. D. Bishop, P. L. Gammel, D. A. Huse, and C. A. Murra, Science **255**, 165 (1992).
29. T. Matsuda, K. Harada, H. Kasai, *et al.*, Science **271**, 1393 (1996).
30. R. Ecke, Y. C. Hu, R. Mainieri, and G. Ahlers, Science **269**, 1704 (1995).
31. H. Stolum, Science **271**, 1710 (1996).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 12, 2003, pp. 1579–1582. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 81–84.

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Self-Noise of Molecular Electronic Transducers

V. A. Kozlov and M. V. Safonov

Moscow Physicotechnical Institute (State University), Institutskiĭ proezd 9, Dolgoprudnyĭ, Moscow oblast, 141700 Russia

e-mail: cme@cme.ffke.mipt.ru

Received March 11, 2003

Abstract—The effect of hydrodynamic fluctuations on noise in molecular electronic transducers is studied. It is shown that turbulent pulsation also makes a considerable contribution to the self-noise of molecular electronic transformers, along with laminar flow fluctuations. A method for qualitative and quantitative calculation of the noise induced by turbulent pulsation that arises when a liquid flows along the electrode surface is proposed. A quantitative relationship that relates the rms pressure pulsation to the liquid head and an expression for the total spectral density of the hydrodynamic noise in molecular electronic transducers are obtained.

INTRODUCTION

Molecular electronic transducers (METs) of the parameters of motion and oscillating fields have recently found wide application in wide-band and low-noise seismic sensors [1]. This is associated largely with the progress in low-noise sensor electronics, where the self-noise referred to the input has become comparable to the minimum signal level measured in seismology. Therefore, the problem of investigating and reducing MET self-noise is of current interest.

The self-noise of METs consists of charge carrier concentration fluctuations and hydrodynamic fluctuations. The latter are due to the local fluctuations of the liquid velocity. The noise related to the concentration fluctuations was investigated in [2]. It was shown that these fluctuations expressed in terms of the acceleration spectral density are much lower than 10^{-9} m/s^2/$\sqrt{\text{Hz}}$. Bearing in mind that the noise of METs is typically at least one order of magnitude higher, one may argue that hydrodynamic fluctuations make a basic contribution to the noise of METs [3]. Electrolyte velocity fluctuations arising when the laminar liquid flows along the transducer channel were studied in [4]. It was shown [4] that, when expressed in terms of acceleration, the spectral density of noise of this is frequency independent:

$$\overline{\delta a_\text{f}^2} = \frac{2k_\text{B}TR_\text{h}}{\rho^2 l^2}, \qquad (1)$$

where R_h is the hydrodynamic impedance in the system, ρ is the electrolyte density, l is the length of the transducer channel, T is the absolute temperature, and k_B is Boltzmann's constant.

For $R_\text{h} = 5 \times 10^8$ (N s)/m^5 and $l = 5 \times 10^{-2}$ m, which are typical of present-day molecular electronic seismometers, the noise associated with laminar fluctuations is 3×10^{-8} m/s/$\sqrt{\text{Hz}}$. It is evident that the noise due to concentration fluctuations may be neglected.

From formula (1), it follows that a decrease in the hydrodynamic impedance R_h or an increase in the transducer channel length l suppresses hydrodynamic fluctuations. This statement was checked in experiments with prototype transducers. It was found that the self-noise of the transducers depends on the above parameters in a more complex manner than predicted by formula (1). In particular, if R_h decreases by a factor of 2000 and l increases threefold, the noise decreases by approximately 20 dB rather than by 40 dB as could be expected. It was also found that the noise increases with the external signal amplitude. Thus, the experimental data suggest that hydrodynamic noise of a type other than that considered in [4] also makes a considerable contribution to the self-noise spectrum of the transducer. Such may be noise due to the vortical pulsation of both pressure and local velocities. The pulsation may be associated with fluctuations that arise when the electrolyte flows over the metal mesh electrodes of the transducer (even for not too large Reynolds numbers). In this study, we investigate noise of this kind from the theoretical standpoint.

MATHEMATICAL STATEMENT OF THE PROBLEM

To investigate vortical pulsation in a molecular electronic transducer, we will first find a relationship between the pressure and local velocity in the near-electrode boundary layer. Since this problem is hard to solve for a real system, we consider a simple model of the boundary layer.

Suppose that an ideal liquid strikes an infinite impermeable plane. At some distance z_0 from the plane, the velocity of the liquid has only the component nor-

mal to the surface. In the cylindrical coordinate system, the normal component has the form

$$v_z(z_0, r) = -U \quad \text{at} \quad 0 < r < R,$$
$$v_z(z_0, r) = 0 \quad \text{at} \quad r \geq R. \tag{2}$$

Here, the z axis is perpendicular to the plane and is directed outward. Within the domain $\{r < R, z < z_0\}$, the distributions of the liquid velocity and pressure have a number of properties typical of the boundary layer (in particular, the derivative $\partial v_r/\partial z$ is significant). In our model, the quantity R is the characteristic size of the real electrode, the plane $z = 0$ (at $r < R$) is the electrode surface, and z_0 is the upper bound of the layer on the electrode surface. This model makes it possible to easily evaluate the potential $\varphi(z, r)$ and velocity distribution. In the polar coordinate system, the equation for potential has the form

$$\frac{\partial^2\varphi}{\partial r^2} + \frac{1}{r}\frac{\partial\varphi}{\partial r} + \frac{\partial^2\varphi}{\partial z^2} = 0. \tag{3}$$

Passing from the spatial variable r to the variable s by means of the Hankel transformation and taking into account boundary conditions (2), we arrived at the following expression for the potential:

$$\varphi(z, r) = -UR\frac{J_1(Rs)\cosh(zs)}{s^2\sinh(z_0 s)}, \tag{4}$$

where $J_1(Rs)$ is the Bessel function of the first kind.

Using the relationships $v_z = \partial\varphi/\partial z$ and $v_r = \partial\varphi/\partial r$, we can now find the components of the liquid velocity. In the boundary layer region of interest ($r < R$, $z < z_0$), the pressure is given by

$$\frac{p(z, r)}{\rho} = -UR\int_0^\infty \frac{1 - \cosh(zs)J_0(rs)J_1(Rs)}{\sinh(z_0 s)s}ds$$
$$-\frac{U^2R^2}{2}\left\{\left[\int_0^\infty \frac{\sinh(zs)}{\sinh(z_0 s)}J_1(Rs)J_0(rs)ds\right]^2\right. \tag{5}$$
$$\left.+\left[\int_0^\infty \frac{\cosh(zs)}{\cosh(z_0 s)}J_1(Rs)J_1(rs)ds\right]^2\right\}.$$

Since the integrals entering into (5) cannot be taken exactly, we will consider approximate estimates. The function $J_1(Rs)$ in the integrands of (5) is the fastest oscillating function; therefore, $s \leq 1/R$ make a dominant contribution to the integrals. Then, assuming that $z < z_0 \leq R$, the hyperbolic functions in (5) can be expanded into the Taylor series up to the terms of the least order of smallness in z/R and z_0/R. Having calculated the resulting integrals, we obtain

$$\frac{p(z, r)}{\rho} = -\frac{\dot{U}}{2z_0}\left(\frac{r^2}{2} - z^2\right) - \frac{U^2}{2z_0^2}\left(\frac{r^2}{2} + z^2\right). \tag{6}$$

The pressure on the electrode surface (averaged over its area) is given by

$$p = -\frac{\rho R^2}{8z_0^2}\left(\dot{U}z_0 + \frac{U^2}{2}\right), \tag{7}$$

where $z_0 \equiv \delta$.

By varying (7), it is easy to find a relation between the spectral density of pulsation of the electrode-area-averaged pressure $\overline{\delta p_\omega^2}$ and the spectral density of pulsation of the velocity $\overline{\delta U_\omega^2}$:

$$\overline{\delta p_\omega^2} = \frac{\rho^2}{64}\frac{R^4}{\delta^4}(\omega^2\delta^2 + U^2)\overline{\delta U_\omega^2}. \tag{8}$$

The presence of two terms in (8) indicates that the pressure pulsation depends on both the flow acceleration and the liquid head variation. From (8), the acceleration-to-head contribution ratio is as follows:

$$\frac{\delta^2\omega^2}{U^2} = \frac{\delta^2}{\lambda^2}. \tag{9}$$

Here, $\lambda = U/\omega$ (where ω is the pulsation frequency) is the characteristic size of a vortex resulting from the velocity pulsation of the liquid.

PRESSURE PULSATION SPECTRUM

As is seen from (8), the pressure pulsation spectrum is defined by the spectrum of velocity pulsation. Suppose that the velocity pulsation in our system is described by the Kolmogorov spectrum [5]. By analogy with [6], we consider the velocity correlator in the form

$$\overline{\delta U_f^2} = \frac{4\sqrt[3]{2}}{27(2\pi)^{\frac{8}{3}}\kappa^{\frac{2}{3}}}\varepsilon^{\frac{2}{3}}U^{\frac{2}{3}}f^{-\frac{5}{3}}, \tag{10}$$

where f is the pulsation frequency, ε is the quantity characterizing the rate of energy dissipation over various scales of turbulence (its value is equal to the mean rate of kinetic energy dissipation per unit mass of the liquid), and κ is a dimensionless constant close to unity.

Similarly to [7], we assume that the velocity fluctuation spectrum has the form of (10) down to the cutoff frequency

$$f_0 \sim \frac{U}{\delta} \tag{11}$$

and then (at frequencies below f_0) tends to a constant. For characteristic liquid velocities in the transducer channel and thicknesses of the boundary layer, the

value of f_0 is roughly equal to 100 Hz. Since frequencies below 100 Hz are of primary interest for seismology, the velocity pulsation spectrum is assumed to be independent of frequency:

$$\overline{\delta U_f^2} = \frac{4\sqrt[3]{2}}{27(2\pi)^{\frac{8}{3}}\kappa^{\frac{2}{3}}}\varepsilon^{\frac{2}{3}}U^{\frac{2}{3}}f_0^{-\frac{5}{3}}. \quad (12)$$

For frequencies $f < f_0$, the vortex size $\lambda = U/2\pi f$ exceeds the thickness δ of the boundary layer; therefore, the term containing ω^2 in (8) may be omitted. For typical values of the MET parameters, the values of δ/R obtained experimentally are equal to 0.25–0.40,[1] where R is the radius of the wire of which the electrode grid is made. The smallness of the ratio δ/R substantiates the use of the two-dimensional model of the boundary layer in a real electrode system.

The rate ε of energy dissipation can be evaluated by using the analogy between the motion of a liquid in a cylindrical channel and the flow of electric current in a conductor. In our case, the hydrodynamic impedance R_h plays the role of ohmic resistance:

$$\varepsilon = \frac{U_0^2 S_0^2 R_h}{m} = \frac{U_0^2 S_0 R_h}{\rho l}, \quad (13)$$

where U_0 is the mean velocity of the electrolyte in the transducer channel, S_0 is the cross-sectional area of the channel, and l is the length of the channel.

Here,

$$U_0 = \frac{\rho l a}{S_0 R_h}, \quad (14)$$

where a is the effective acceleration of gravity near the Earth's surface at the point of observation.

The velocity U_0 is related to the characteristic liquid velocity U near the surface of the metal mesh electrode by the relationship

$$\frac{U}{U_0} \approx \frac{S_0}{S}, \quad (15)$$

where S is the total area of holes in the metal mesh electrode.

Thus, for the spectral density of the rms pressure pulsation at frequencies below the cutoff frequency f_0, we have

$$\overline{\delta p_f^2} \approx \frac{\sqrt[3]{2}}{108(2\pi)^{\frac{8}{3}}}\left(\frac{R}{\delta}\right)^4\left(\frac{S^2 R_h}{\rho S_0 l\kappa}\right)^{\frac{2}{3}}\left(\frac{\delta}{U}\right)^{\frac{5}{3}}\left(\frac{\rho U^2}{2}\right)^2. \quad (16)$$

[1] I.S. Zakharov, private communication.

As is customary in hydroacoustics [7], we recast formula (16) into the form

$$\sqrt{\overline{\delta p^2}} = \alpha\frac{\rho U^2}{2}, \quad (17)$$

where α is the factor depending on the properties of the flow and a body in stream.

