

Reflection of a Longitudinal Strain Soliton from the End Face of a Nonlinearly Elastic Rod

G. V. Dreiden, A. V. Porubov, A. M. Samsonov, and I. V. Semenova

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

Received July 4, 2000

Abstract—Reflection of a solitary longitudinal strain wave (soliton) from the end face of a nonlinearly elastic rod is investigated theoretically and experimentally. It is shown that the wave reflected from the free end of the rod has a reversed amplitude sign, which results in dispersion of the wave. If the end of the rod is fixed, the reflected wave retains its polarity and properties of the incident solitary wave and propagates back to the input end. © 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

A soliton is a localized long nonlinear quasi-steady-state wave propagating either along the interface of two media (a wave of elevation or an internal wave) or inside a waveguide (a density soliton). Such a wave can propagate over long distances and transfer energy without distortion of the waveform. In a homogeneous isotropic rod, the soliton propagates without distortion, while in the presence of inhomogeneities, its form is distorted. In particular, focusing of the wave may take place, which is an increase of the amplitude of the solitary strain wave accompanied by a decrease of its half-width. This process can initiate regions of plasticity or cracks in a waveguide. Hence, there is the practical importance of investigating the properties of solitary strain waves in solving the problems of the strength of materials, nondestructive testing, and determining the elastic properties of materials.

In our previous theoretical and experimental studies, we demonstrated the possibility of the propagation of a solitary strain wave through a rod with a constant cross section [1, 2], as well as wave focusing in a tapered rod [3]. In the references cited here, the wave processes were investigated apart from the waveguide ends, which is in an infinite rod approximation. However, in real designs, the finite sizes of the wave-guiding components, as well as the way in which the component are coupled with other parts of the unit, are important. Therefore, it is important to study the reflection of a solitary strain wave from the rod end face under various fixation conditions.

THEORY OF REFLECTION OF A LONG NONLINEAR STRAIN WAVE FROM THE END FACE OF AN ELASTIC ROD

(1) Formulation of the problem and derivation of the model equation. Consider a semi-infinite homoge-

neous isotropic rod of a nonlinearly elastic compressible material in a system of cylindrical Lagrangian coordinates (x, r, φ) with the x axis coinciding with the rod axis, and $\varphi \in [0, 2\pi]$, $0 \leq r \leq R$, and $-\infty \leq x \leq L$, where R and L are the rod radius and length, respectively. With no rod torsion, the displacement vector of the rod points is $\mathbf{V} = (u, w, 0)$. The basic equations describing the evolution of nonlinear waves in a strainless structure can be derived using the Hamilton variational principle

$$\delta S = \sigma \int_{t_0}^{t_1} dt \left[2\pi \int_{-\infty}^L dx \int_0^R r \mathcal{L} dr + A \right] = 0, \quad (1)$$

where δS is the variation of the stress functional and A is the work performed by an external force on the rod end, $x = l$.

The inner integration procedure in (1) is performed for a time instant $t = t_0$, and the initial rod state is supposed to be natural. The density of the Lagrangian \mathcal{L} is found as the difference between the density K of the kinetic energy and the volume density Π of the internal energy under adiabatic deformation:

$$\mathcal{L} = K - \Pi = \frac{\rho_0}{2} \left[\left(\frac{\partial u}{\partial t} \right)^2 + \left(\frac{\partial w}{\partial t} \right)^2 \right] - \Pi(I_k). \quad (2)$$

Here, ρ_0 is the rod material density at the time instant $t = t_0$ and I_k ($k = 1, 2, 3$) are the invariants of the Cauchy–Green tensor C of finite deformations. To approximate the specific strain energy, let us choose the Murnaghan model appropriate for a wide range of nonlinearly elastic materials [4]:

$$\Pi = \frac{\lambda + 2\mu}{2} I_1^2 - 2\mu I_2 + \frac{l + 2m}{3} I_1^3 - 2m I_1 I_2 + n I_3. \quad (3)$$

The coefficients in (3) depend on the second-order moduli of elasticity or the Lamé coefficients (λ, μ), and on the third-order or Murnaghan moduli (l, m, n).

Our objective is to study the nonlinear longitudinal waves with a characteristic amplitude of B and wavelength L , such that the strain wave is elastic ($B \ll 1$) and sufficiently long ($R/L \ll 1$). These constraints allow simplifications to be made by relating the longitudinal and transverse displacements, u and w , respectively. Let us derive these relations taking into account that at the free lateral surface of the rod (at $r = R$), at any instant there must be no normal or tangential stresses and the corresponding components P_{rr} and P_{rx} of the Piola-Kirchhoff tensor P must be zeros [4]. The condition of smallness of the wave amplitude $B \ll 1$ allows one to consider linear and nonlinear terms in the expressions for u and w separately, whereas expansion into series in powers of r is apparent in the limit of long waves. Then, we have

$$u = u_L + u_{NL},$$

$$u_L = u_0(x, t) + ru_1(x, t) + r^2u_2(x, t) + \dots, \quad (4)$$

$$u_{NL} = u_{NL_0}(x, t) + ru_{NL_1}(x, t) + \dots,$$

$$w = w_L + w_{NL},$$

$$w_L = w_0(x, t) + rw_1(x, t) + r^2w_2(x, t) + \dots, \quad (5)$$

$$w_{NL} = w_{NL_0}(x, t) + rw_{NL_1}(x, t) + \dots$$

Substituting (4) and (5) into the conditions $P_{rr} = 0$ and $P_{rx} = 0$ at $r = R$ and equating the summands with the same powers of r to each other, we obtain

$$u = U(x, t) + \nu r^2 U_{xx} + \dots, \quad (6)$$

$$w = -\nu r U_x - \frac{\nu^2}{2(3-2\nu)} r^3 U_{xxx} - \left(\frac{\nu(1+\nu)}{2} + \frac{(1-2\nu)(1+\nu)}{E} [l(1-2\nu)^2 + 2m(1+\nu) - n\nu] \right) r U_x^2 + \dots, \quad (7)$$

where E is Young's modulus and ν is the Poisson coefficient.

The higher-order terms can be found in a similar way. However, they do not alter the form of the resulting equation for the nonlinear longitudinal strain waves that will be derived below; therefore, they are omitted here. The equation for longitudinal strain waves in the dimensional form is derived with the help of the Hamilton principle (1). Having substituted (6) and (7) into the expressions for K and Π , respectively, we obtain

$$K = \frac{\rho_0}{2} (U_t^2 + \nu r^2 [U_t U_{xxt} + \nu U_{xt}^2]) + \dots, \quad (8)$$

$$\Pi = \frac{1}{2} \left(E U_x^2 + \frac{\beta}{3} U_x^3 + \nu E r^2 U_x U_{xxx} \right) + \dots, \quad (9)$$

where $\beta = 3E + 2l(1-2\nu)^3 + 4m(1+\nu)^2(1-2\nu) + 6n\nu^2$ is the nonlinearity factor.

Substituting these expressions into (1) and putting $\delta S = 0$, we obtain a nonlinear equation with two dispersions (ETD) for the strain wave $\nu = U_x$:

$$\nu_{tt} - \frac{E}{\rho_0} \nu_{xx} - \left(\frac{\beta}{2\rho_0} (\nu^2) + aR^4 \nu_{tt} - bR^2 \nu_{xx} \right)_{xx} = 0, \quad (10)$$

where $a = -(\nu(1-\nu))/2$, $b = -\nu E/2\rho_0$.

Also, from the Hamilton principle, the boundary conditions with respect to x namely follow ν , and its derivatives with respect to x and t tend to zero when $x \rightarrow -\infty$. At the other end, $x = X$ and the conditions depend on fixation. If the rod end is free, then the work performed by external forces is zero and we obtain from (1)

$$\nu = 0, \quad \nu_{xx} = 0. \quad (11)$$

If the rod end is fixed, the work of external forces is not defined, and on the basis of kinematic reasoning, the displacement and displacement velocity are taken equal to zero:

$$U = 0, \quad U_x = 0, \quad (12)$$

or, in terms of strain,

$$\nu_x = 0, \quad \nu_{xt} = 0. \quad (13)$$

(2) Asymptotic consideration of the interaction between solitary waves. To investigate the reflection of a nonlinear wave, let us reduce ETD (10) to a nondimensional form having chosen as the scales L for x , L/c_0 for t , and B/L for ν , where $c_0 = \sqrt{E/\rho_0}$ is the so-called rod velocity. Let us introduce a small parameter ϵ such that the nonlinear and dispersive terms in the ETD are of the same order of magnitude:

$$\epsilon = B = \frac{R^2}{L^2} \ll 1. \quad (14)$$

Let us assume that the function ν depends, in addition to x and t , on a slow time $\tau = \epsilon t$. We will seek a solution to the nondimensional Eq. (10) in the form of a series in integer powers of ϵ :

$$\nu = \nu_0 + \epsilon \nu_1 + \dots \quad (15)$$

Substituting (15) into Eq. (10) and equating the terms with the same powers of ϵ leads to the D'Alembert solution for ν_0 :

$$\nu_0 = \nu_{01}(\Theta, \tau) + \nu_{02}(\Psi, \tau), \quad (16)$$

where $\Theta = x + t$ and $\Psi = x - t$.