Experimental values of α found from hydroacoustic measurements lie between 1.9×10^{-3} and 5.7×10^{-3} [7, 8]. Integrating the spectral density of pressure pulsation over the entire frequency range yields the rms pressure

$$\sqrt{\overline{\delta p^2}} = 0.014\left(\frac{R}{\delta}\right)^2\left(\frac{S^2 R_h}{\rho S_0 l\kappa}\right)^{\frac{1}{3}}\left(\frac{\delta}{U}\right)^{\frac{5}{6}}\frac{\rho U^2}{2}; \quad (18)$$

hence,

$$\alpha = 0.014\left(\frac{R}{\delta}\right)^2\left(\frac{S^2 R_h}{\rho S_0 l\kappa}\right)^{\frac{1}{3}}\left(\frac{\delta}{U}\right)^{\frac{5}{6}}. \quad (19)$$

Substituting typical values of the system's parameters into (19), we find that $\alpha = 5.5 \times 10^{-3}$, which agrees with the above experimental data.

Next, from the spectral density of turbulent pressure pulsation, we derive the spectral density of turbulent noise (expressed in terms of acceleration):

$$\overline{\delta a_f^2} = \overline{\delta p_f^2}/\rho^2 l^2. \quad (20)$$

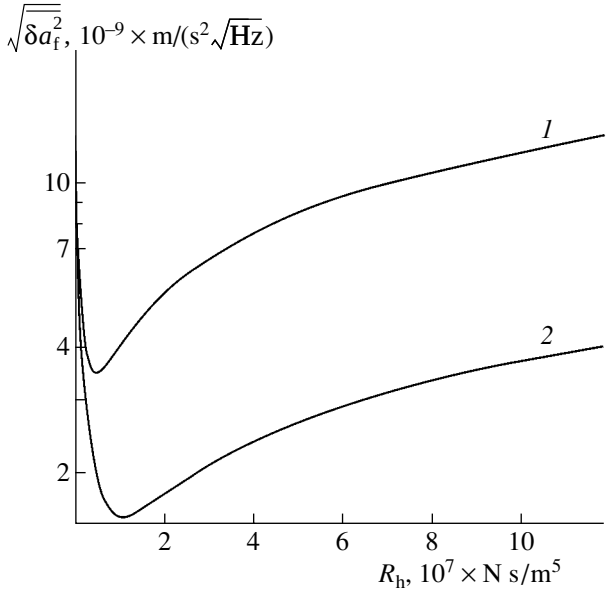
With allowance for laminar fluctuation noise (1), we arrive at the final formula for the spectral density of total hydrodynamic noise (in terms of acceleration) in METs at frequencies below $f_0 = 100$ Hz:

$$\overline{\delta a_f^2} = \frac{\sqrt[3]{2}}{432(2\pi)^{\frac{8}{3}}}\left(\frac{R}{\delta}\right)^4\left(\frac{Sa}{S_0}\right)^{\frac{2}{3}}\left(\frac{\delta\rho a}{S}\right)^{\frac{5}{3}}\frac{1}{l^{\frac{1}{3}}R_h^{\frac{5}{3}}} + \frac{2k_B T R_h}{\rho^2 l^2}. \quad (21)$$

From (21), it follows that the spectral density of the noise (expressed in acceleration) does decrease monotonically with increasing length l of the transducer channel but more slowly than could be expected from formula (1). This is consistent with the experimental data. For a fixed length l, the spectral density as a function of R_h has a minimum. The hydrodynamic impedance R_{h0} at which the noise is minimum varies as

$$R_{h0} \sim l^{\frac{5}{8}}. \quad (22)$$

For the channel lengths $l = 0.05$ and 0.16 m, $\sqrt{\overline{\delta a_f^2}}$ as a function of R_h is plotted in the figure. With these values, the optimum values of the hydrodynamic impedance are 5×10^6 and 1×10^7 (N s)/m^5, respectively.



Spectral density of electrolyte acceleration fluctuations in the channel of a molecular electronic transducer vs. hydrodynamic impedance. l = (*1*) 0.05 and (*2*) 0.16 m.

CONCLUSIONS

The investigation of self-noise in METs that is performed in this study enables one to conclude that turbulent fluctuations contribute considerably to the total hydrodynamic noise beginning from a level of $(1.5–3.0) \times 10^{-8}$ m/s$^2/\sqrt{\text{Hz}}$, while the spectral density of total noise remains frequency independent down to frequencies on the order of 100 Hz. It is important that this noise increases with signal amplitude. This circumstance is significant when molecular electronic seismic detectors are used in observatories where the natural seismic background is low. The quantitative estimate of the factor α, which relates the rms pressure pulsation to the liquid head, agrees well with hydroacoustic measurements. Thus, it is hoped that noise of such a kind may be separated out from the total noise of the system. Of basic importance is the expression for the spectral density of the MET self-noise as a function of the physical and geometrical parameters of the transducer. This expression makes it possible to optimize the parameters of the transducer in terms of noise minimization. Since the noise as a function of the hydrodynamic impedance of the transducer has a pronounced minimum (at a fixed length of the transducer channel), the accuracy of evaluating the hydrodynamic impedance becomes of crucial importance.

REFERENCES

1. I. A. Abramovich, V. M. Agafonov, S. K. Daragan, *et al.*, Seismicheskie Prib. **31**, 56 (1999).
2. A. Yu. Antokhin and V. A. Kozlov, Elektrokhimiya **25**, 1631 (1989).
3. A. V. Kharlamov, Candidate's Dissertation (Moscow, 2001).
4. V. A. Kozlov and K. A. Sakharov, *Fundamentals of Liquid- and Solid-State Measuring Systems and Data Processing Devices* (Mosk. Fiz.-Tech. Inst., Moscow, 1994), pp. 37–42.
5. A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics: Theory of Turbulence* (Gidrometeoizdat, St. Petersburg, 1992; MIT, Cambridge, 1975).
6. V. I. Blokhintsev, *Acoustics of Inhomogeneous Moving Medium* (Nauka, Moscow, 1981).
7. G. P. Haddle and E. J. Skudrzyk, J. Acoust. Soc. Am. **46**, 130 (1969).
8. V. S. Petrovskiĭ, *Fluid Mechanics of Turbulent Noise* (Sudostroenie, Moscow, 1966).

Translated by Yu. Vishnyakov

Technical Physics, Vol. 48, No. 12, 2003, pp. 1583–1585. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 85–86.

SHORT COMMUNICATIONS

Stability Analysis of a Spherical Electrostatic Suspension in a Liquid in the Induction Approximation

A. V. Limonov and V. A. Semenov

Perm State University, ul. Bukireva 15, Perm, 614600 Russia

e-mail: semenov@psu.ru

Received May 14, 2003

Abstract—In [1] the image force was shown to impose additional conditions for the electrostatic suspension of a sphere without dynamic control of the electrode potential, and the dependence of the critical voltage between the electrodes on the sphere radius was derived experimentally. In this work, this dependence is found analytically by calculating electrical forces in the third-order approximation in shift of the sphere from equilibrium. © *2003 MAIK "Nauka/Interperiodica".*

Let a potential distribution $U_0 f(\theta)$ be given on the surface of a spherical cavity filled with a dielectric fluid with a permittivity ε_m. The distribution $U_0 f(\theta)$ is such that the field inside the cavity is centrosymmetric (here, U_0 is the characteristic potential drop across the electrodes). The origin of the spherical coordinate system (r, θ, φ) is placed at the center of the cavity. The angle θ is measured relative to the polar axis z.

Let a sphere of radius R and permittivity ε_p be placed at the center of the cavity. The potentials u_m and u_p in the fluid and sphere, respectively, satisfy the set of equations

$$\Delta u_m = 0, \quad \Delta u_p = 0,$$
$$r = 1, \quad u_m = u_p, \quad \varepsilon\frac{\partial u_m}{\partial n} = \frac{\partial u_p}{\partial n}, \tag{1}$$
$$r = c, \quad u_m = f(\theta),$$

where n is the outer normal to an element ds of the spherical surface and Δ is the Laplacian. Also, $r \longrightarrow R$, $u \longrightarrow U_0$, $c = b/R$, and $\varepsilon = \varepsilon_m/\varepsilon_p$.

By virtue of the field symmetry, the sphere is in equilibrium. Let us find the electrical force acting on the sphere when it shifts from equilibrium along and normal to the symmetry axis.

When the sphere shifts along the symmetry axis z by $\delta_z \ll 1$, the changes in the potentials u_m and u_p have the same order of magnitude as δ_z. A solution to problem (1) is represented as the third-order expansion in the shift δ_z:

$$u_m = \sum_{n=0}^{\infty}\left[\left(\frac{r}{c}\right)^n (a_n^{(0)} + a_n^{(1)}\delta_z + a_n^{(2)}\delta_z^2 + a_n^{(3)}\delta_z^3)\right.$$
$$\left. + r^{-(n+1)}(b_n^{(0)} + b_n^{(1)}\delta_z + b_n^{(2)}\delta_z^2 + b_n^{(3)}\delta_z^3)\right]P_n(\cos\theta),$$

$$u_p = \sum_{n=0}^{\infty} r^n(d_n^{(0)} + d_n^{(1)}\delta_z + d_n^{(2)}\delta_z^2 + d_n^{(3)}\delta_z^3)P_n(\cos\theta), \tag{2}$$

$$a_n^{(0)} = \frac{2n+1}{2\left(1 + \dfrac{(\varepsilon-1)n}{(n+1)\varepsilon + n}c^{-(2n+1)}\right)} \times \int_0^{\pi} f(\theta)P_n(\cos\theta)\sin\theta d\theta.$$

Here, $P_n(\cos\theta)$ are Legendre polynomials. The coefficients $a_n^{(1)}$, $a_n^{(2)}$, and $a_n^{(3)}$ are calculated from the condition that the potential on the cavity surface is undisturbed:

$$a_n^{(i)} + c^{-(n+1)}b_n^{(i)} = 0, \quad i = 1, 2, 3. \tag{3}$$

When the sphere is displaced normally to the symmetry axis (i.e., along the x axis), we use the coordinate system (r, θ', φ') where the angle θ' is counted relative to the x axis and the angle φ', relative to the z axis in the yz plane. The potential distribution function in the new coordinate system is designated as $f(\theta', \varphi')$. Note that $\cos\theta = -\sin\theta'\cos\theta'$ in this case.

When the sphere shifts along the x axis by $\delta_x \ll 1$, a solution to (1) is represented in the form

$$u_m = \sum_{n=0}^{\infty}\sum_{k=0}^{n}\left(\left(\left(\frac{r}{c}\right)^n (\alpha_{n,k}^{(0)} + \alpha_{n,k}^{(1)}\delta_x + \alpha_{n,k}^{(2)}\delta_x^2 + \alpha_{n,k}^{(3)}\delta_x^3)\right.\right.$$
$$\left.\left. + r^{-(n+1)}(\beta_{n,k}^{(0)} + \beta_{n,k}^{(1)}\delta_x + \beta_{n,k}^{(2)}\delta_x^2 + \beta_{n,k}^{(3)}\delta_x^3)\right)\cos k\varphi'$$

1063-7842/03/4812-1583$24.00

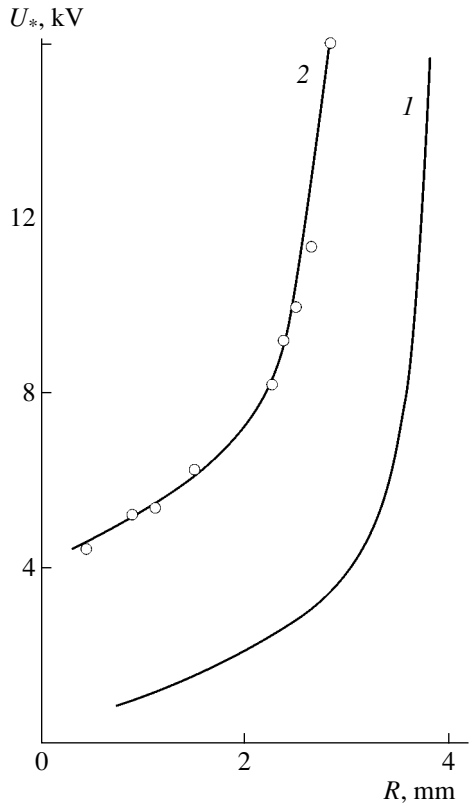


Figure.

$$+\left(\left(\frac{r}{c}\right)^n(\gamma_{n,k}^{(0)}+\gamma_{n,k}^{(1)}\delta_x+\gamma_{n,k}^{(2)}\delta_x^2+\gamma_{n,k}^{(3)}\delta_x^3)\right.$$

$$\left.+r^{-(n+1)}(\chi_{n,k}^{(0)}+\chi_{n,k}^{(1)}\delta_x+\chi_{n,k}^{(2)}\delta_x^2+\chi_{n,k}^{(3)}\delta_x^3)\right)$$

$$\times\sin k\varphi')\Big)P_n^k(\cos\theta'),$$

$$u_p=\sum_{n=0}^{\infty}\sum_{k=0}^{n}(r^n((\lambda_{n,k}^{(0)}+\lambda_{n,k}^{(1)}\delta_x+\lambda_{n,k}^{(2)}\delta_x^2+\lambda_{n,k}^{(3)}\delta_x^3)\cos k\varphi' \tag{4}$$

$$+(\mu_{n,k}^{(0)}+\mu_{n,k}^{(1)}\delta_x+\mu_{n,k}^{(2)}\delta_x^2+\mu_{n,k}^{(3)}\delta_x^3)\sin k\varphi')P_n^k(\cos\theta')),$$

$$\alpha_{n,k}^{(0)}=\frac{\dfrac{(2n+1)(n-k)!}{2\pi\omega_k(n+k)!}}{1+\dfrac{(\varepsilon-1)nc^{-(2n+1)}}{\varepsilon(n+1)+n}}$$

$$\times\int_0^{2\pi}\int_0^{\pi}f(\theta',\varphi')P_n^k(\cos\theta')\cos k\varphi'\sin\theta'd\theta'd\varphi',$$

$$k=0:\ \omega_k=2,\quad k\neq 0:\ \omega=1.$$

Here, $P_n^k(\cos\theta')$ are associated Legendre polynomials.

Solving problem (1) by expanding in shift of the sphere, we find the coefficients in (2) and (4) under the condition that the potential disturbance on the cavity surface vanishes.

It is known [2] that an external electrostatic field acts on a body with a force given by the formula

$$\mathbf{F}=\frac{\varepsilon_m}{4\pi}\oint\left[\mathbf{E}(\mathbf{n}\cdot\mathbf{E})-\frac{1}{2}E^2\mathbf{n}\right]ds. \tag{5}$$

Substituting the potential u_m on the spherical surface from (2) and (4) into (5) yields expressions for the forces acting on the sphere shifted along and normally to the symmetry axis of the field:

$$F_z=F_z^{(1)}\delta_z+F_z^{(3)}\delta_z^3+\ldots,$$

$$F_x=F_x^{(1)}\delta_x+F_x^{(3)}\delta_x^3+\ldots,$$

$$F_z^{(1)}=-2\sum_{n=0}^{\infty}(n+1)c^{-(n+1)}(a_{n+1}^{(0)}b_n^{(1)}+a_{n+1}^{(1)}b_n^{(0)}),$$

$$F_z^{(3)}=-2\sum_{n=0}^{\infty}(n+1)c^{-(n+1)}$$

$$\times(a_{n+1}^{(0)}b_n^{(3)}+a_{n+1}^{(1)}b_n^{(2)}+a_{n+1}^{(2)}b_n^{(1)}+a_{n+1}^{(3)}b_n^{(0)}),$$

$$F_x^{(1)}=-2\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{(n+1+k)!}{(n-k)!}c^{-(n+1)} \tag{6}$$

$$\times(\alpha_{n+1}^{(0)}\beta_n^{(1)}+\alpha_{n+1}^{(1)}\beta_n^{(0)}+\gamma_{n+1}^{(0)}\chi_n^{(1)}+\gamma_{n+1}^{(1)}\chi_n^{(0)}),$$

$$F_x^{(3)}=-2\sum_{n=0}^{\infty}\sum_{k=0}^{n}\frac{(n+1+k)!}{(n-k)!}c^{-(n+1)}$$

$$\times(\alpha_{n+1}^{(0)}\beta_n^{(3)}+\alpha_{n+1}^{(1)}\beta_n^{(2)}+\alpha_{n+1}^{(2)}\beta_n^{(1)}+\alpha_{n+1}^{(3)}\beta_n^{(0)}$$
$$+\gamma_{n+1}^{(0)}\chi_n^{(3)}+\gamma_{n+1}^{(1)}\chi_n^{(2)}+\gamma_{n+1}^{(2)}\chi_n^{(1)}+\gamma_{n+1}^{(3)}\chi_n^{(0)}).$$

Estimating from (6) the shifts δ_z^* and δ_x^* for which the forces acting on the sphere along and normally to the symmetry axis of the field reach maximal values, we find expressions for the maximal forces:

$$F_z^*\approx F_z^{(1)}\left(-\frac{F_z^{(1)}}{3F_z^{(3)}}\right)^{\frac{1}{2}},$$

$$F_x^*\approx F_x^{(1)}\left(-\frac{F_x^{(1)}}{3F_x^{(3)}}\right)^{\frac{1}{2}}. \tag{7}$$

Equating (7) to an external body force acting on the sphere, one may estimate the minimal (critical) voltage U_* across the electrodes that provides a balance of forces. The figure shows the analytical dependence (curve *1*) of the critical voltage on the sphere radius in the gravitational field. The parameters of the experiment were the same as in [1], and the potential distribution function on the cavity surface was taken to be

$$f(\theta) = \frac{1}{2}(3\cos^2\theta - 1) = P_2(\cos\theta).$$

The experimental (curve *2*) and calculated results are seen to be in qualitative agreement.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (grant no. 01-01-00512).

REFERENCES

1. V. A. Semenov, Zh. Tekh. Fiz. **57**, 2056 (1987) [Sov. Phys. Tech. Phys. **32**, 1244 (1987)].
2. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8*: Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 12, 2003, pp. 1586–1589. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 87–90.

SHORT COMMUNICATIONS

Dynamics of Energy Release in a Low-Voltage Pulsed Arc in Air

A. N. Dovbnya*, K. V. Korytchenko, M. A. Krasnogolovets*, Yu. Ya. Volkolupov*, Yu. D. Tur*, S. M. Shkirida*, and A. I. Kosoĭ***

** Uskoritel' Research Complex, Kharkov, 61108 Ukraine*

e-mail: dovbnya@nik.kharkov.ua

*** Kharkov Institute of Armoured Forces, Kharkov, 61034 Ukraine*

e-mail: entropia@rambler.ru

Received February 5, 2003; in final form, May 27, 2003

Abstract—The influence of the discharge conditions on the dynamics of energy release is considered for pulsed arcs in air at initial pressures from 10^5 to 8×10^5 Pa and a low-voltage capacitor voltage of up to 400 V. A novel method for determining the resistance of the discharge channel in the final stage of a spark discharge is proposed. The method is applied to estimating the discharge channel parameters. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

Thermodynamic equilibrium that is achieved in high-pressure steady-state arcs provides optimum conditions for the Joule heating of the gas medium. Obviously, studying the transition from a spark to an arc discharge assists in solving the problem of rapidly and efficiently heating gas media. The short duration of the transient process makes it possible to produce a plasma–wave system for the formation of intense shock waves, which, in turn, will allow one to create new technological devices, such as gasdynamic pulsed detonation-combustion lasers and direct-flow pulsed air propulsion engines [1].

PROBLEMS SOLVED IN STUDYING PULSED ARCS IN GASES

The transition from a spark to an arc is accompanied by a change in the character of collisions leading to gas ionization in the discharge channel. In the initial stage of the transition, electron-impact ionization of neutral molecules is dominant, which is followed by step ionization and then ionization via collisions of excited molecules with neutral ones. The latter type of ionization takes place when the gas temperature becomes sufficiently high. By step ionization, we mean the ionization of molecules via subsequently passing through excited states. The ionization processes are accompanied by a change in the electric field strength in the discharge channel.

Hence, by setting the electrode voltage (which only slightly differs from the potential drop across the positive column of an equilibrium arc), it is possible to generate an arc discharge.

It should be noted that a fairly high degree of ionization is achieved in the spark channel. Nevertheless, a necessary condition for an electric discharge to occur is that the threshold voltage should be maintained at the electrodes. This voltage depends, in particular, on the gas temperature in the discharge channel. It is believed that, at a low electrode voltage, the electric field strength is insufficient to provide electron-impact ionization of the gas molecules in the channel. If energy release in a spark discharge is low and, accordingly, the gas kinetic temperature in the channel is also low, then the threshold voltage appears to be higher than the electrode voltage corresponding to a steady-state arc discharge. If the discharge current continues to flow after the spark discharge stage is finished, while the electrode voltage is maintained at a level close to the potential drop across the positive column of an equilibrium arc, then we can conclude that a high gas kinetic temperature in the discharge channel has been reached in the final stage of the spark. Having determined the resistance of the spark channel in the final stage of its evolution in a given discharge gap, one can analyze to what extent the gas parameters in the discharge channel at this time correspond to the gas parameters in a steady-state arc.

In view of the above, the problem arises of determining the resistance of the discharge channel by estimating the threshold voltage at which the current continues to flow in a given electric circuit at a fixed energy released in a spark discharge. As applied to plasma–wave systems, this parameter should be estimated for the conditions under which the energy deposited in the spark discharge is lower than 1 J per 1 cm of the gap length. In addition, it is necessary to determine how the period and dynamics of energy release in a pulsed arc depend on the charging voltage at a low-voltage capacitor, the interelectrode distance, and the initial gas pressure in the discharge gap. The determination of these parameters makes it possible to estimate the fields of

1063-7842/03/4812-1586$24.00

application of pulsed arcs in technological devices from the standpoint of the required rate of energy release.

EXPERIMENT

1. Characteristics of the Experimental Device and Range of Investigation

The electric circuit of the experimental device is shown in Fig. 1. This circuit allows the amplitude of the high-voltage pulse at the secondary winding of transformer T to be varied in the range up to 26 kV. The charge voltage at low-voltage capacitor C_1 was varied in the range 0–400 V. The transformation ratio of pulsed transformer T was 1 : 2. The transformer core was made of an ET 3424 electrical steel with a cross section of 7.5 cm^2 and an average length of the magnetic field line of 0.44 m. As high-voltage capacitor C_2, we used KVI-3 capacitors with a total capacitance of 6×680 pF. The maximum value of the energy deposited in the spark discharge was lower than 0.7 J. Since capacitor C_2 was included in a circuit with a completely discharging capacitive storage and the transformer efficiency was nearly 90%, the energy deposited in the spark discharge could be determined with a fairly good accuracy. Our experiments were carried out for discharges in air at an initial temperature of 293 K and gas pressures in the range 10^5 to 8×10^5 Pa. The interelectrode distances were 0.9, 1.9, and 3 mm. The electrodes were made of Kh18N10T steel. In order to take into account the effect of the electric field inhomogeneity, we have measured the static breakdown voltages at atmospheric conditions for the above interelectrode distances. These voltages were found to be 3, 5, and 9 kV, respectively. At longer distances, discharges were not ignited because of the limited duration of the high-voltage pulse [2].

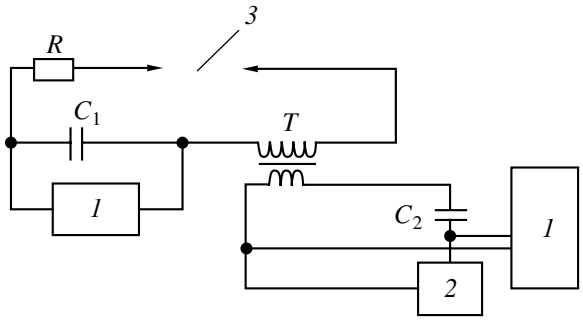


Fig. 1. Diagram of the pulsed-arc circuit: (*1*) charging device, (*2*) switch, R measuring shunt, and (*3*) discharge gap.