In ε approximation we obtain the equation

$$2v_{01, \Theta\psi} = 2v_{01, \Theta\tau} - 2v_{02, \psi\tau} + \frac{\beta}{2E}((v_{01}^2)_{\Theta\Theta} + 2v_{01, \Theta}v_{02, \psi} + (v_{02}^2)_{\psi\psi}) + \frac{v^2}{2}(v_{01, \Theta\Theta\Theta\Theta} + v_{02, \psi\psi\psi\psi}). \quad (17)$$

The condition of the absence of secular terms leads to two Korteweg–de Vries equations (KdV) for the functions v_{01} and v_{02} ,

$$2v_{01, \tau} - \frac{\beta}{2E}(v_{01}^2)_{\Theta} - \frac{v^2}{2}v_{01, \Theta\Theta\Theta} = 0, \quad (18)$$

$$2v_{02, \tau} - \frac{\beta}{2E}(v_{02}^2)_{\psi} - \frac{v^2}{2}v_{02, \psi\psi\psi} = 0, \quad (19)$$

while the solution of the problem to a first approximation (17) is

$$v_1 = \frac{\beta}{E}v_{01}v_{02} + v_{11}(\Theta, \tau) + v_{12}(\psi, \tau). \quad (20)$$

Substituting the known soliton solutions of the KdV equations [5] into the expression for a zeroth approximation, we obtain

$$v_0 = \frac{6Ev^2}{\beta}k^2[\cosh^{-2}(k[x + (1 + \varepsilon v^2 k^2)t - x_{01}]) + \cosh^{-2}(k[x - (1 + \varepsilon v^2 k^2)t - x_{02}])]. \quad (21)$$

From (21) it follows that, regardless of the propagation direction, the type of the strain wave (compression or tension) is governed only by the sign of the nonlinearity coefficient β , which depends on the elastic properties of the rod material. Both conditions (13) for the fixed end can be satisfied at $x_{02} = 2X - x_{01}$. In this case, reflection of the solitary strain wave occurs with no change of its form. On the contrary, soliton solution (21) does not satisfy the conditions (11) for a free rod end. Therefore, the wave reflected from the rod end will not retain its shape, but will either disappear due to dispersion or will overturn because of the nonlinearity. These processes can possibly be described either in the framework of an asymptotic solution using the approach of the inverse scattering problem to the KdV equation [5] or by numerical simulation of the initial nonintegrable ETD (10).

(3) Numerical study of the wave reflection from the end face of the rod. Numerical simulation of the ETD was performed using a conventional implicit finite-difference scheme. The set of algebraic equations resulting from the finite-difference approximation was solved using the procedure of Gelfand and Lokutsievskii [6], also known as the Thomas algorithm. Numerical solutions of the ETD for an infinite rod of invariable cross section were previously obtained using this approach in [7].

Implementation of the boundary conditions at the end of a rod of length X is effected by means of symmetrical continuation of the calculation area beyond the rod end face, so that this area would occupy the interval $0 < x < 2X$. If the rod end is free, then in the interval $X < x < 2X$, the ETD is calculated using the nonlinearity factor β opposite in sign to that used in the interval $0 < x < X$. The initial conditions are postulated in the form of solitons located centrosymmetrically relative to the real rod end at $x = X$. The type of the initial soliton is chosen depending on the sign of β . So, if $\beta > 0$ in the interval $0 < x < 2X$, then we have a tension soliton, whereas an initial compression soliton takes place in the interval $X < x < 2X$. The initial velocities of the solitons are taken to be directed towards each other. The initial condition $v(x = X) = 0$ is used in explicit form while the condition of the zero second derivative is realized automatically in the calculation scheme. Results on the soliton reflection from the free end of the rod $0 < x < X$ are presented in Fig. 1. The right-hand side of the figure corresponds to the free end. The lowermost profile corresponds to the initial time instant. It is seen that the amplitude of the solitary wave propagating from left to right decreases as the wave reaches the end of the rod. The reflected wave has the opposite polarity. First, its amplitude grows, then the wave is dispersed. As a result, no localized strain wave is observed near the input end of the rod.

In the case of the fixed rod end, an asymmetrical continuation into the interval $0 < x < 2X$ is implemented. In both intervals, ETDs with the same coefficients are calculated with the initial solitons of the same

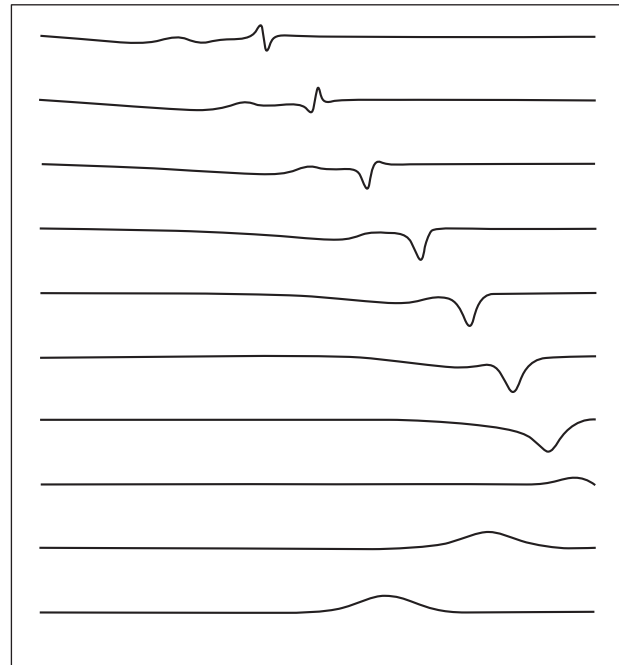


Fig. 1. Numerical simulation of the decay of a longitudinal strain soliton reflected from the free end of the rod.

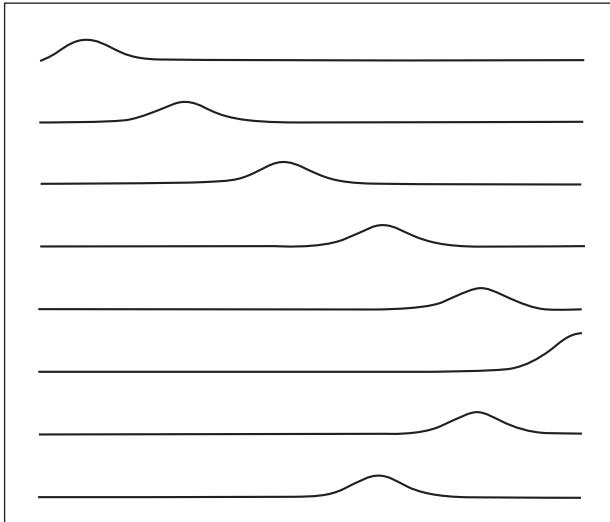


Fig. 2. Numerical simulation of the reflection of a longitudinal strain soliton reflected from the fixed end of the rod.

polarities at equal distances from the point $x = X$. In this case, the boundary conditions (13) are automatically satisfied. Results of the numerical simulation are presented in Fig. 2. The amplitude of the soliton is seen to be nearly twice as large when it reaches the rod end.

The reflected wave has the same amplitude and velocity as the incident one. Keeping its shape, the reflected wave propagates towards the input end of the rod.

EXPERIMENT

(1) Experimental technique and processing of the experimental data. As the waveguide material, SD-3 transparent polystyrene was chosen. The optical scheme of the experimental setup for the generation and recording of strain solitons was already described in our previous papers [2, 3]. The setup consists of a device for generating the initial shock wave, a holographic interferometer for recording the wave pattern, a synchronization unit, and a laser radiation energy meter. The probing light beam of the interferometer has a diameter of 50 mm.

The soliton parameters were calculated using data obtained from processing of the holographic interferograms. Note that the interference pattern is not the usual image of a strain soliton because this soliton is a longitudinal density wave. However, the amplitude of the strain soliton can be calculated using the shift Δk of the interference fringes [3]

$$v = \frac{\Lambda \Delta k}{2h[(n_1 - 1)(1 - 2v) + v(n_1 - n_0)]}. \quad (22)$$

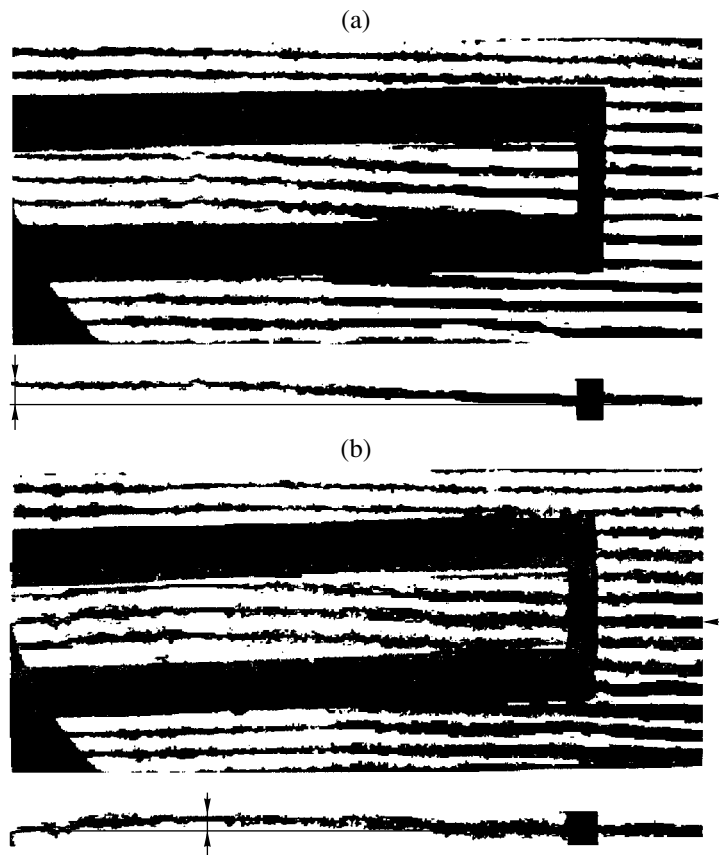


Fig. 3. Experimental observation of the decay of a longitudinal strain soliton reflected from the free end of the rod.

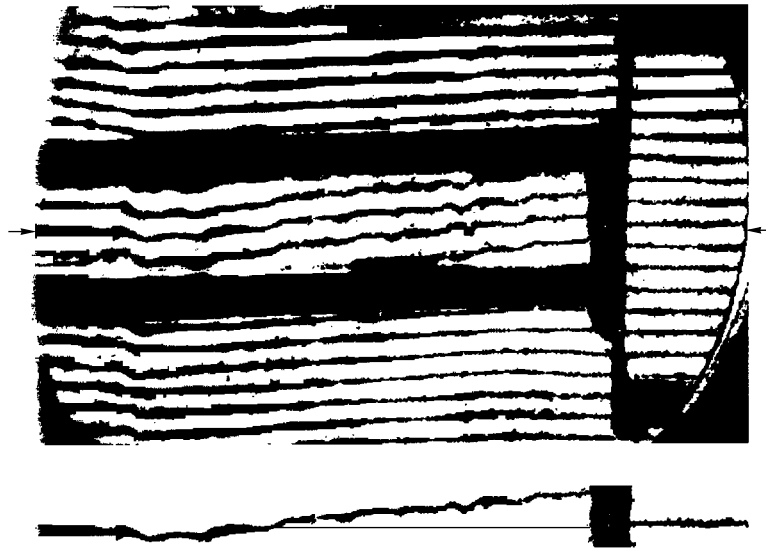


Fig. 4. Total reflection of a strain wave from the rod end attached to a quartz plate.