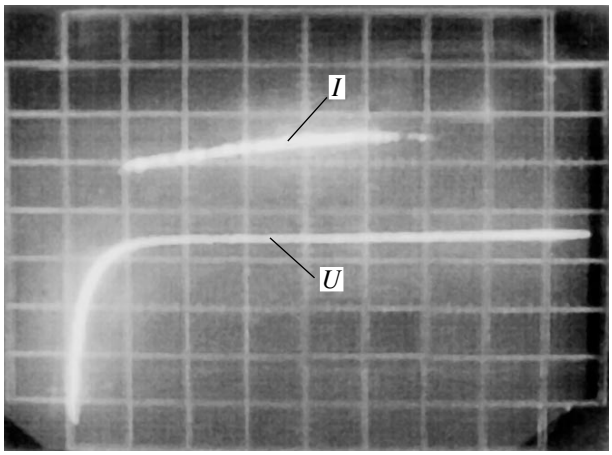


Fig. 2. Current and voltage oscillograms. The sweep speed is 50 μs/division, the current scale is 5.4 A/division, and the voltage scale is 500 V/division.

2. Estimate of the Resistance of the Discharge Channel in the Final Stage of Spark Evolution

The resistance of the discharge channel was estimated as follows. Spark discharges were produced across a fixed discharge gap. Varying the charging voltage at low-voltage capacitor C_1, we measured the time evolution of the current in the discharge circuit. Thereby, for a fixed length of the discharge gap, we could determine the minimum voltage U_{min} at which the discharge current continued to flow. Since the minimum discharge current I_{min} is determined by the threshold current of the cathode spot, the channel resistance R in the final stage of the spark discharge can be determined from the formula $R = U_{min}/I_{min}$. According to the data of [3], the threshold spot current for an iron cathode is $I_{min} = 1.5$ A.

In studying the generation of a pulsed arc across a 0.9-mm discharge gap, we determined the minimum voltage required for the current to flow through the gap. It can be seen from Fig. 2 that the current continues to flow at a discharge voltage of 35 V and the amplitude of the discharge current for the given discharge capacitance is about 3 A. When capacitor C_1 was charged below 30 V, no current was observed in the discharge circuit. To exclude the opportunity of disrupting the discharge due to the limitation of the discharge current to the starting arc current by the circuit parameters, we produced discharges through relay contacts with the same electric circuit. The current amplitude in this case was up to 100 A (Fig. 3). Measurements of the above minimum voltage across a 1.9-mm discharge gap gave a value of about 45 V at a discharge-current amplitude of up to 6 A. In discharges through relays, the amplitude of the discharge current at this voltage was larger than 160 A. In both cases, the energy deposited in the spark discharge was slightly above 0.1 J. With this energy deposition in the spark discharge, the resistance of the spark channel in the final stage of its evolution was found to be R = 23 and 30 Ω for the 0.9- and 1.9-mm gaps, respectively. A comparison of these values with the resistance of the positive columns of steady-state arcs shows that, under identical discharge conditions, the latter resistance is lesser by one order of magnitude [3]. It follows from here that further devel-

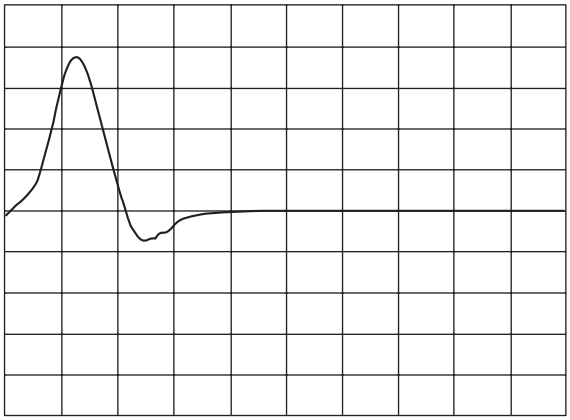


Fig. 3. Current oscillogram for a discharge through relay contacts. The discharge voltage is U_1 = 35 V, the sweep speed is 0.2 ms/division, and the current scale is 27 A/division.

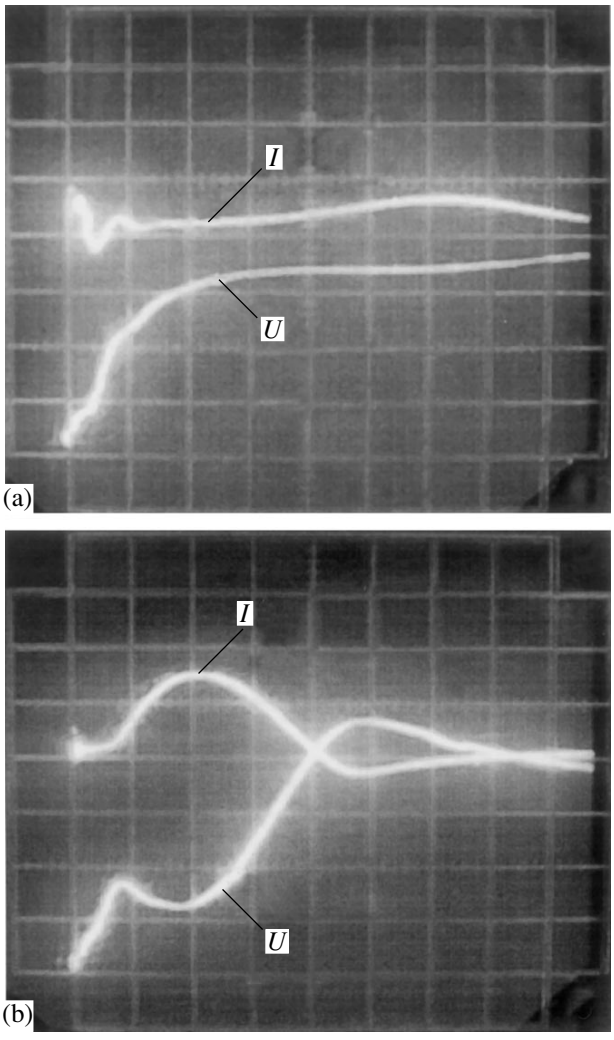


Fig. 4. Current and voltage oscillograms for a charging voltage of (a) 100 and (b) 400 V. The sweep speed is 20 μs/division. The current scales are (a) 270 and (b) 1351 A/division, and the voltage scales are (a) 1000 and (b) 200 V/division.

opment of the discharge at a low voltage across the discharge gap proceeds via step ionization of the gas.

We note that, at the minimum voltage ensuring the current to flow through the discharge gap, low-voltage capacitor C_1 is discharged only partially. Consequently, this electrode voltage does not sustain the degree of ionization that was produced in the discharge channel by the spark discharge. As a result, the discharge channel resistance increases, and, accordingly, the discharge terminates.

Therefore, from the known voltage at low-voltage capacitor C_1 at which it completely discharges, we can determine the value of the threshold voltage for igniting a pulsed arc. It was found that the threshold voltage for an energy deposition in a spark of less than 1 J per 1 cm of the gap length is higher than 60 V/mm. It should be noted that the threshold voltage is affected by the inductance of the transformer's secondary winding and the capacitor's self-inductance. These inductances increase the current rise time, thereby affecting the balance of the ionization and recombination processes. Hence, the above value of the threshold voltage is only an approximate estimate.

3. Influence of the Charging Voltage of the Low-Voltage Capacitor and the Interelectrode Distance on the Dynamics of Energy Release in a Pulsed Arc

Studying the waveforms of the current and voltage in the discharge circuit allowed us to compare the discharge duration and the dynamics of energy release under different discharge conditions. A characteristic feature revealed in comparing these curves was that the discharge duration decreased with increasing charging voltage at the low-voltage capacitor at a fixed length of the discharge gap. For instance, when the charging voltage of capacitor C_1 was 100 V, the duration of the pulsed arc was nearly 0.14 ms, whereas at a voltage of 400 V, this duration was nearly 0.06 ms (Fig. 4).

We note that, in studying the dynamics of energy release with this discharge circuit, it is rather difficult to determine exactly the influence of each of the discharge parameters. This is because the parameters of both the load and some elements of the discharge circuit are nonlinear. For this reason, the influence of the external circuit parameters was examined by changing the discharge gap with a lower resistance load. It was found that the current rise time and the extent to which the discharge duration depends on the charging voltage of capacitor C_1 are determined, in particular, by the circuit parameters. Therefore, the influence of particular discharge conditions was examined at fixed values of the other parameters.

At fixed charging voltages of the capacitors, the discharge duration insignificantly increased and the amplitude of the discharge current decreased as the interelectrode distance increased. Thus, when the voltage at capacitor C_1 was 200 V, the duration of the pulsed arc

across the 0.9-mm gap was 0.075 ms, whereas it was 0.08 ms at a gap length of 3 mm. In this case, the decrease in the current amplitude was 30 A, the current amplitude being larger than 1000 A.

4. Influence of the Initial Gas Pressure in the Discharge Gap on the Duration and Dynamics of Energy Release in the Discharge Circuit

This study was carried out with air discharges in "open" gaps of lengths 0.9 and 1.9 mm at a gas pressure in the range 10^5 to 8×10^5 Pa and an initial gas temperature of 293 K. The term "open" here implies that the volume of the discharge cavity is several orders of magnitude larger than the volume of the discharge channel. To decrease the measurement error, we only varied the air pressure in the discharge cavity; the charging voltages of the capacitors were fixed.

The results of these studies showed that, as the initial gas pressure in the discharge cavity increased, the discharge amplitude decreased, but the discharge duration increased. This was especially pronounced when we increased the voltage U_1 of capacitor C_1. For instance, at $U_1 = 200$ V, the increase in the initial pressure to 5.8×10^5 Pa resulted in a decrease in the current amplitude of more than 30 A. At $U_1 = 60$ V and the same increase in pressure, the current amplitude decreased by nearly 12 A. The increase in the discharge duration in the former case was about 5 ms, whereas in the latter case it was nearly 20 ms. Taking into account the voltage applied to the discharge gap, such an influence of the initial pressure can be explained by the fact that this parameter affects the resistance of the discharge channel.

CONCLUSIONS

The proposed method for determining the resistance of the discharge channel at the final stage of a spark discharge allows one to estimate the gas parameters in the discharge channel. It has been found how the discharge duration and the dynamics of energy release in the pulsed-arc circuit depend on the charging voltage of the low-voltage capacitor, the interelectrode distance, and the initial gas pressure in the discharge gap. An analysis of the dynamics of energy release shows that such a pulsed-arc circuit can be used in plasma–wave systems for the formation of intense shock waves.

REFERENCES

1. K. V. Korytchenko, Yu. Ya. Volkolupov, M. A. Krasnogolovets, *et al.*, Zh. Tekh. Fiz. **72**, 124 (2002) [Tech. Phys. **47**, 495 (2002)].
2. J. M. Meek and J. D. Craggs, *Electrical Breakdown of Gases* (Clarendon Press, Oxford, 1953; Inostrannaya Literatura, Moscow, 1960).
3. Yu. P. Raizer, *Physics of Gas Discharge* (Nauka, Moscow, 1987; Springer-Verlag, Berlin, 1991).

Translated by N. Larionova

Technical Physics, Vol. 48, No. 12, 2003, pp. 1590–1593. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 91–94.

SHORT COMMUNICATIONS

Method for Generating a Low-Voltage Pulsed Arc in Gas: Study of the Influence of the Circuit Parameters

A. N. Dovbnya*, K. V. Korytchenko*, Yu. Ya. Volkolupov*, S. M. Shkirida, M. A. Krasnogolovets, and V. S. Demin****

** Uskoritel' Research Center, Kharkov, 61108 Ukraine*
e-mail: dovbnya@nik.kharkov.ua
*** Kharkov Institute of Armoured Forces, Kharkov, 61034 Ukraine*
e-mail: entropia@rambler.ru

Received February 5, 2003; in final form, May 25, 2003

Abstract—A novel method for generating a pulsed arc in gas is considered. The method can be used in devices in which rapid and efficient heating of the working gaseous medium is required. Electric circuits for generating pulsed arcs are analyzed. For the chosen discharge circuit, the dynamics of energy release in the column of a pulsed arc is experimentally investigated. Methods for estimating the discharge circuit parameters are proposed. Results of estimating these parameters in an existing experimental device are presented. © *2003 MAIK "Nauka/Interperiodica".*

INTRODUCTION

In recent years, the problem of creating devices in which the electric-discharge energy is used to produce intense shock waves has attracted considerable interest. To solve this problem, a novel method for generating a pulsed arc is proposed. The method can be used in a plasma–wave system for the formation of intense shock waves [1].

REQUIREMENTS FOR THE CONDITIONS OF AN ELECTRIC DISCHARGE IN A PLASMA–WAVE SYSTEM

The specific features of a discharge in a plasma–wave system stem from the requirement for the gas in the positive column of a pulsed arc to rapidly relax to a quasi-equilibrium thermodynamic state. The study of the breakdown mechanisms and the development of discharges in gases allowed one to determine the conditions for generating discharges in such systems. These conditions are the following. A high-voltage pulse should provide spark breakdown of a gas-filled discharge gap. The further development of the discharge occurs at a lower electrode voltage. The low-voltage power supply should provide a high discharge current until a quasi-equilibrium thermodynamic state of the working gas is reached in the discharge. In this case, the energy deposited in the spark discharge is expected to be at least one order of magnitude lower than the energy released during the short-duration arc discharge. It is believed that, for such a discharge scenario, the discharge energy balance, which is governed by the electron–molecule interaction, can be changed by varying the electric field strength. A high electrode voltage ensures a high ionization rate of gas molecules. When the voltage is low, the fraction of the energy that is transferred from electrons to the rotational degrees of freedom of molecules and is spent on elastic losses substantially increases [2]. As a result, the rate and efficiency with which electric energy is converted into the kinetic energy of the gas molecules also increases.

In addition to the achievement of a quasi-equilibrium thermodynamic state of the gas in the discharge, it is necessary to provide the required rate of energy deposition in the discharge channel. This rate can be limited by the parameters of the circuit elements.

As applied to plasma–wave systems, the total discharge duration should be shorter than 5×10^{-5} s and the required amount of energy released during the discharge should be larger than 10 J. The voltage applied to the electrodes of the discharge gap during the arc discharge should differ slightly from the steady-state arc voltage.

CHOICE OF THE CIRCUIT FOR GENERATING A PULSED ARC WITH REQUIRED PARAMETERS

When electric circuits with an additional trigger electrode for the preliminary ionization of the working gas in the discharge gap are used, the voltage at the main electrodes should be substantially higher than the minimum arc voltage. This stems from the necessity of applying an additional potential between the electrodes in order to form the main current channel. A decrease in this potential results in an increase in the time delay of the formation of the main current channel or even in the disruption of the discharge. In particular, we should mention the three-electrode circuit [3] in which a dis-

1063-7842/03/4812-1590$24.00

charge initiated in the semi-enclosed volume produces a directed flow of the ionized gas, which bridges the main discharge gap (Fig. 1a). For the rapid formation of the directed gas flow, a rapid and efficient heating of the working gas is also required, which makes this circuit inapplicable in our case.

For electric circuits in which high-voltage and low-voltage sources connected in parallel to the discharge gap are used, precautions must be taken to ensure the simultaneous operation of the voltage sources (Fig. 1b). To protect the low-voltage source from a high-voltage pulse, active and reactive elements (e.g., a protective choker) are connected in series with the circuit. As a result, requirements for either the efficiency or the energy-release rate are not satisfied.

It is evident that the use of a pulse transformer directly connected to the discharge gap is hardly possible. The reason is that very different requirements are imposed upon the discharge dynamics at different stages, so that it is technically impossible to realize a transformer with the necessary output parameters (Fig. 1c).

Having examined different circuits for generating pulsed arcs, we chose an electric circuit that best matched our requirements (see Fig. 1d). This decision was made on the following grounds. Applying the high and low voltages in series ensures a continuous current throughout the entire discharge phase, so that the total discharge duration shortens. The circuit provides a decrease in the reactive component of the circuit because the pulse transformer operates in the switch regime. This makes it possible to achieve the required rate of energy release.

EXPERIMENT

Dynamics of the Energy-Release in the Electric Discharge Circuit under Study

The electric circuit presented in Fig. 1d was put into practice. The pulse transformer produced high-voltage pulses with an amplitude of up to 26 kV. The charging voltage of the low-voltage capacitor was varied in the range 0–400 V. Such parameters of the device allowed us to study the dynamics of energy release in air gaps with a length of up to 3 mm under normal atmospheric conditions. The use of a capacitor of the MBGN type with a rating value of 200 mF as low-voltage energy storage provided the required discharge dynamics (Fig. 2).

An analysis of the energy-release dynamics shows that this electric circuit can be used in plasma–wave systems for the formation of intense shock waves.

In view of the nonlinear characteristic of the load, the question arises as to how the discharge circuit parameters influence the energy-release dynamics. The results of the study of these parameters are presented below.

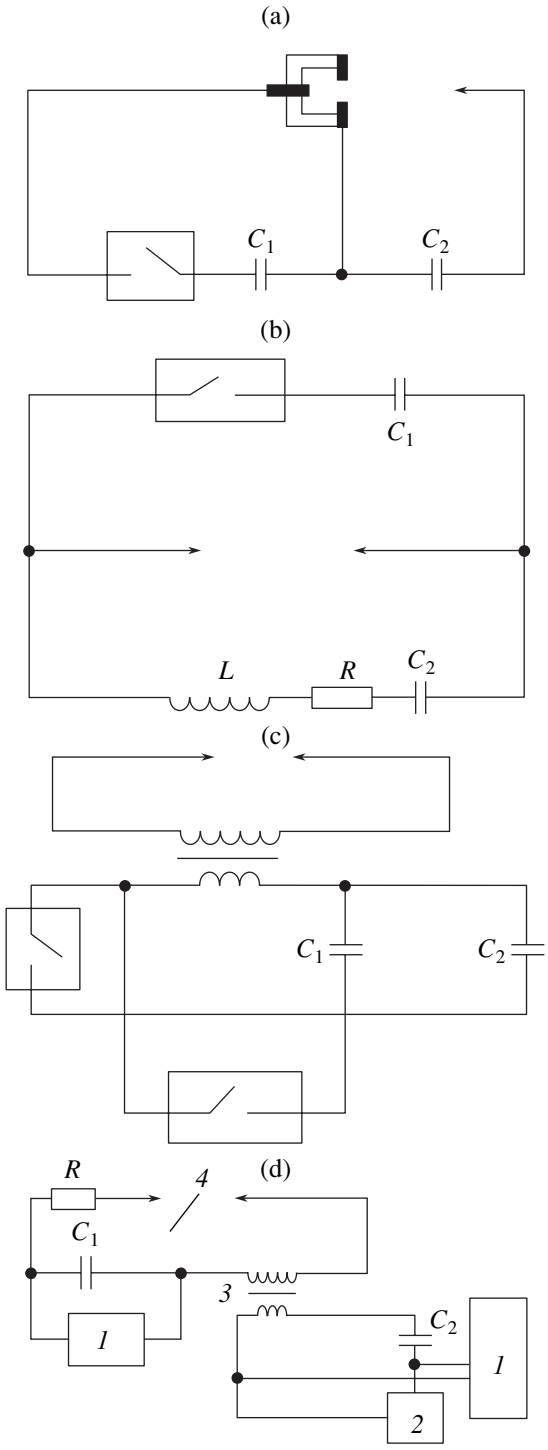


Fig. 1. (a–c) Electric discharge circuits and (d) pulsed arc circuit: (*1*) charging devices, (*2*) switch, (*3*) transformer, (*4*) discharge gap, and (*R*) measuring shunt.

Study of the Influence of the Self-Inductance of the Capacitive Storage on the Discharge Dynamics

One of the capacitor parameters that affects the discharge duration is the self-inductance of the capacitor. In order to determine how the self-inductance of the

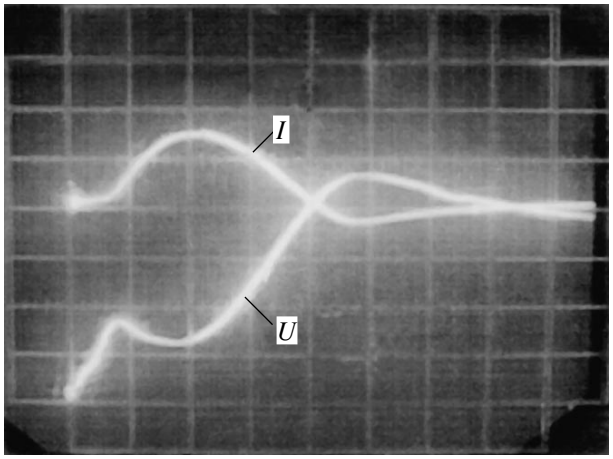


Fig. 2. Oscillograms of the current I and voltage U. The sweep speed is 20 ms/division, the current scale is 1351 A/division, and the voltage scale is 200 V/division.

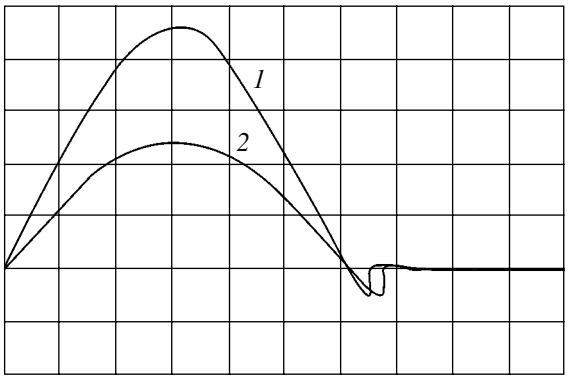


Fig. 3. Oscillograms of the current for a capacitor charging voltage of (*1*) 200 and (*2*) 100 V. The sweep speed is 10 ms/division, and the current scale is 270 A/division.

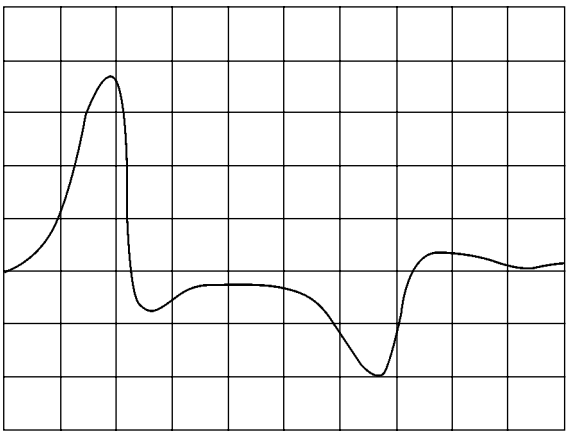


Fig. 4. Oscillograms of the current. The capacitor charging voltage is 100 V, the current scale is 270 A/division, and the sweep speed is 0.1 ms/division.

available capacitor influences the discharge duration, it was proposed to trace the dynamics of the discharge current when the capacitor discharges through a thyristor and a low-resistance shunt.

The type of thyristor was chosen taking into account the maximum values of the voltage and current in the discharge circuit and the value of the potential drop in the cathode and anode regions (this value should be much lower than the potential drop across the discharge gap). When the current growth rate is high, the effect of nonlinearity of the current–voltage characteristic of the thyristor can be ignored. In this case, the current rise time should be at least one order of magnitude shorter than the total discharge duration.

The MBGN capacitor discharged through a TChI-100 thyristor. The capacitor charging voltage was varied in the range 0–400 V. Using this type of thyristor, we succeeded in substantially decreasing the influence of the characteristic of the electronic switch on the energy-release dynamics.

By comparing the waveforms of the current for the cases in which the capacitor discharged through the thyristor and through the pulsed arc circuit, we could determine the influence of the capacitor's self-inductance on the energy-release dynamics in the given discharge circuit. It was found that, in our case, the capacitor's self-inductance results in a gradual current rise during an arc discharge (Fig. 3).

Study of the Combined Effect of the Nonlinear Resistances of the Circuit Elements on the Discharge Dynamics

This study was performed by including a relay (instead of the discharge gap) in the pulsed arc circuit. The discharge circuit with capacitor C_2 was not used. In this version, no "external" saturation of the transformer core occurred, which resulted in distorted information about the actual influence of the circuit parameters on the discharge dynamics. However, these studies made it possible to determine the maximum attainable amplitude of the discharge current in the given circuit at a fixed initial charging voltage of capacitor C_1. This estimate allowed us to separate the influence of the state of the gas in the discharge channel from the influence of the circuit parameters on the energy-release dynamics in the given discharge circuit.

In this case, the main requirements for the relay parameters are a minimal influence of the relay flutter on the energy-release dynamics and a low resistance of the relay contacts. The latter implies the use of a relay with a large contact area.