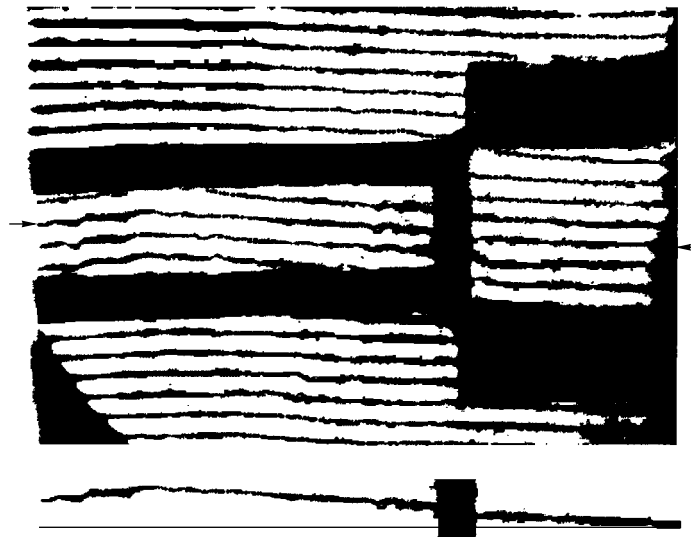


Fig. 5. Penetration of a strain wave through a polystyrene-plexiglass interface.

Here, $2h$ is the light path transverse to the rod, and n_0 and n_1 are the refractive indices of the water and the strainless rod, respectively. The amplitude is determined from the maximum shift of the fringes. The derivation of relation (22) shows that the characteristic length L of a solitary pulse (soliton) can be directly determined from an interferogram as the length of the interference fringe distortion. In our experiments, the shape of the longitudinal strain wave is reconstructed using the following parameters: $n_0 = 1.33$, $n_1 = 1.6$, $\Lambda = 7 \times 10^{-7}$ m, and $\nu = 0.35$ [8]. Due to the negative refractive index β for polystyrene, the upward fringe shift from the undisturbed fringe position corresponds, according to (22), to compression of the rod material.

(2) Observation of the soliton reflection from the rod end face. Figure 3 shows holographic interferograms of the reflection of a longitudinal-strain soliton from the free end. The interference fringes from the liquid beyond the rod end remain undisturbed (horizontal), demonstrating total reflection of the incident wave. At the bottom of the figure panels there are fragments of the interferograms containing single fringes chosen for the analysis. In perfect agreement with the theory, in the fragment at the bottom of Fig. 3b, the amplitude of the initial incident compression wave is seen to be less as compared with a similar fragment in Fig. 3a. The experiment demonstrates the absence of any localized waves near the input rod end. This confirms the theoretical conclusion of the impossibility of propagation of a

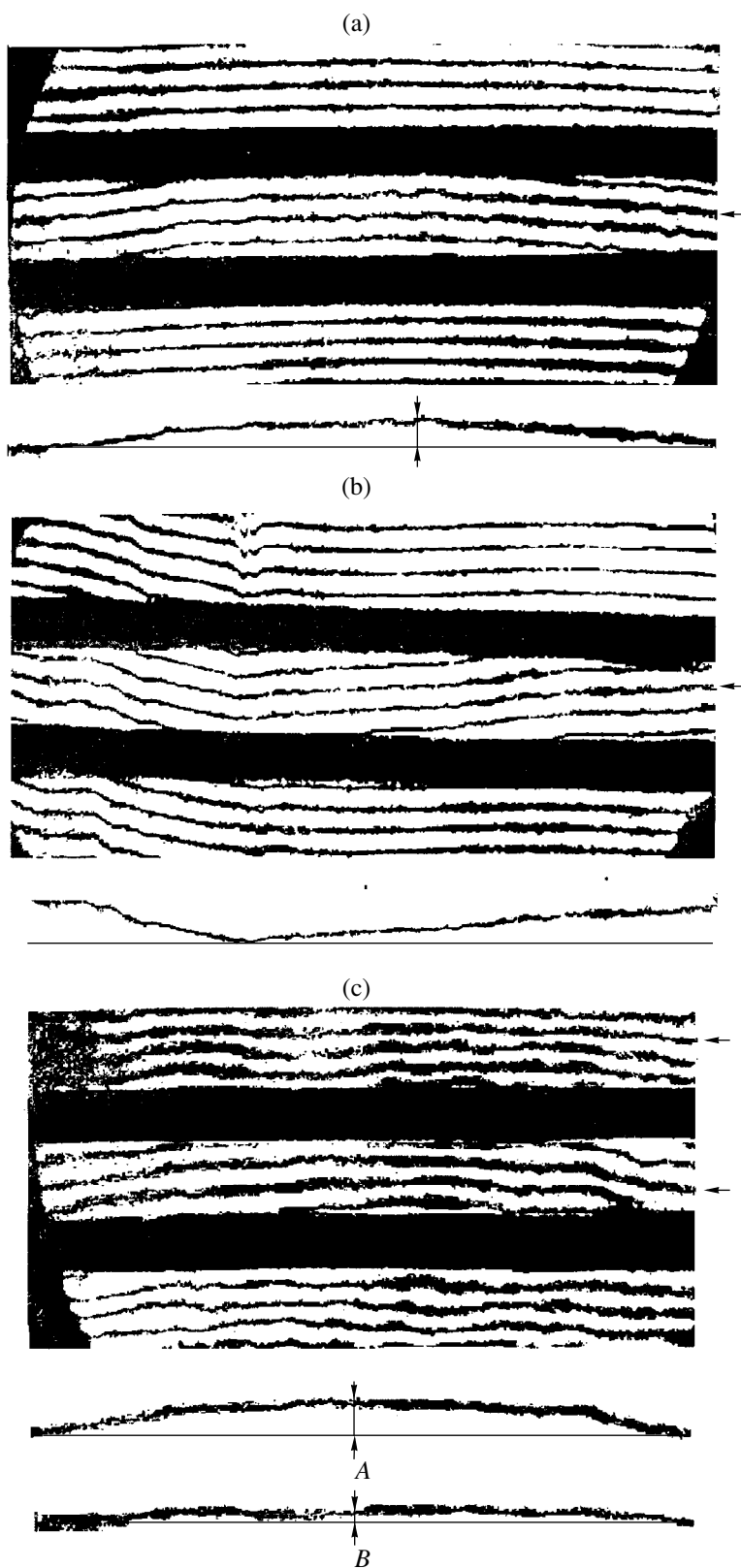


Fig. 6. Experimental observation of the reflection of a longitudinal strain soliton from the fixed end of the rod.

localized tension wave through polystyrene. Positions in the interferograms of the interference fringes displayed separately in this and subsequent figures are indicated by horizontal arrows.

In the experiment, the rod end fixing was implemented by gluing it to a massive brass plate. To determine whether or not the wave could propagate beyond the fixed end, we repeated the experiment with a plate of transparent quartz whose acoustic resistance is six times that of polystyrene. In the fragment of the holographic interferogram in Fig. 4, it is clear that the compression wave does not penetrate into the quartz from the polystyrene as evidenced by the horizontal interference fringes from the quartz plate corresponding to an undisturbed state. At the same time, replacing the quartz by plexiglass with an acoustic resistance close to that of polystyrene, we observed an unattenuated passage of the wave beyond the rod end, as is seen in the fragment in Fig. 5. The acoustic resistance of brass is 15 times that of polystyrene. Therefore, one can presume that the wave is reflected from the rod end face with no loss of energy.

Figure 6 demonstrates the wave reflection from the rod end attached to a massive brass plate; the right side of the interferograms corresponds to the fixed rod end. It is seen that the amplitude of the initial compression soliton propagating from the input end, shown also in the fragment in Fig. 6a, has almost doubled by the fixed rod end (Fig. 6b). Figure 6c shows the reflected compression wave at a distance of 140 mm from the fixed end. At this point, the water surrounding the rod is disturbed; therefore, the optical path difference caused by the strain wave itself can be found by subtracting the fringe shifts in the image outside the rod (B in Fig. 6c) from shifts measured for the rod (A in Fig. 6c). The result is that the amplitude of the reflected solitary strain wave is practically equal to that of the incident wave in Fig. 6a.

CONCLUSION

For the first time, reflection of a solitary strain wave from the free and fixed ends of a nonlinearly-elastic rod has been studied. It has been shown that the amplitude of the wave reflected from the free end changes sign

and the reflected wave is dispersed. If the end is fixed, the reflected wave has the same polarity as the incident one and, keeping all the properties of the incident solitary wave, propagates towards the input rod end.

It is demonstrated analytically and numerically that in our experiments, the reflection of a localized wave from the fixed end occurs in perfect accordance with the theory of collisions of solitary waves. Consequently, the observed wave exhibits the property of the soliton to keep its shape after collisions, which serves as one more confirmation that the waves we previously observed are in fact strain solitons. In addition, the applicability of the ETD model for description of the evolution of solitary strain waves in a nonlinearly elastic compressible rod has been confirmed.

ACKNOWLEDGMENTS

This work was supported by INTAS (grant no. 99-0167).