In the relay used in our experiments, the total contact area was 39.25 mm^2. Figure 4 shows the waveform of the discharge current. We note that the use of a relay causes the problem of synchronization with the measuring equipment. This can result in a shift of the start time of scanning. However, on the whole, the time evo-

lution of the discharge current remained stable. An analysis of curves shows that, when an additional transformer-core saturation produced by an external current source was absent, an oscillatory process occurred in the discharge circuit. This means that the discharge circuit shown in Fig. 1b is unsuitable. The absence of sharp current spikes indicates a rather short period of relay flutter. A comparison of the waveforms of the discharge current through the discharge gap and through the relay contacts shows that threshold values of the arc voltage are determined by the state of the gas in the discharge channel, rather than the disruption of the discharge current due to the influence of the discharge-circuit parameters. The observed waveforms of the discharge current through the relay contacts allow us to determine how the nonlinear resistances of the electric circuit elements influence the current dynamics in the pulsed arc circuit.

CONCLUSIONS

In this paper, we studied an electric circuit for generating a low-voltage pulsed arc. It was found that the given circuit can be used in plasma–wave systems for the formation of intense shock waves and other devices in which fast and efficient heating of the working gaseous medium is required. The proposed versions of experimental studies of the influence of resistances of the electric circuit elements on the discharge dynamics allow one to estimate the applicability range of a device based on this circuit and find a way of optimizing the device parameters.

REFERENCES

1. K. V. Korytchenko, Yu. Ya. Volkolupov, M. A. Krasnogolovets, *et al.*, Zh. Tekh. Fiz. **72**, 124 (2002) [Tech. Phys. **47**, 495 (2002)].
2. Yu. P. Raizer, *Physics of Gas Discharge* (Nauka, Moscow, 1987; Springer-Verlag, Berlin, 1991).
3. S. A. Romanenko, *Pulsed Arcs in Gases* (Nauka, Moscow, 1968).

Translated by E. Satunina

Technical Physics, Vol. 48, No. 12, 2003, pp. 1594–1597. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 95–98.

SHORT COMMUNICATIONS

Elimination of Styrene Vapor from Air by a Pulsed Electron Beam

Yu. N. Novoselov and I. E. Filatov

Institute of Electrophysics, Ural Division, Russian Academy of Sciences, ul. Amundsena 106, Yekaterinburg, 620016 Russia

e-mail: nov@iep.uran.ru

Received March 20, 2003

Abstract—Styrene vapor contained in air in small amounts is decomposed when the air is subjected to a pulsed electron beam and non-self-maintained space discharge. The physical laws of the process and the final products of styrene vapor conversion are found. Experimental data make it possible to consistently describe the styrene vapor elimination from air exposed to a pulsed electron beam and relate the beam parameters to the properties of the gas flow being irradiated. © *2003 MAIK "Nauka/Interperiodica".*

The process of organic synthesis, as well as the production of plastics and plastic products, inevitably generates effluents that sometimes contain toxic volatiles in dangerous concentrations. The vapors of aromatic and nonsaturated compounds, such as styrene, benzene, and the like, are the most dangerous. One way of reducing the toxicant concentration in the gaseous effluents is to irradiate them by pulsed electron beams. It has been shown [1, 2] that irradiation may reduce significantly the concentration of inorganic substances without consuming much energy. In this work, we report experimental data for styrene vapor removal by irradiating the air with pulsed nanosecond electron beams.

EXPERIMENTAL CONDITIONS

We experimented with model nitrogen : oxygen = 80 : 20 gas mixtures at atmospheric pressure and room temperature. The styrene vapor concentration was varied from 50 to 1000 ppm. The concentration of the styrene and styrene conversion products in the gas phase was measured chromatographically [3]. The error in measuring the impurity concentration was no more than 5% in the 100–1000 ppm range and 12% in the 10–100 ppm range.

The experiments were carried out with a setup based on the RADAN nanosecond electron accelerator [4]. The accelerator generated a 180-keV pulsed electron beam with a current of 800 A, cross-sectional area of 1 cm^2, half-height duration of 3 ns, and repetition rate of 10 s^{-1}. The beam irradiated a 1-cm-long gas-filled gap with a volume of about 10 cm^3 between a grid electrode, through which the beam was injected into the gas, and a solid metallic electrode, connected to a 13-nF storage capacitor. The voltage across the storage capacitor was varied from 0 to 10 kV. The energy delivered to the test gas volume (about 4.2×10^{-3} J per pulse) was measured by standard film dose meters. This energy plus the energy of the storage capacitor initiated a non-self-maintained discharge. The total volume of the discharge chamber was 3 l. A fan built in the chamber made the gas flow through the gap with a mean velocity of 0.5 m/s. Such an experimental scheme simulates multistage purification, where the gas flow is repeatedly irradiated. In addition, it improves the reproducibility and decreases the error of measurements by averaging the styrene vapor concentration in the air.

The gas was irradiated by trains of pulses. In the course of experiments, we determined the absolute change ΔC in the impurity concentration, the degree of impurity conversion (i.e., the degree of purification), and the energy spent to remove a styrene molecule. The conversion η and the energy consumption per molecule ε (eV/mol) were calculated as

$$\eta = \Delta C/C_0, \quad \varepsilon = (W_b + W_d)N/e\Delta C.$$

Here, $\Delta C = C_0 - C$ (cm^{-3}) is the change in the styrene concentration per train; C_0 and C (cm^{-3}) are the initial and current impurity concentrations in the mixture, respectively; W_b is the electron beam energy introduced into the gas; W_d (J/cm^3) is the discharge energy introduced into the gas; N is the number of pulses per train; and e (C) is the electron charge. The total error in determining ε (including the error in determining the absorbed energy $W = W_b + W_d$) was no more than 0.3 eV/mol.

EXPERIMENTAL RESULTS

It was found that the styrene concentration drops with increasing number of nanosecond irradiating pulses (Fig. 1). The value of ε depends on the initial styrene impurity concentration. Typically, ε lies in the range from 2 to 10 eV/mol for C_0 = 100–1000 ppm. At low impurity concentrations (from 50 to ≈300 ppm),

1063-7842/03/4812-1594$24.00 © 2003 MAIK "Nauka/Interperiodica"

C varies with N by a near-exponential law (Fig. 1). Typical dependences of the logarithm of the initial-to-current concentration ratio, $\ln(C_0/C)$, on the energy W delivered to the gas are shown in Fig. 2. At low concentrations (curves *1*, *2*), the curves approach straight lines, indicating that the concentration varies exponentially.

The slope of these straight lines may serve as an energy coefficient β (e.g., in units of J/cm^3) that relates the purification efficiency to the energy consumption [5]:

$$\ln(C_0/C) = W/\beta. \quad (1)$$

Physically, this coefficient shows the amount of energy that is necessary to apply to the gas for the impurity concentration to decrease by a factor of e = 2.718 ….

It turned out that β depends on the initial impurity concentration C_0. As C_0 grows, the dependence becomes nonlinear (Fig. 3).

The final product of styrene vapor conversion is polystyrene, which covers the gas chamber walls in the form of thin film. A small amount of benzaldehyde, the product of styrene oxidation, was also detected in small amounts. Benzaldehyde forms when styrene combines with ozone and other active oxygen species generated by the electron beam. It was found experimentally that the concentration C_{ba} of benzaldehyde is much lower than the initial styrene concentration C_0. After irradiation, the benzaldehyde concentration was estimated with the parameter $\sigma = (C_0 - C)/C_{ba}$, which shows the number of benzaldehyde molecules produced per removed styrene molecule.

Provided that the number of irradiating pulses (hence, the energy introduced into the gas) is the same, this parameter was found to have a constant value. For example, if the number of irradiating pulses is $N = 8 \times 10^3$, $\sigma = 0.18 \pm 0.02$. This means that, if the initial concentration of the styrene is C_0 = 1000 ppm, the benzaldehyde concentration after irradiation by N pulses will be $C_{ba} \approx 180$ ppm; if C_0 = 300 ppm, $C_{ba} \approx 54$ ppm.

We carried out experiments with a non-self-maintained discharge in the electrode gap. The current concentrations of styrene and benzaldehyde vapors were measured as a function of the electric field strength E in the discharge column. The parameters β and σ were found to depend on the field strength. The experimental dependence of β on E in the discharge column for C_0 = 300 ppm is shown in Fig. 4 (curve *1*). It is seen that the energy consumption is minimal for $E \approx 500$ V/cm.

Curve *2* in Fig. 4 shows σ vs. E for the number of irradiating pulses $N \approx 5 \times 10^3$. There is an optimal field strength at which the number of benzaldehyde molecules produced decreases severalfold compared with the case when the field is absent. If the initial styrene concentration is C_0 = 300 ppm, the parameter σ equals 0.06 after irradiation by $N \approx 5 \times 10^3$ pulses of the non-self-maintained discharge with a field strength E = 800–1000 V/cm in the discharge column. This value is much less than 0.19 in the absence of the discharge.

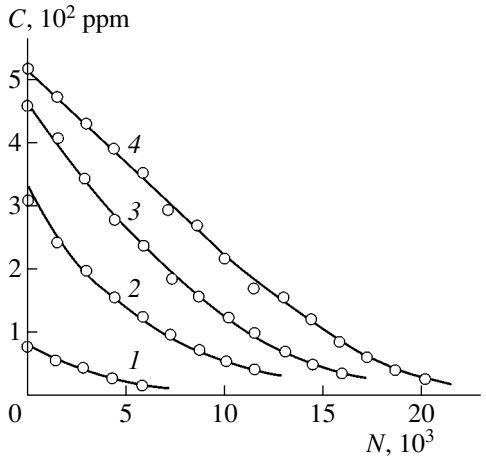


Fig. 1. Styrene vapor concentration C vs. the number N of irradiating pulses. The initial concentration C_0 = (*1*) 75, (*2*) 310, (*3*) 460, and (*4*) 520 ppm.

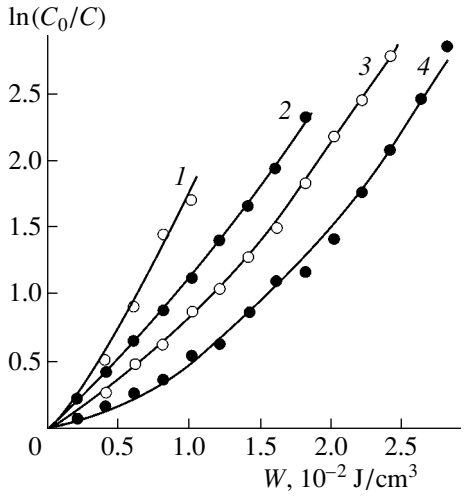


Fig. 2. Dependence of $\ln(C_0/C)$ on the energy W applied to the gas for C_0 = (*1*) 75, (*2*) 310, (*3*) 460, and (*4*) 520 ppm.

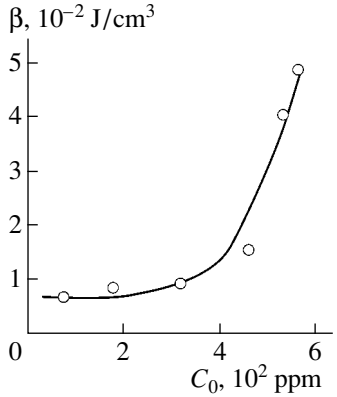


Fig. 3. Parameter β vs. initial styrene vapor concentration C_0.

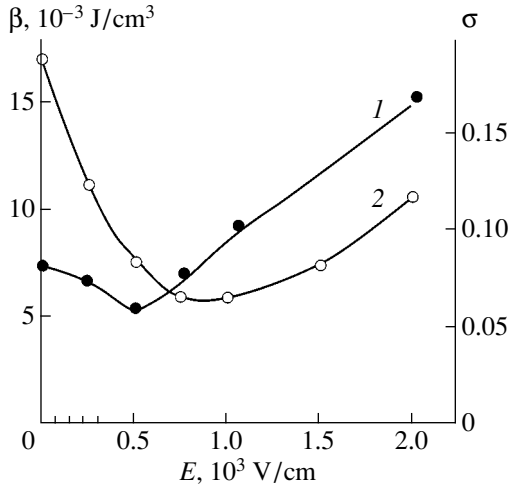


Fig. 4. Parameters (*1*) β and (*2*) σ vs. electric field strength *E*.

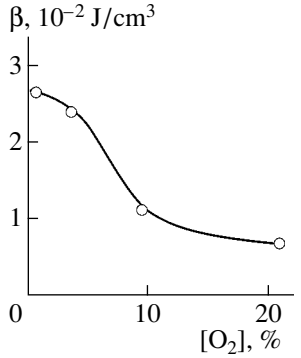


Fig. 5. Parameter β vs. O_2 concentration in the mixture.

One more series of experiments was concerned with the effect of oxygen concentration on styrene vapor removal. If oxygen is absent in the mixture, the styrene is virtually not eliminated. An increase in the O_2 concentration results in a drastic reduction of β, and the related experimental curve (Fig. 5) is described well by the relationship

$$\beta = KC_0[O_2]^{1/2}, \quad (2)$$

where *K* is a numerical conversion factor.

FORMAL DESCRIPTION OF THE PROCESS

Simple expressions (1) and (2), which fit experimental data for styrene vapor removal from air by means of pulsed electron beams, allow one to suggest a formal approach to the process. The most extensively employed approach in this case is the formal kinetic description of kinetic reactions involving two or more reagents [6]. The process in this case is described by simple equations where reagent concentrations and running time are parameters. Our empirical equations relate impurity concentration and energy introduced into the gas; therefore, the description which follows may be called the formal energy approach by analogy.

Without considering complex reactions of styrene conversion, the parameter β may be formally termed the characteristic energy by analogy with the formal kinetic approach. The smaller β, the higher the efficiency of the approach.

As follows from Fig. 3, β depends on the initial impurity concentration C_0. When the styrene vapor concentration tends to zero, the energy needed to achieve a given degree of purification tends to a definite nonzero value,

$$\lim_{C \to 0}(\beta) = \text{const} = \beta_0, \quad (3)$$

which is called the initial characteristic energy. Based on the curves shown in Fig. 3, one may assume that β_0 falls into the range $(7–10) \times 10^{-3}$ J/cm^3 when styrene is irradiated by a pulsed electron beam. At low initial concentrations (C_0 = 50–300 ppm), $\beta_0 = 7 \times 10^{-3}$ J/cm^3 and remains practically unchanged.

Using the analogy of the description suggested and the formal kinetic approach used in [6], it may be inferred (see (2)) that one oxygen molecule is responsible for the removal of one styrene molecule. The most plausible explanation for this fact is the participation of atomic oxygen or ozone in the basic route of the styrene conversion reaction.

The above experimental data suggest that, when a pulsed electron beam acts on styrene vapor to initiate the styrene conversion reaction, the characteristic energy β depends on the initial styrene concentration C_0 and oxygen concentration $[O_2]$. Generally, β may also depend on the concentration of other reagents that appear under the action of the beam. A detailed formal description of the process involving several reagents is given in [7].

In terms of the characteristic energy, the basic parameters of impurity conversion in air (the degree of purification and the energy needed to remove one molecule) are given by

$$\eta \equiv (1 - C/C_0) = 1 - \exp(-W/\beta), \quad (4)$$

$$\varepsilon = -A\beta(\eta C_0)^{-1}\ln(1-\eta), \quad (5)$$

where *A* is a numerical conversion factor.

If β is expressed in J/cm^3; the concentrations *C* and C_0, in ppm; and ε, in eV/mol, the factor *A* roughly equals 0.236 (eV cm^3 ppm)/J.

From (5), it follows that one cannot judge the efficiency of different techniques of impurity conversion in air based on the energy consumption ε alone. Correct comparison also includes the initial characteristic energy β_0.

The rules and parameters of the process found in our experiments allow one to estimate the throughput *V* (l/s) of a purifier, i.e., the flow rate of a gas being puri-

fied, and its power $P(W)$ for a given initial styrene vapor concentration C_0 and a desired degree of purification η:

$$V = P/W \equiv -P/[\beta \ln(1-\eta)]. \quad (6)$$

For example, if it is necessary to remove styrene of concentration $C_0 = 200$ ppm from air flowing with a rate of 100 l/s and attain a degree of purification of 50% with an electron accelerator (in this case, $\beta = \beta_0 = 7 \times 10^{-3}$ J/cm^3), the electron energy must be equal to about 500 W according to (6). With a streamer or any ionizer other than a non-self-maintained discharge or with an increase in C_0, this value will certainly change.

Thus, as follows from our experiments on removing styrene vapor by a pulsed electron beam, the initial characteristic energy β_0 may be used as an efficiency parameter. This parameter defines the amount of energy that must be applied to polluted air for the pollutant concentration (tending to zero) to decrease by a factor of $e = 2.718\ldots$. The parameters β and σ, along with expressions (4)–(6), completely characterize the process of styrene vapor removal from air with a pulsed electron beam.

REFERENCES

1. Yu. N. Novoselov, G. A. Mesyats, and D. L. Kuznetsov, J. Phys. D **34**, 1248 (2001).
2. G. V. Denisov, Yu. N. Novoselov, A. I. Suslov, and A. M. Uster, Zh. Tekh. Fiz. **71**, 136 (2001) [Tech. Phys. **46**, 129 (2001)].
3. V. G. Berezkin and V. S. Tatarinskii, *Gas Chromatographic Analysis of Trace Impurities* (Nauka, Moscow, 1970).
4. Yu. N. Novoselov, Yu. S. Surkov, and I. E. Filatov, Pis'ma Zh. Tekh. Fiz. **26** (13), 23 (2000) [Tech. Phys. Lett. **26**, 550 (2000)].
5. G. A. Mesyats, Yu. N. Novoselov, and I. E. Filatov, Pis'ma Zh. Tekh. Fiz. **27** (19), 25 (2001) [Tech. Phys. Lett. **27**, 813 (2001)].
6. V. N. Kondrat'ev and E. E. Nikitin, *Kinetics and Mechanism of Vapor-Phase Reactions* (Nauka, Moscow, 1974).
7. Yu. N. Novoselov and I. E. Filatov, Dokl. Akad. Nauk **382**, 1 (2002).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 12, 2003, pp. 1598–1599. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 99–100.

SHORT COMMUNICATIONS

Inertial Navigation Method Based on Absolute Acceleration Measurements

A. S. Devyatisilny

Institute of Automation and Control, Far East Division, Russian Academy of Sciences, Vladivostok, 690041 Russia
e-mail: devyatis@iacp.dvo.ru

Received March 19, 2003

Abstract—The physical grounds of an inertial navigation method that measures the resultant vector of specific forces (absolute acceleration) are stated. The problem of combining this method with the conventional method based on measuring the vector of the apparent acceleration is discussed. © *2003 MAIK "Nauka/Interperiodica".*

The design and operation of a device for measuring the absolute linear acceleration of a moving object, as well as the potentiality for using this meter in inertial navigation systems (INSs), were described in [1]. The measurement accuracy was predicted to be as high as $\sim 10^{-4}$ m/s^2, which meets well practical demands at present and in the immediate future. Without questioning the conclusions drawn in [1], we will point to a number of features of the new meters that should be taken into account when viewing them as candidates for INS elements.

As is well known [2], an INS functions according to equations that describe the motion of a unit-mass material point (dynamic equations) and the evolution of a frame of reference (kinematic equations, which are sometimes called Poisson's equations) where the motion of the object (point) is considered. Below, we will refer to equations of the first group.

In the absolute (inertial) frame of reference, the equation of motion of a point unit mass has the form

$$\ddot{\mathbf{R}} = \mathbf{g} + \mathbf{f}.$$

Here, overcircles above the radius vector of the position of the mass mean absolute derivatives. Accordingly, $\ddot{\mathbf{R}}$ is the absolute acceleration and $\mathbf{g}$ and $\mathbf{f}$ are the vectors of specific gravitational and nongravitational (e.g., drag, engine thrust, etc.) forces. Thus, measuring the absolute acceleration $\ddot{\mathbf{R}}$ is nothing but measuring the total (resultant) specific force ($\mathbf{g} + \mathbf{f}$). It is precisely this way of measuring this force and the device used that are dealt with in [1]. This device will be called a *gf* meter in order to distinguish it from a standard newtonmeter [3], or *f* meter, which measures the force $\mathbf{f}$.

In a frame of reference rotating with an absolute angular velocity $\boldsymbol{\omega}(t)$, the dynamic equations for INS perfect operation have the form [2]

$$\begin{aligned} \dot{\mathbf{R}} &= -\boldsymbol{\omega} \times \mathbf{R} + \mathbf{V}, \quad \mathbf{R}(0) = \mathbf{R}_0, \\ \dot{\mathbf{V}} &= -\boldsymbol{\omega} \times \mathbf{V} + \mathbf{g}(\mathbf{R}) + \mathbf{f}, \quad \mathbf{V}(0) = \mathbf{V}_0, \end{aligned} \tag{1}$$

where $\mathbf{R}$ is the radius vector of the point (objective), $\mathbf{V}$ is the vector of its absolute linear velocity, $\mathbf{g}(\mathbf{R})$ is the vector of gravitational forces, $\mathbf{f}$ is the vector of all the other forces, and overcircles stand for local time derivatives.

When designing a conventional INS, one must know the gravitational field model (that is, the model of specific force $\mathbf{g}(\mathbf{R})$) and the way of measuring the specific force $\mathbf{f}$ (and, in general, the vector $\boldsymbol{\omega}(t)$; this issue is disregarded here). For *gf* measurements, a model of $\mathbf{g}(\mathbf{R})$ is not necessary (as was justly noted in [1]), since the vector $\mathbf{g}(\mathbf{R}) + \mathbf{f}$ is assumed to be known (with instrumental accuracy). Without going into detail, we also note that the three-component INS scheme is unstable in both cases, although the character of instability is different [2, 5].

Let us identify the rotating frame of reference with a right-handed moving (ideal) Cartesian trihedron $0x = 0x_1x_2x_3$ such that the $0x_3$ axis is aligned with the vector $\mathbf{R}$. We assume that $0y = 0y_1y_2y_3$, a device trihedron where all measurements are taken, is the physical model of the trihedron $0x$.

Consider the case when both inertial measurements (*f* and *g* measurements) are taken simultaneously. Assuming that the mutual orientation of the trihedra $0x$ and $0y$ is described through a small angle of rotation α, we may represent *f* and *gf* measurements as

$$\mathbf{J}_1 = (E + \hat{\alpha})\mathbf{f} + \Delta\mathbf{f}_1,$$

$$\mathbf{J}_2 = (E + \hat{\alpha})(\mathbf{g}(\mathbf{R}) + \mathbf{f}) + \Delta\mathbf{f}_2,$$

where E is the unity matrix and $\Delta\mathbf{f}_1$ and $\Delta\mathbf{f}_2$ are the vectors of the instrument errors.

1063-7842/03/4812-1598$24.00

Comparing both available measurements yields a new one:

$$\mathbf{J}_3 = \mathbf{J}_2 - \mathbf{J}_1 = (E + \hat{\alpha})\mathbf{g}(\mathbf{R}) + \Delta\mathbf{f}, \quad (2)$$

where $\Delta\mathbf{f} = \Delta\mathbf{f}_2 - \Delta\mathbf{f}_1$.

For small enough $\Delta\mathbf{f}$ and α, measurement (2) can be used as a test measurement of gravitational fields that are poorly understood.

If a gravitational field is well studied and its model is known, the solution of (1) may numerically yield a vector $\mathbf{g}(\mathbf{R}')$, where $\mathbf{R}' = \mathbf{R} - \delta\mathbf{R}$, $\delta\mathbf{R} = (\alpha_2 r, -\alpha_1 r, \delta r)^T$ is the vector of inaccuracies involved in $\mathbf{R}$ calculated by model (1) in the projections onto the axes of the trihedron $0y$, and $r = |R|$. Comparing $\mathbf{J}_3$ and $\mathbf{g}(\mathbf{R}')$, we then find

$$\mathbf{J} = \mathbf{J}_3 - \mathbf{g}(\mathbf{R}') = \hat{\alpha}\mathbf{g}(\mathbf{R}) - \frac{\partial \mathbf{g}(\mathbf{R})}{\partial \mathbf{R}}\delta\mathbf{R} + \Delta\mathbf{f}. \quad (3)$$

Apparently, measurement (3) can be used to correct a dynamic group of INS malfunctions. For navigation in the central gravitational field (the exterior field of terrestrial gravitation may be considered as central, since the central component essentially prevails in the expansion of the Earth's gravitational potential), (3) takes the form

$$J = 2\omega_0^2(r)\delta r + \Delta\tilde{f}, \quad (4)$$

where $\tilde{f}$ is the third component of the vector $\Delta\mathbf{f}$, $\omega_0 = (\mu/r^3)^{1/2}$ is the Shuler frequency, and μ is the Earth's gravitational potential.