REFERENCES

1. A. M. Samsonov, Dokl. Akad. Nauk SSSR **299**, 1083 (1988) [Sov. Phys. Dokl. **33**, 298 (1988)].
2. G. V. Dreiden, A. V. Porubov, A. M. Samsonov, *et al.*, Pis'ma Zh. Tekh. Fiz. **21** (11), 42 (1995) [Tech. Phys. Lett. **21**, 415 (1995)].
3. A. M. Samsonov, G. V. Dreiden, A. V. Porubov, and I. V. Semenova, Phys. Rev. B **57**, 5778 (1998).
4. A. I. Lur'e, *Nonlinear Theory of Elasticity* (Nauka, Moscow, 1980).
5. M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (SIAM, Philadelphia, 1981; Mir, Moscow, 1987).
6. S. K. Godunov and V. S. Ryaben'kiĭ, *Difference Schemes* (Nauka, Moscow, 1973).
7. E. V. Sokurinskaya, Candidate's Dissertation (St. Petersburg, 1991).
8. I. N. Frantsevich, F. F. Voronov, and S. A. Bakuta, *Elastic Constants and Modules of Elasticity of Metals and Nonmetals* (Naukova Dumka, Kiev, 1982).

Translated by N. Mende

Inelastic Processes in the Collision of Fast Multiply Charged Ions with a Hydrogen Molecule

V. I. Matveev, V. A. Pazderskiĭ, and Kh. Yu. Rakhimov

Department of Thermal Physics, Academy of Sciences of Uzbekistan, Tashkent, 700135 Uzbekistan

e-mail: khamdam@hpd.silk.org

Received May 11, 1999; in final form, June 30, 2000

Abstract—The collision of a fast (including relativistic) multiply charged ion with the simplest (hydrogen) molecule is considered on the basis of the eikonal approximation. The analytical formula for the reaction cross section, i.e., the total cross section of all inelastic electron processes in the domain where the Born approximation is inapplicable, is derived. The cross section is compared to the corresponding doubled inelastic cross sections for collisions of multiply charged ions with hydrogen atoms calculated within perturbative and nonperturbative approaches. © 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

Studies of elementary processes arising in the interaction of atomic systems with fast multiply charged ions, which have received considerable attention in the last few years (see, e.g., [1, 2] and recent reviews [3–5] with their references), are of interest from the standpoint of both general physics and their possible practical application. In the interaction of multiply charged ions with atomic systems, the arising effective electric fields may substantially exceed the characteristic atomic field strength. Studying atom and molecule behavior in such superintense fields is one of the topical problems of modern physics. The applicability condition of the Born approximation $Z/v \ll 1$ (Z and v are the ion charge and velocity, respectively; hereafter we employ the atomic system of units) may be violated in the case of the collision of an atom or molecule with an ion of large charge even for relativistic velocities. This makes it impossible to use the perturbation theory for calculating cross sections for various elementary processes. In such cases, various nonperturbative methods are used: the sudden perturbation approximation [6–9], the eikonal approximation (Glauber approximation) and its various modifications [10–14], numerical methods [15], and the recently found exact solution of the Dirac equation in the ultrarelativistic limit [16]. A considerable part of nonperturbative calculations are confined to collisions of fast multiply charged ions with atoms, whereas collisions with molecules are considered more rarely [1–5] because of the cumbersome analytical calculations and time-consuming numerical simulations. However, from the practical point of view, it is precisely studies of collisions between fast ions and molecules that are of considerable interest.

Whereas collisions of fast protons and ions of small charges with molecules can be considered within the ordinary perturbation theory (see, e.g., [17–20]), in the

case of collisions between molecules and fast, heavy, multiply charged ions, when $Z \sim v$ and the Born approximation is no longer valid, one should use nonperturbative methods. If the perturbation theory is employed for calculating the cross sections for electron transitions in inelastic collisions of fast charged particles with multielectron systems, one-electron excitations and ionization are the first-order effects with respect to the interaction between the incident particle and atomic electrons [3]. Two-electron transition corresponds to the second order of the perturbation theory, when the interaction of the incident particle with electrons and interelectron interaction are taken into account only once each. However, the situation evidently changes when the interaction of electrons with an incident particle is much stronger than the interelectron one, as it takes place in the interaction of an atom or molecule with a multiply charged ion. In this case, a multielectron transition should be considered as a result of direct action of the incident-particle strong field [3, 21]. The above-mentioned nonperturbative methods correspond to just this mechanism of direct excitation.

In this paper, we consider the collision of a fast (including relativistic) multiply charged ion with the simplest (hydrogen) molecule on the basis of the eikonal approximation. The analytical formula for the reaction cross section, i.e., the total cross section of all inelastic electron processes in the domain where the Born approximation is inapplicable, is derived. The cross section is compared to the corresponding doubled inelastic cross sections for collisions of multiply charged ions with hydrogen atoms calculated according to perturbative and nonperturbative approaches.

GENERAL FORMALISM

In the Glauber approximation, the general expression for the amplitude of an inelastic collision of a fast

(including relativistic) multiply charged ion with a non-relativistic (both before and after collision)¹ system of particles (atom or molecule), which is accompanied by the transition of the system from a state with a wave function $|\Psi_i\rangle$ to a state with a wave function $|\Psi_f\rangle$, can be represented in the form (compare to [12, 13, 22])

$$f_{if}(\mathbf{q}) = \frac{ik_i}{2\pi} \int e^{-i\mathbf{q}\mathbf{b}} \langle \Psi_f | \left[1 - \exp \left\{ -\frac{i}{v} \int U dx \right\} \right] | \Psi_i \rangle d^2 b, \quad (1)$$

where $\mathbf{q} = \mathbf{k}_f - \mathbf{k}_i$ is the change in the ion momentum, \mathbf{b} is the impact parameter, and the x axis is directed along the initial ion momentum \mathbf{k}_i .

By implication of the Glauber approximation, the scattering potential U in (1) is the sum of electrostatic Coulomb potentials produced by static (during the collision) electrons and nuclei of the molecule. The scattering potential U is a function of ion coordinates $\mathbf{R}_{\text{ion}} = (x, \mathbf{b})$ and the instantaneous positions of a molecule's nuclei and electrons, whose coordinates are denoted by $\{\mathbf{R}_A\}$ ($A = 1, 2, \dots, N$, where N is the number of a molecule's nuclei) and $\{\mathbf{r}_a\}$ ($a = 1, 2, \dots, n$, where n is the number of electrons), respectively. Thus, $U = U(x, \mathbf{b}; \{\mathbf{r}_a\}, \{\mathbf{R}_A\})$. A feature of collisions between multiply charged ions, on the one hand, and atoms and molecules, on the other hand, is that the cross sections of inelastic processes are rather large and generally exceed atomic sizes substantially. Therefore, the range of large impact parameters makes the major contribution to the cross sections. Using the small-angle approximation for large impact parameters (as in [22]), the integration over the ion scattering angles can be performed in the corresponding expression for the cross section. Finally, for the cross section of the transition considered, we find

$$\sigma = \int d^2 b \left| \langle \Psi_f | \left[1 - \exp \left\{ -\frac{i}{v} \int U dx \right\} \right] | \Psi_i \rangle \right|^2. \quad (2)$$

It is clear that the expression $\left| \langle \Psi_f | \exp \left\{ -\frac{i}{v} \int U dx \right\} | \Psi_i \rangle \right|^2$ in (2) should be interpreted as the probability of the transition of a molecule from a state with a wave function $|\Psi_i\rangle$ to a state with a wave function $|\Psi_f\rangle$ in its collision with an ion moving along a straight path with the impact parameter \mathbf{b} . It is important that the expression for the transition probability satisfies the unitarity condition, i.e., the sum of the

probabilities of transitions to all possible final states (a complete set) is evidently equal to unity.

Let us discuss the contribution to the cross section of the interaction between an incident ion and the molecule's nuclei. In a collision having an impact parameter b ($b \gg 1$), one nucleus gains a momentum of $q_1 \sim 2Z/(vb)$ and another, which is spaced at a distance $\sim R$ from the first one, gains a momentum of $q_2 \sim q_1 + 2RZ/(vb^2)$. The difference between these momenta, i.e., the relative momentum, is $\delta q \sim 2RZ/(vb^2) \ll 1$ for $b \gg 1$. Since $\delta q^2/(2M) \ll \omega_0$ (M is the mass of the molecule's nucleus, and ω_0 is the natural oscillation frequency; in our case of a hydrogen molecule, M is the proton mass and $\omega_0 \sim 0.02$), the vibrational excitation of nuclei due to the interaction of an incident ion with the molecule's nuclei can be neglected (compared to characteristic cross sections of electron transitions). The excitations of rotational degrees of freedom can be neglected for similar reasons. Indeed, the angular momentum transferred to a molecule $\sim \delta q R \sim 2Z/(vb^2)R^2 \ll 1$ (compare to analogous estimations in [23]). Thus, we can keep only the interaction of the incident ion with electrons in the scattering potential U and consider the nuclei positions to be fixed, i.e., $U = U(x, \mathbf{b}; \{\mathbf{r}_a\})$. Then,

$$\frac{1}{v} \int_{-\infty}^{+\infty} U dx = \sum_{a=1}^N \chi_a(\mathbf{b}, \mathbf{s}_a),$$

$$\chi_a(\mathbf{b}, \mathbf{s}_a) = \frac{2Z}{v} \ln \frac{|\mathbf{b} - \mathbf{s}_a|}{b},$$

where the two-dimensional vector \mathbf{s}_a lies in the plane perpendicular to the direction of ion motion and has the coordinates $\mathbf{s}_a = (y_a, z_a)$.

Generally speaking, the integral with respect to the impact parameter in (2) diverges at large impact parameters.² However, such a divergence turns out to be insignificant [4, 8, 13]: for large impact parameters, the ion field is small and the Born approximation is applicable with overlapping ranges of applicability of the Born and eikonal approximations. This makes it possible to match the correctly corresponding cross section with respect to the impact parameter. Let us denote by b_0 the upper limit of integration with respect to the parameter b in (2). For large $b \gg s$ and orthogonal states $|\Psi_f\rangle$ and $|\Psi_i\rangle$, the generalized inelastic form factor (the

² Note that (just as in the case of collisions between ions and atoms [13]), if in the collision of a fast ion with a molecule either the states of more than one electron change or transitions are forbidden in the dipole approximation, the integration with respect to the impact parameter in (2) can be extended to the whole plane of the impact parameter (since the integrand provides the convergence of the integral) and the need for matching with the result obtained by the perturbation theory disappears.

¹ Strictly speaking, electrons or nuclei of a molecule thrown into the continuum as a result of a collision with a relativistic ion may gain relativistic velocities. However, such processes occur at small impact parameters and the corresponding contribution to the total cross section can be neglected (compare to [4]).

amplitude of inelastic transition)

$$\begin{aligned} & \langle \Psi_f | 1 - \exp \left\{ -i \frac{2Z}{v} \ln \frac{|\mathbf{b} - \mathbf{s}|}{b} \right\} | \Psi_i \rangle \\ & \approx \langle \Psi_f | \exp \{ i \mathbf{q} \mathbf{s} \} | \Psi_i \rangle, \quad \mathbf{q} = \frac{2Z\mathbf{b}}{vb^2} \end{aligned} \quad (3)$$

tends to $i\mathbf{q} \langle \Psi_f | \mathbf{s} | \Psi_i \rangle$ for small \mathbf{q} and, hence, the integral (2) with respect to d^2b depends logarithmically on b_0 . Therefore, the contribution of the region $b < b_0$ to the cross section can be represented as (compare to [13])

$$\sigma(b < b_0) = 8\pi \frac{Z^2}{v^2} \lambda_{if} \ln \frac{2\alpha_{if}}{q_0}, \quad q_0 = 2Z/(vb_0), \quad (4)$$

where the variables λ_{if} and α_{if} depend only on the properties of the electron states of the molecule and are independent of the impact parameters, namely, the incident ion charge and its velocity.

In the range of large $b > b_0$, we will use the Born approximation to calculate the corresponding cross section, since the ion field in this case is a small perturbation. We will calculate σ in each of the impact-parameter ranges and obtain the total cross section by summing the contributions of each range. In doing so, the exact value of b_0 is insignificant,³ since the b_0 -dependence of σ turns out to be logarithmical. This results in a correct matching of the contributions of each range and dropping out of the b_0 -dependence in the final result.

TOTAL INELASTIC CROSS SECTION

Consider the collision of a fast multiply charged ion with a hydrogen molecule in the reference frame in which the molecule's center of mass is at rest at the origin. The coordinate axes are chosen so that the \mathbf{b} vector is directed along the z axis and the ion velocity \mathbf{v} , along the x axis, as before. We assume that $\mathbf{R}_a = \mathbf{R}/2$ and $\mathbf{R}_b = -\mathbf{R}/2$ are the radius vectors of the static protons (the molecule's nuclei), and $\mathbf{r}_{1,2}$ are the radius vectors of electrons. Then, in the approximation of large impact parameters, the amplitude of the transition of the system from a state with a wave function $|\Psi_0\rangle$ to a state with a wave function $|\Psi_k\rangle$ is given by (see (3)) the expression

$$A_{0k} = \langle \Psi_k | e^{i\mathbf{q}(\mathbf{r}_1 + \mathbf{r}_2)} | \Psi_0 \rangle. \quad (5)$$

Hereafter, we will consider only inelastic processes that are connected with changes in the state of the electron subsystem of a hydrogen molecule and will not touch upon possible processes related to changes in its nucleus subsystem. The cross section of all the inelastic processes that occur in collisions of a multiply charged ion with a hydrogen molecule will be found in the fol-

lowing way. First, we obtain the probability of a collision having an impact parameter b in which the molecule retains its state. In order to find it, one should insert the wave function of the ground state $|\Psi_0\rangle$ instead of $|\Psi_k\rangle$ in (5). We choose the following form for this wave function:

$$\Psi_0(\mathbf{r}_1, \mathbf{r}_2) = N(\varphi(r_{1a})\varphi(r_{2b}) + \varphi(r_{1b})\varphi(r_{2a})), \quad (6)$$

where $\varphi(r)$ are the hydrogen-like functions

$$\varphi(r) = \sqrt{\frac{Z_a^3}{\pi}} e^{-Z_a r}; \quad (7)$$

r_{1a} , r_{1b} , r_{2a} , and r_{2b} are the distances of the first and second electrons to nuclei a and b , respectively; and N is the normalization factor.

At $Z_a = 1$, this function corresponds to the Heitler-London approximation (with a fixed distance R between the nuclei); $Z_a = 1.166$ corresponds to the variational calculation [24] of the ground state of a hydrogen molecule with the wave function (6) for the equilibrium distance between the nuclei, $R = 1.4$. The normalization factor in (6) is conventionally expressed in terms of the overlap integral $S(\rho)$:

$$\begin{aligned} N &= [2(1 + S^2(Z_a R))]^{-1/2}, \\ S(\rho) &= e^{-\rho} \left(1 + \rho + \frac{\rho^2}{3} \right). \end{aligned} \quad (8)$$

The calculation of the transition amplitude yields

$$A_{00} = 2N^2 \{ A_1^2(q) + A_2^2(\mathbf{q}) \}, \quad (9)$$

where

$$\begin{aligned} A_1(q) &= \frac{16Z_a^4}{(q^2 + 4Z_a^2)^2}, \\ A_2(\mathbf{q}) &= 6Z_a^5 \int_{-1/2}^{1/2} dz \left(\frac{1}{4} - z^2 \right) \frac{\cos(z(\mathbf{q}\mathbf{R})) S(\beta R)}{\beta^5}, \quad (10) \\ \beta^2 &= Z_a^2 + q^2 \left(\frac{1}{4} - z^2 \right). \end{aligned}$$

The expression obtained depends on the molecule orientation [through the factor $\cos(z(\mathbf{q}\mathbf{R}))$ in $A_2(\mathbf{q})$]. In order to obtain the probability of the process in which the molecule stays in the ground state in its collision with a multiply charged ion, one should average the squared absolute value of the amplitude A_{00} over angles. This yields an expression involving an integral of the product of four factors of the form $\cos(z_i(\mathbf{q}\mathbf{R}))$ (with different z_i). The expression obtained in this way is rather cumbersome; therefore, we will use the following approximation to get a more compact expression. For large impact parameters, when $q \leq 1$, we may

³ Although, as is done for collisions with atoms [9], similar estimations may be easily obtained for b_0 .

substitute

$$\cos(z(\mathbf{qR})) \cong 1 - \frac{z^2(\mathbf{qR})^2}{2}$$

into (10) and perform averaging over molecule orientations after this substitution.⁴ The final expression for the averaged probability depends on two integrals $I_{1,2}(q)$,

$$I_1(q) = 3 \int_{-1/2}^{1/2} dz \left(\frac{1}{4} - z^2 \right) \frac{S(\beta R)}{\beta^5},$$

$$I_2(q) = 3q^2 R^2 \int_{-1/2}^{1/2} dz z^2 \left(\frac{1}{4} - z^2 \right) \frac{S(\beta R)}{\beta^5},$$

which may be calculated numerically.

The final expression for the averaged probability is

$$\begin{aligned} \overline{W_{00}(q)} &= 4N^4 \left\{ A_1^4(q) + 8Z_a^{10} A_1^2(q) \right. \\ &\times \left[I_1^2(q) - \frac{1}{3} I_1(q) I_2(q) + \frac{1}{20} I_2^2(q) \right] \\ &+ 16Z_a^{20} \left[I_1^4(q) - \frac{2}{3} I_1^3(q) I_2(q) + \frac{3}{10} I_1^2(q) I_2^2(q) \right. \\ &\left. \left. - \frac{1}{14} I_1(q) I_2^3(q) + \frac{1}{144} I_2^4(q) \right] \right\}. \end{aligned} \quad (11)$$

Then, the reaction probability is $W_r(q) = 1 - \overline{W_{00}(q)}$. After multiplying $W_r(q)$ by $2\pi b db$ and performing integration with respect to the impact parameter from zero to b_0 , we find $\sigma_r(b < b_0)$, i.e., the reaction cross section in the impact-parameter range $0 < b < b_0$, which makes a major contribution to the total inelastic cross section. However, such a procedure can be directly performed only numerically. Therefore, in order to obtain an analytical expression for the cross section, we employ the method described in [13] and represent the inelastic cross section in form (4) with $\lambda_{if} = \lambda_r$ and $\alpha_{if} = \alpha_r$. In this expression, the parameter λ_r is defined by the asymptotic of the probability $W_r(q)$ at small q

$$W_r(q) = \lambda_r q^2, \quad q \rightarrow 0 \quad (12)$$

and is equal to 2.098 for $Z_a = 1.0$ and 1.555 for $Z_a =$

⁴ We numerically calculated the cross section in view of the subsequent terms of the cosine expansion and found that the corrections were $\leq 1\%$.

1.166. Formula (5) shows that

$$\begin{aligned} \lambda_r &= \sum_n \left| \langle \Psi_n | (z_{1a} + z_{2b}) | \Psi_0 \rangle \right|^2 \\ &= \langle \Psi_0 | (z_{1a} + z_{2b})^2 | \Psi_0 \rangle, \end{aligned} \quad (13)$$

where the summation \sum_n is performed over the complete set.

The value α_r has been obtained numerically by the formula (compare to [13])

$$\alpha_r = \lim_{b_0 \rightarrow \infty} \frac{Z}{v b_0} \exp \left\{ \frac{1}{\lambda_r} \frac{v^2}{8\pi Z^2} \int_0^{b_0} 2\pi b db W_r(q) \right\} \quad (14)$$

and it is equal to 0.415 for $Z_a = 1.0$ and 0.482 for $Z_a = 1.166$. Thus, the contribution of the impact-parameter range $b < b_0$ to the cross section has the form

$$\sigma(b < b_0) = 8\pi \frac{Z^2}{v^2} \lambda_r \ln \frac{\alpha_r v b_0}{Z}, \quad (15)$$

where α_r does not depend on the parameters of the collision v and Z , just as in the case of collisions with atoms [13].

At large impact parameters $b > b_0$, the ion field is weak and the Born approximation may be used for calculating the cross sections. In the first order of the perturbation theory, the amplitude of the molecule transition from the initial state $|\Psi_0\rangle$ into the final $|\Psi_k\rangle$ state in the time-variable field of the moving multiply charged ion is given by the expression (see, e.g., [4, 12, 13])

$$\begin{aligned} A_{0k}^{(1)} &= -i \frac{2Z}{v^2} \omega_{k0} [i(z_{1a} + z_{2b})_{k0} \sqrt{1 - \beta^2} K_1(\xi) \\ &+ (x_{1a} + x_{2b})_{k0} (1 - \beta^2) K_0(\xi)], \end{aligned} \quad (16)$$

where ω_{k0} is the transition frequency; $K_0(x)$ and $K_1(x)$ are the McDonald functions; and $\xi = \omega_{k0} b \sqrt{1 - \beta^2} / v$, $\beta = v/c$, and $(z_{1a})_{k0} = \langle \Psi_k | z_{1a} | \Psi_0 \rangle$, etc. are the dipole matrix elements of the transition.