From (4) it is easy to see that a pair of inertial measurements (*f* and *gf* measurements) is equivalent to the noninertial (e.g., radar, barometric, etc.) measurement of the object's altitude with instrumental inaccuracy

$$\Delta h = \frac{\Delta\tilde{f}}{2\omega_0^2}.$$

For example, with $|\Delta\mathbf{f}_2| = 10^{-4}$ m/s^2 (as in [1]) and $|\Delta\mathbf{f}_1| = |\Delta\mathbf{f}_2|$, we have $|\Delta h| \le 70$ m at the Earth's surface.

It should be noted that the state–measurement system produced by measurement equation (4) and equations for INS dynamic errors (the latter are obtained by varying Eqs. (1)) is observable (at least when the object moves with a constant velocity along parallels of latitude) in the general (Kalman's [4]) sense. This implies the possibility of constructing asymptotically robust correction algorithms for this group of INS malfunctions [5].

Thus, the basic conclusions that can be drawn from our analysis are as follows. (i) When the inaccuracies Δf and α meet obvious conditions, the use of a *gf* meter (together with an *f* meter) seems to be promising for probing little-studied gravitational fields; (ii) the combined use of *gf* and *f* meters raises the information independence of an INS, since INS malfunction (in particular, dynamic errors) can be reliably corrected based solely on inertial information; (iii) the use of a precision (for example, radar) altimeter virtually eliminates the problem of deciding between *f* and *gf* meters as inertial information sources (in the case of comparable accuracies): one may choose the one which is preferable based on technical, economical, service, etc., parameters.

REFERENCES

1. Yu. M. Kolyada, S. V. Sokolov, and S. A. Olenev, Izmer. Tekh., No. 4, 33 (2001).
2. V. D. Andreev, *Theory of Inertial Navigation: Correctable Systems* (Nauka, Moscow, 1967).
3. A. Yu. Ishlinskiĭ, *Classical Mechanics and Inertial Forces* (Nauka, Moscow, 1987).
4. R. Kalman, P. Falb, and M. Arbib, *Topics in Mathematical System Theory* (McGraw-Hill, New York, 1969; Mir, Moscow, 1971).
5. A. S. Devyatisilny, Dal'nevostochnyi Mat. Zh. **3**, 227 (2002).

Translated by V. Isaakyan

Technical Physics, Vol. 48, No. 12, 2003, pp. 1600–1602. Translated from Zhurnal Tekhnicheskoĭ Fiziki, Vol. 73, No. 12, 2003, pp. 101–103.

SHORT COMMUNICATIONS

Holographic Screen To Project a 3D Image by the Method of Point-Focused Aspects

Yu. N. Denisyuk, N. M. Ganzherli, and D. F. Chernykh

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

e-mail: nina@holo.ioffe.rssi.ru

Received May 19, 2003

Abstract—A holographic screen that makes it possible to display 3D images by the method of point-focused aspects (of a scene to be displayed) or by projecting narrow line holograms is suggested and implemented. The image thus formed reconstructs only horizontal parallax, which allows one to cut the body of information necessary for 3D imaging. Projection through the screen and a way of recording the screen are considered. © *2003 MAIK "Nauka/Interperiodica".*

Earlier, a method of 3D imaging where an image is formed with an array of point-focused aspects of a scene to be displayed was reported [1–3]. At the first step, a set of aspects of a scene to be displayed is recorded. The array is produced by a scanning laser beam into which the scene's aspects recorded from different points are introduced. At the stage of projection, the aspects are illuminated by a point-focused laser beam. Scanning the screen, this beam generates an array of luminous points through each of which one of the aspects of the 3D scene is seen. The observer sees a local 2D image whose configuration varies according to the point of observation as in the observation of a real 3D object. Thus, the illusion that the image projected is three-dimensional arises. It was noted that the problem of recording and projecting a 3D image may be considerably simplified if only the horizontal parallax of a scene being displayed is reconstructed. A drawback of this approach is that the square field of vision through which the observer sees the image is transformed into a narrow horizontal line, making observation difficult. This line may be extended in the vertical direction by introducing a special 1D diffuser screen into the projecting system. However, designing a two-component optical scheme consisting of a very large lens and 1D diffuser screen is a challenge.

Here, we suggest a special holographic screen that focuses and diffuses the light in the vertical direction simultaneously. The scheme for recording such a screen is shown in Fig. 1. A cylindrical lens focuses laser radiation to form a narrow vertical line in the plane of diffuser *D*. Then, the diffuse light falls on light-sensitive medium (photoplate) *P*. Wave front *W* is also directed to the light-sensitive medium and converges at point *S* behind it. Hologram *H*, which is recorded in the area where the two beams meet, has the form of an ellipse extended in the horizontal direction. Holograms were recorded on PFG-03M photoplates and developed in a GP-3 holographic developer.

The holographic optical element (screen) thus recorded is capable of "imaging" the radiation from point *S* as vertical diffuse line *DL* (Fig. 1). When a 3D image is projected through the screen, the luminous points where different aspects of the scene are focused lie along horizontal line *LL* (Fig. 2). The screen, as was noted above, projects each of the points onto field of vision *FV* in the form of vertical diffuse line *DL*. Through each of the points of this line, observer *h* sees

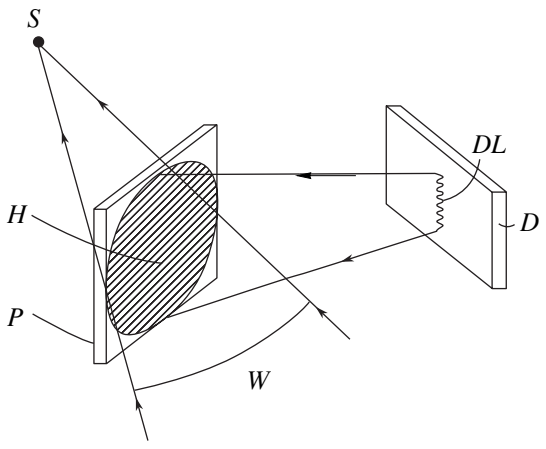


Fig. 1. Recording of the holographic screen: *W*, wave front converging to point S; *D*, diffuser on which the cylindrical lens projects vertical diffuse light line *DL*; and *P*, light-sensitive material (photoplate) with recorded hologram *H*.

1063-7842/03/4812-1600$24.00 © 2003 MAIK "Nauka/Interperiodica"

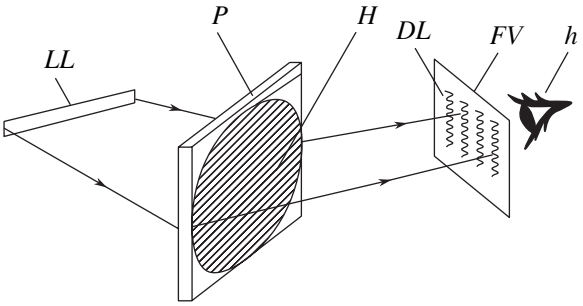


Fig. 2. Reconstruction of the holographic screen reproducing horizontal parallax. *P*, light-sensitive material (photoplate) with recorded hologram *H*; *LL*, reconstructing light line; *FV*, field of vision produced by a set of reconstructed vertical diffuse lines *DL*; and *h*, observer.

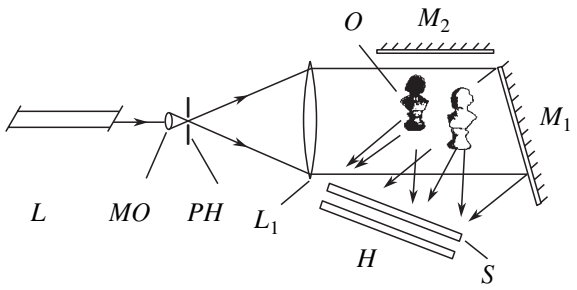


Fig. 3. Recording a set of double-beam line holograms of the 3D scene. *L*, laser; M_1 and M_2, mirrors; *MO*, microobjective lens; *O*, object; L_1, lens; *H*, light-sensitive medium; *PH*, pinhole; and *S*, split diaphragm.

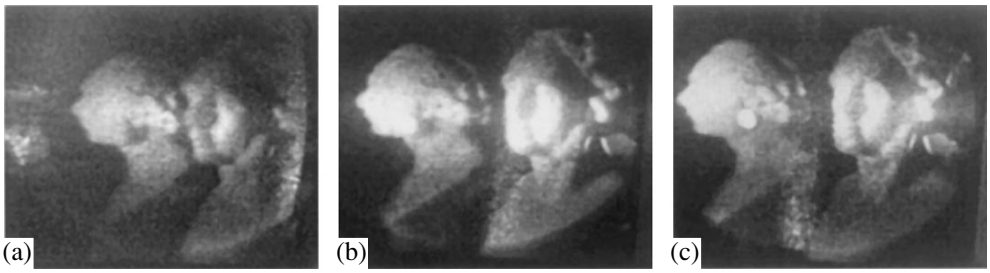


Fig. 4. Images of the 3D scene on the holographic screen for three positions of the observer's eye.

the image of the same aspect that is projected from a given point of horizontal light line *LL*. Other aspects of the 3D scene are seen through their corresponding vertical lines *DL*.

Thus, the points of field of vision *FV* are extended vertically. This enables the observer to see the image projected even if the eye is considerably shifted in the vertical direction.

Such a screen may be used to project 3D images by projecting aspects obtained with conventional photography or by projecting hologram-generated aspects (since the radiation from each of the hologram points may be considered as an aspect). In the latter case, the screen provides large-scale images with the help of a narrow line hologram. In experiments, we studied this type of projection by recording line double-beam holograms of a 3D scene that consisted of two objects *O* and mirrors M_1 and M_2 (Fig. 3). Additional illumination from the mirrors enhanced the apparent depth of the scene. The objects, two 6-cm-high busts made of white mat marble, were arranged at different distances from the mirrors and recording medium *H*. Objects *O* were illuminated by a part of the collimated beam, which then fell on mirror M_1 and reflected from it, additionally illuminating the objects. The part of the collimated beam that reflects from the mirror M_1 forms a reference beam. Recording medium *H* (PFG-03M holographic photoplates) was placed in the plane where the reference and object beams meet. Split diaphragm *S*, which cuts a narrow horizontal strip from the incident beams, was fixed in front of the photoplate. As a result, holograms had the form of 1.5- to 2-mm-high (wide) strips of length to 75 mm. Narrower strips had a worse recording resolution. The photoplate was moved vertically relative to the horizontal slot, so that a set of line holograms was recorded. Up to 40 holograms were recorded on a 9×12-cm photoplate, each registering either a change in the mutual arrangement of the objects or the appearance of a new one. The line holograms were reconstructed with a beam in the form of a luminous line, and the reconstructed image was directed on holographic screen *H* recorded as shown in Fig. 1. The 3D scene was observed in the plane of holographic screen *H*, and the horizontal motion of the eye created the effect of an exploded view of the scene. In the horizontal direction, the viewing angle of the scene was somewhat restricted both by the geometry of the recording scheme and by the small effective size of the hologram due to the insufficient intensity of the exposing radiation.

The screen-projected image, while recovering such a 3D effect as an exploded view, is slightly defocused at points out of the screen plane (Fig. 4).

Thus, our experiments demonstrated the possibility of projecting a 3D image with a one-component screen made using simple technology.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (grant no. 01-02-17854) and by the Scientific School (grant no. 00-15-96771).

REFERENCES

1. Y. N. Denisyuk, V. B. Markov, and N. M. Ganzherli, Proc. SPIE **3011**, 45 (1997).
2. Yu. N. Denisyuk, V. B. Markov, and N. M. Ganzherli, Opt. Spektrosk. **84**, 104 (1998) [Opt. Spectrosc. **84**, 94 (1998)].
3. Yu. N. Denisyuk, N. M. Ganzherli, V. V. Orlov, *et al.*, Opt. Spektrosk. **86**, 864 (1999) [Opt. Spectrosc. **86**, 775 (1999)].

Translated by V. Isaakyan