Having summed $|A_{0k}^{(1)}|^2$ over all final states k and performed integration with respect to the impact parameter b from b_0 to ∞ , we find the contribution of the range of large impact parameters to the total inelastic cross section. In the expression for the cross section, besides the squared z and x components of the dipole moment, the cross terms involving their product will appear. However, they vanish after averaging and, hence, we will not take them into account. In view of the fact that the following equality is satisfied for the squared z and x components of the dipole moment averaged over molecule orientations,

$$\overline{|(z_{1a} + z_{2b})_{k0}|^2} = \overline{|(x_{1a} + x_{2b})_{k0}|^2},$$

the reaction cross section is given by the formula

$$\sigma_r(b > b_0) = 8\pi \frac{Z^2}{v^2} \sum_k \overline{|(z_{1a} + z_{2b})_{k0}|^2} \times \left[\xi_0 K_0(\xi_0) K_1(\xi_0) - \frac{1}{2} \xi_0^2 \beta^2 (K_1^2(\xi_0) - K_0^2(\xi_0)) \right], \quad (17)$$

where $\xi_0 = \omega_{k0} b_0 \sqrt{1 - \beta^2} / v$.

In the limit of large velocities $\xi_0 \rightarrow 0$, taking into account the asymptotics [25] of the McDonald functions for $x \rightarrow 0$,

$$K_0(x) \rightarrow -B + \ln \frac{2}{x}; \quad x K_1(x) \rightarrow 1; \quad x K_0(x) \rightarrow 0, \quad (18)$$

σ_r takes the form of the so-called Bethe asymptotic

$$\sigma_r = 8\pi \frac{Z^2}{v^2} \sum_k \overline{|(z_{1a} + z_{2b})_{k0}|^2} \times \left(\ln \frac{2v}{\eta b_0 \omega_{k0} \sqrt{1 - \beta^2}} - \frac{\beta^2}{2} \right). \quad (19)$$

Here, $\eta = e^B = 1.781$ ($B = 0.5772$ is the Euler constant). We define the mean transition frequency $\bar{\omega}$ by the relation

$$\ln \bar{\omega} = \frac{\sum_k \overline{|(z_{1a} + z_{2b})_{k0}|^2} \ln \omega_{k0}}{\sum_k \overline{|(z_{1a} + z_{2b})_{k0}|^2}} \quad (20)$$

with the quantity $\sum_k \overline{|(z_{1a} + z_{2b})_{k0}|^2} = \lambda_r$ determined above [see (13)]. Generally, formula (20) makes it possible to calculate the value of $\bar{\omega}$. However, this requires knowing the wave functions and energies of all the electron states of a molecule. Therefore, in view of the weak (logarithmic) $\bar{\omega}$ -dependence of the cross section, we estimate it in the following way. Since we consider collisions with large impact parameters, the field of the molecule's nuclei can be described by the Coulomb field of a point charge Z_a . In this case, the value of $\bar{\omega}$ can be calculated exactly (compare to [13]) and is equal to $0.465 Z_a^2$, where 0.465 is the value of mean transition frequency in the hydrogen atom.

Then, the final expression for the contribution of the range $b > b_0$ to the cross section of all inelastic processes is given by

$$\sigma_r(b > b_0) = 8\pi \frac{Z^2}{v^2} \lambda_r \left(\ln \frac{2v}{\eta b_0 \bar{\omega} \sqrt{1 - \beta^2}} - \frac{\beta^2}{2} \right). \quad (21)$$

Summing the cross sections $\sigma_r(b < b_0)$ and $\sigma_r(b > b_0)$, we obtain the total inelastic cross section

$$\sigma_r = 8\pi \frac{Z^2}{v^2} \lambda_r \left(\ln \frac{2v^2 \alpha_r}{\eta Z \bar{\omega} \sqrt{1 - \beta^2}} - \frac{\beta^2}{2} \right). \quad (22)$$

Note that the b_0 -dependence disappears after matching. In the nonrelativistic limit, the obtained expression for the cross section satisfies the well-known empirical scaling law [1, 26]

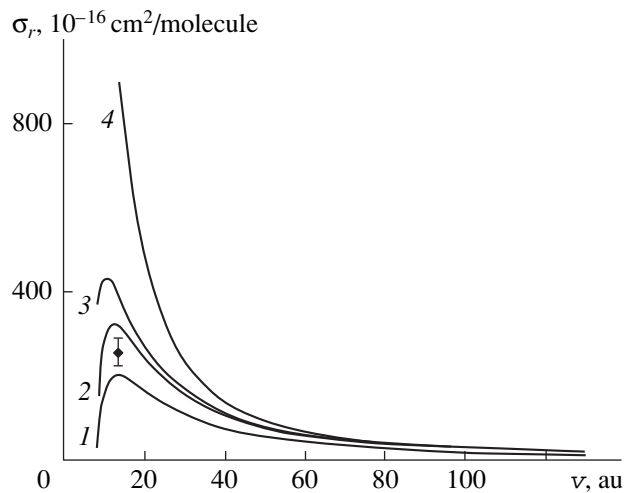
$$\sigma = ZQ(E/Z), \quad (23)$$

where E is the energy of an incident ion divided by the unit of atomic mass and Q is a universal function.

Generally, in measurements and calculations of the cross sections of inelastic processes (σ_{H_2}) accompanying collisions of fast charged particles with a hydrogen molecule, the cross sections are compared to the corresponding doubled cross sections ($2\sigma_H$) for collisions of the same particles with a hydrogen atom [27]. Note that deviations from this rule are registered both experimentally and theoretically (in calculations in the Born approximation) [17]. Such a comparison has not been performed yet for calculations by nonperturbative

methods in spite of the fact that the estimate $\sigma_H \approx \frac{1}{2} \sigma_{H_2}$ of the cross sections for collisions of hydrogen atoms with fast multiply charged ions is widely used (see, e.g., [26, 27]) in the domain where the perturbation theory is not valid.

The figure shows the result of calculating the total inelastic cross section by formula (22) for $Z_a = 1.166$ and 1, as well as the doubled reaction cross sections for



Reaction cross section (22) obtained in this paper for $Z_a = (1) 1.166$ and $(2) 1$; (3) doubled cross section for a hydrogen atom in eikonal approximation (24) and (4) in Born approximation (25). The experimental value of the cross section σ_{loss} was obtained in [26].

Cross sections for collisions of Pb^{Z+} ions of energy $E = 4.65$ MeV/nucleon with molecular hydrogen (in units of 10^{-16} cm²/molecule)

Z	σ_{loss}	σ_{r1}	σ_{r2}	σ_{re}	σ_{rB}
52	220_{-30}^{+24}	178	274	330	722
53	220_{-33}^{+24}	182	280	338	750
54	220_{-33}^{+24}	186	287	347	779
55	230_{-35}^{+25}	189	293	356	808
57	250_{-38}^{+28}	196	306	373	868
59	260_{-39}^{+30}	203	318	391	930

Note: σ_{loss} is the experimental data [26]; σ_{r1} and σ_{r2} are the calculations by (22) with $Z_a = 1.166$ and 1, respectively; σ_{re} is the calculation by (24); and σ_{rB} is the calculation by (25).

the collision of fast multiply charged ions with hydrogen atoms in the eikonal approximation [13]

$$\sigma_{re} = 2 \times 8\pi \frac{Z^2}{v^2} \left(\ln \frac{1.4v^2}{Z\sqrt{1-\beta^2}} - \frac{\beta^2}{2} \right) \quad (24)$$

and in the Born approximation (the Bethe asymptotic) [22]

$$\sigma_{rB} = 2 \times 4\pi \frac{Z^2}{v^2} \left(\ln \frac{v^2}{0.16(1-\beta^2)} - \beta^2 \right) \quad (25)$$

for $Z = 59$ and $E > 1$ MeV/nucleon. Here, the experimental value for σ_{loss} obtained in [26] for the collision of a Pb^{59+} ion at $E = 4.65$ MeV/nucleon ($v = 13.6$) with molecular hydrogen is shown. The table lists the experimental data obtained in [26] for σ_{loss} in collisions of Pb^{Z+} ions of various multiplicity with molecular hydrogen at $E = 4.65$ MeV/nucleon. Here, the inelastic cross sections calculated by formulas (22), (24), and (25) are also presented.

The results of our calculations presented in the figure and table illustrate the corresponding errors ($\sigma_{H_2} = 2\sigma_H$) and make it possible to draw the following conclusions. The correct allowance for the strong ion field yields a substantial difference between the cross sections calculated in the Born approximation and in the eikonal approximation in the cases of collisions of multiply charged ions with both atoms and molecules. The effects resulting from allowing for the molecular structure of electron states may result in significant (more than 30%) deviations from the prevailing rule of estimation, according to which the cross section for a collision with a two-atom molecule is approximately equal to the sum of the corresponding cross sections for

isolated atoms. The violation of this rule increases with decreasing relative collision velocity.

REFERENCES

1. L. P. Presnyakov, V. P. Shevel'ko, and R. K. Yanev, *Elementary Processes with the Participation of Multiply Charged Ions* (Énergoatomizdat, Moscow, 1986).
2. J. Eichler and W. E. Meyrhof, *Relativistic Atomic Collisions* (Academic, New York, 1995).
3. J. H. McGuire, *Adv. At., Mol., Opt. Phys.* **29**, 217 (1992).
4. V. I. Matveev, *Fiz. Élem. Chastits At. Yadra* **26**, 780 (1995) [*Phys. Part. Nucl.* **26**, 329 (1995)].
5. J. Eichler, *Phys. Rep.* **193** (4–5), 167 (1990).
6. J. Eichler, *Phys. Rev.* **15** (5), 1856 (1977).
7. G. L. Yudin, *Zh. Éksp. Teor. Fiz.* **80** (3), 1026 (1981) [*Sov. Phys. JETP* **53**, 523 (1981)].
8. A. Salop and J. Eichler, *J. Phys. B* **12** (2), 257 (1979).
9. V. I. Matveev and M. M. Musakhanov, *Zh. Éksp. Teor. Fiz.* **105** (2), 280 (1994) [*JETP* **78**, 149 (1994)].
10. J. H. McGuire, *Phys. Rev. A* **26** (1), 143 (1982).
11. D. S. F. Grothers and S. H. McCann, *J. Phys. B* **16** (17), 3229 (1983).
12. V. I. Matveev and S. G. Tolmanov, *Zh. Éksp. Teor. Fiz.* **107** (6), 1780 (1995) [*JETP* **80**, 989 (1995)].
13. V. I. Matveev and Kh. Yu. Rakhimov, *Zh. Éksp. Teor. Fiz.* **114**, 1646 (1998) [*JETP* **87**, 891 (1998)].
14. P. K. Khabibullaev, V. I. Matveev, and D. U. Matrasulov, *J. Phys. B* **31** (14), L607 (1998).
15. U. Becker, N. Grün, W. Scheid, and G. Soff, *Phys. Rev. Lett.* **56** (19), 2016 (1986).
16. A. J. Baltz, *Phys. Rev. Lett.* **78** (7), 1231 (1997).
17. L. Nagy and L. Vegh, *Phys. Rev. A* **46** (1), 284 (1992).
18. L. Nagy and L. Vegh, *Phys. Rev. A* **46** (1), 290 (1992).
19. M. Inokuti, *Rev. Mod. Phys.* **43** (3), 297 (1971).
20. L. Nagy and L. Vegh, *Phys. Rev. A* **50** (5), 3984 (1994).
21. J. H. McGuire, A. Mueller, B. Shuch, *et al.*, *Phys. Rev. A* **35** (6), 2479 (1987).
22. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics, Vol. 3: Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989, 4th ed.; Pergamon, New York, 1977, 3rd ed.).
23. A. B. Migdal, *Qualitative Methods in Quantum Theory* (Nauka, Moscow, 1975; Benjamin, Reading, Mass., 1977).
24. S. C. Wang, *Phys. Rev.* **28**, 279 (1928).
25. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, 1964; Nauka, Moscow, 1979).
26. A. S. Schlachter, K. H. Berkner, W. G. Graham, *et al.*, *Phys. Rev. A* **24** (2), 1110 (1981).
27. M. E. Rudd, Y.-K. Kim, D. H. Madison, and J. V. Gallagher, *Rev. Mod. Phys.* **57** (4), 965 (1985).

Translated by M. Fofanov

Multistream Instability of Crossed Electron Beams in a Plasma

A. E. Dubinov

All-Russia Research Institute of Experimental Physics, Russian Federal Nuclear Center, Sarov,
Nizhni Novgorod oblast, 607188 Russia

e-mail: dubinov@ntc.vniief.ru

Received February 21, 2000; in final form, May 3, 2000

Abstract—Multistream instabilities of crossed electron beams in a plasma are investigated analytically. By using successively more complicated models, the instability of a finite number of crossed monoenergetic beams and the instability of an infinite number of crossed monoenergetic beams with different distributions in the azimuthal angle are studied. The conclusion is drawn that an isotropic plasma is unstable against the excitation of longitudinal waves. © 2001 MAIK “Nauka/Interperiodica”.

The familiar theory of the two-stream instability of electron beams in a plasma [1] can naturally be generalized to the multistream instability theory, which, on the one hand, serves as a discrete model of the kinetic instability of a plasma with a continuous electron distribution function and, on the other hand, describes actual experiments on the injection of many electron beams into a plasma.

There are a number of papers aimed at theoretical [2, 3] and experimental [4] investigations of the multistream instability of a system of parallel electron beams with different densities and different electron velocities. However, it is also of interest to study the multistream instability of nonparallel beams. For this reason, our main objective here is to develop a model of the instability of a finite number of crossed beams. We then improve our model step by step in order to investigate the instability of an infinite number of crossed monoenergetic beams with different distributions in the azimuthal angle and the instability of an isotropic plasma against the excitation of longitudinal waves. Note that the latter problem is analogous to the problem treated by Weibel [5] in connection with the instability of a plasma with an anisotropic electron temperature (see also [6]). The formalism developed here is also applicable to one-dimensional isotropic electron distributions that appear in microwave [7] or optical gas breakdowns [8, 9] in superstrong fields such that the electron oscillatory energy is much higher than the ionization energy of the atoms.

We start by considering a finite number of straight monoenergetic electron beams that are unbounded in the transverse direction and propagate against the background of immobile neutralizing ions. Let the veloci-

ties of all of the beams lie in the same plane (Fig. 1). We use the dispersion relation derived by Stix [3]:

$$1 = \sum_{j=1}^N \frac{\omega_{bj}^2}{(\omega - \mathbf{k}\mathbf{v}_j)^2} = \sum_{j=1}^N \frac{\omega_{bj}^2}{(\omega - k v_j \cos \theta_j)^2}, \quad (1)$$

where ω_{bj} is the plasma frequency of the j th beam, \mathbf{v}_j is the velocity vector of the electrons from the j th beam, N is the number of beams, ω and \mathbf{k} are the frequency and wave vector of the excited wave, and θ_j is the angle between the vectors \mathbf{v}_j and \mathbf{k} .

It is convenient to rewrite Eq. (1) multiplied by k^2 as

$$k^2 = \sum_{j=1}^N \frac{\omega_{bj}^2}{(v_{ph} - v_j \cos \theta_j)^2}, \quad (2)$$

where v_{ph} is the wave phase velocity.

Equation (2) can be analyzed in a standard way. As an example, we consider eight electron beams with the same velocity ($v_0 = v_j$) and the same density ($\omega_{b0} = \omega_{bj}$). Let the beams divide the total angle 2π about the point of their intersection into eight equal parts (Fig. 2a). The plot of the function $k^2(v_{ph})$ for this beam

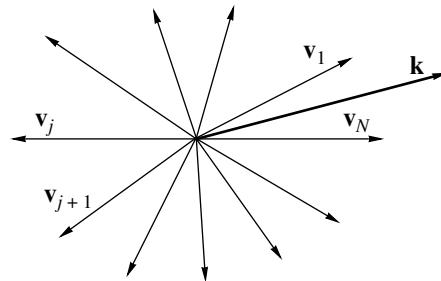


Fig. 1. General configuration of crossed electron beams.

system is presented in Fig. 3a and is seen to be similar in shape to the corresponding plots obtained in [2, 3] for a system of parallel monoenergetic electron beams with different velocities. However, the plots shown in Fig. 3a and calculated in [2] differ in that the envelope of the local minima of the function $k^2(v_{ph})$ in Fig. 3a (the dashed curve) is convex downward, while the corresponding envelope in the model of [2] is convex upward. This discrepancy can be explained as follows. In our model, the interaction of the excited wave with the beams that are nearly parallel to the wave propagation direction is more efficient than with the beams that are nearly perpendicular to the wave propagation direction. In contrast, in the model of parallel beams [2], the excited wave interacts most efficiently with an electron beam whose velocity is closest to the average velocity of the beams.

The roots of Eq. (2) can be readily found by intersecting the plot in Fig. 3a by a horizontal straight line $k^2 = \text{const}$. If the line passes below some of the local minima of the function $k^2(v_{ph})$, then Eq. (2) has complex roots whose imaginary parts determine the instability growth rates.

We can easily see that, for sufficiently small values of k^2 , Eq. (2) always has complex roots. However, in real bounded beam-plasma systems, no instability can occur if their characteristic dimension L is less than the shortest wavelength at which Eq. (2) still has complex roots.

In the general case (and in our case of eight beams), the complex roots cannot be determined analytically. However, they can be either calculated numerically or evaluated graphically using the approximate method described by Smirnov [9].

Let us point out one more feature that should be taken into account when analyzing Eq. (2). With a special choice of the beam velocity vectors, the velocities of some of the beams may have the same projections onto the direction of the wave vector \mathbf{k} . In this case, the vertical asymptotics of the plot of the function $k^2(v_{ph})$ that correspond to the resonances of the excited wave with these beams merge into one, because the corresponding denominators of the fractions in Eq. (2) coincide; the corresponding partial densities in the numerators ω_{bj}^2 are added together. In addition, the values of the local minima near such degenerate resonances change. As a result, the number of resonances decreases; moreover, for a system of N beams distributed uniformly over the azimuthal angle, we can distinguish between two cases: the wave vector of the excited wave is parallel to one of the beams (case A in Fig. 2b) and the wave vector is parallel to the bisector of the angle between the neighboring beams (case B in Fig. 2c). For even and odd values of N , the number of resonances decreases in different ways (see table). As an example, the plot of the function $k^2(v_{ph})$ for $N = 8$ and for case A with five resonances is shown in Fig. 3b.

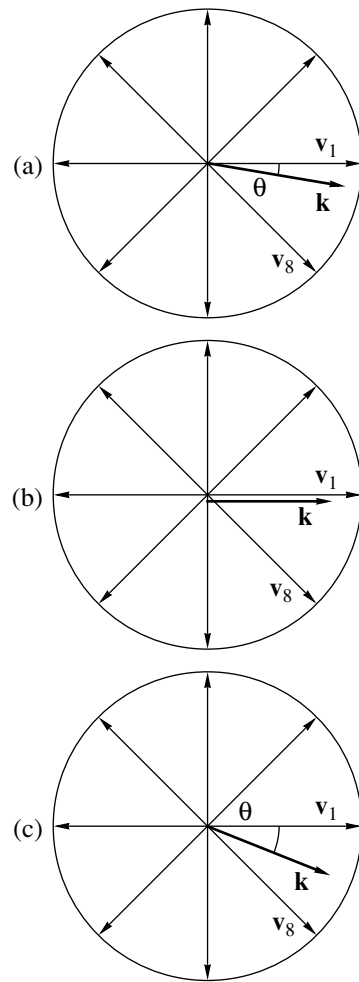


Fig. 2. Configuration of eight electron beams with the same density and velocity: (a) general case, (b) case A, and (c) case B.

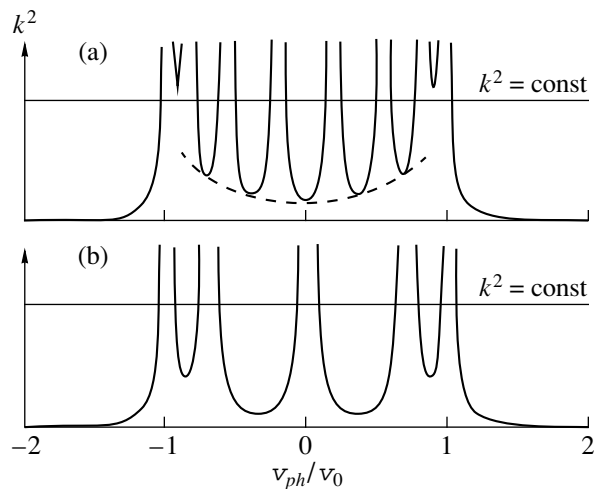


Fig. 3. Plots of the function $k^2(v_{ph})$ (a) in the general case and (b) in case A.

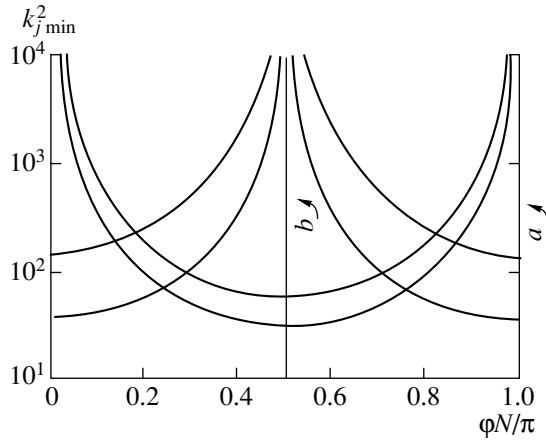


Fig. 4. Dependence of the values of the local minima of the function $k^2(v_{ph})$ on the orientation of the wave vector \mathbf{k} .

It is also of interest to understand how the values of the local minima depend on the orientation of the wave vector \mathbf{k} relative to the beams. For a system of beams distributed uniformly in the azimuthal angle, it is sufficient to consider such angles between the wave vector \mathbf{k} and the velocity \mathbf{v}_j of one of the beams that lie in the range $\varphi \in [0; \pi/N]$, in which the smallest and the largest angles correspond to cases A and B, respectively.

For the above symmetric system of eight electron beams, the dependence of the values of the local minima of the function $k^2(v_{ph})$ on the orientation of the wave vector of the excited wave is illustrated in Fig. 4. This dependence is characterized by four branches, because, by virtue of the symmetry of the beam system, some of the minima are equal in magnitude, so that we can speak of a partial degeneration of the local minima. For a general system of beams distributed arbitrarily in the azimuthal angle, the local minima are nondegenerate; however, one can construct a beam system with a desired (and even fairly high) degree of degeneration.

It is worth noting that, for the wave vector \mathbf{k} oriented approximately in the same manner as in cases A and B, some of the values of the local minima can be arbitrarily large, thereby providing the possibility of generating very short waves. Physically, this can be explained by the fact that the excited wave resonates

with several (rather than one) beams, the number of beams being determined by the degree of degeneration.

Now, we proceed to an analysis of an infinite number of crossed beams distributed continuously in the azimuthal angle. We assume that a two-dimensional space is penetrated by electron beams with the same speed v_0 but with different densities. We describe the continuous angular distribution of the beams by the directionality function $\omega_p^2 f(\theta)$. We also assume, as before, that the space charge of the beams is neutralized by an immobile ion background.

The wave properties of such a beam system are governed by the complex dielectric function

$$\epsilon = 1 - \frac{\omega_p^2}{k^2} \int_0^{2\pi} \frac{f(\theta) d\theta}{(v_{ph} - v_0 \cos \theta)^2}. \tag{3}$$

Mathematically, the problem reduces to that of calculating the integral in expression (3) and determining the imaginary part of ϵ . For a number of particular directionality functions, this problem can be solved exactly. The results of relevant calculations carried out for different orientations of the wave vector \mathbf{k} are presented in the Appendix and are illustrated in Fig. 5.

The most interesting case is that in which the directionality function is uniform. Calculations show that, as expected, the imaginary part of the dielectric function (3) is nonzero under the Cherenkov radiation condition for longitudinal waves:

$$|v_{ph}/v_0| < 1. \tag{4}$$

The excitation of a longitudinal wave gives rise to the Cherenkov instability. Note that this situation is fairly general and does not require that the system be anisotropic, as is the case in [5]. It is also an easy matter to generalize this example with an isotropic directionality function and monoenergetic electron beams to the case of isotropic discrete or continuous electron distribution functions (see below).

All of the examples presented below refer to anisotropic directionality functions, so that the results of calculations depend strongly on the orientation of the wave vector \mathbf{k} relative to the directionality function $f(\theta)$.

In particular, for a single-lobed directionality function, we consider three different directions of the wave vector \mathbf{k} . For the transverse direction of \mathbf{k} (\mathbf{k}_2), the imaginary part of the dielectric function is exactly zero. For the longitudinal direction of \mathbf{k} (\mathbf{k}_4), there is only direct Cherenkov resonance; in this case, passing over to the frame of reference that moves at a speed $v_0/2$ with respect to the original frame, we arrive at the above example with an isotropic directionality function and the electron velocity $v_0/2$. For the wave vector \mathbf{k} that bisects the angle between the transverse and longitudinal directions, there is also Cherenkov resonance at

The number of resonances for different orientations of the wave vector \mathbf{k}

General case	Even N	Odd N
	N	N
Case A	$\frac{N+2}{2}$	$\frac{N+1}{2}$
Case B	$\frac{N}{2}$	$\frac{N+1}{2}$

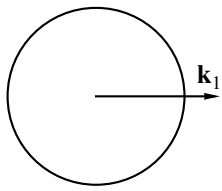
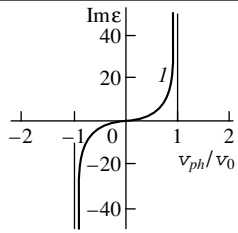
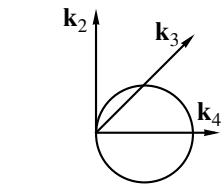
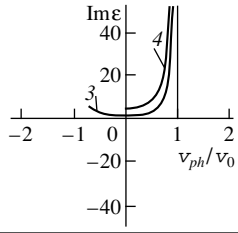
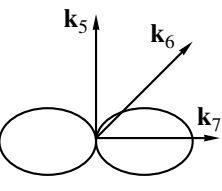
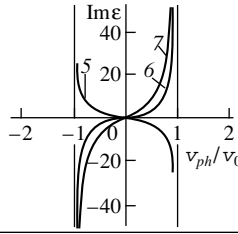
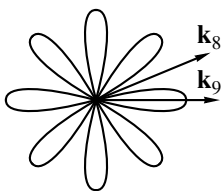
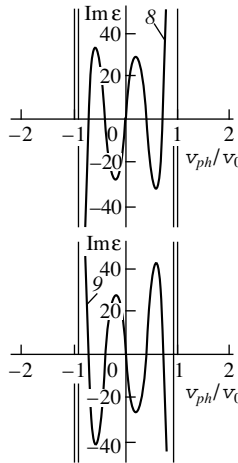
$f(\theta)$ 1	$f(\theta); \mathbf{k}$ 2	Equation (3); A. 3	$\text{Im}\epsilon(v_{ph})$ 4
$f(\theta) = \text{const}$		$\mathbf{k}_1:(A.1)$	
$f(\theta) = \cos\theta$		$\mathbf{k}_2:(A.2)$ $\mathbf{k}_3:(A.3)$ $\mathbf{k}_4:(A.4)$	
$f(\theta) = \cos^2\theta$		$\mathbf{k}_5:(A.5)$ $\mathbf{k}_6:(A.6)$ $\mathbf{k}_7:(A.7)$	
$f(\theta) = \cos^2 4\theta$		$\mathbf{k}_8:(A.8)$ $\mathbf{k}_9:(A.9)$	

Fig. 5. Results of the calculations of $\text{Im}\epsilon(v_{ph})$. (1) The directionality function $f(\theta)$: $f(\theta) = \text{const}$ is an isotropic function, $f(\theta) = \cos\theta$ is a single-lobed function, $f(\theta) = \cos^2\theta$ is a double-lobed function, and $f(\theta) = \cos^2 4\theta$ is a function with eight lobes. (2) Polar diagrams of the function $f(\theta)$ and the orientations of the wave vector \mathbf{k} . (3) Formulas derived in the Appendix for the exact values of the integral in expression (3). (4) Plots of the function $\text{Im}\epsilon(v_{ph})$.

$|v_{ph}/v_0| = 1$, in the case of which the function $\text{Im}\epsilon(v_{ph})$ has finite-amplitude jumps at $v_{ph}/v_0 = \pm\sqrt{2}/2$.

The case with a double-lobed directionality function is a generalization of the case of a two-stream instability, and the case with eight lobes is a generalization of the above model of eight unstable electron beams. These examples, which are illustrated in Fig. 5 for several different directions of the wave vector \mathbf{k} , are also characterized by the Cherenkov instability; moreover,

in the case with eight lobes, the imaginary part $\text{Im}\epsilon(v_{ph})$ is an alternating-sign function.

In conclusion, we briefly discuss the electron beam distribution functions that are not monoenergetic. For an electron distribution function that is discrete in velocity and uniform in the azimuthal angle, we have to analyze the imaginary part of the sum in the expression

$$\epsilon = 1 - \sum_{j=1}^N \frac{\omega_{pj}^2}{k^2} \int_0^{2\pi} \frac{d\theta}{(v_{ph} - v_j \cos\theta)^2}. \quad (5)$$

Clearly, the sum in formula (5) is equal to the sum of the terms of the form of (A.1), and the imaginary part $\text{Im}(\epsilon)$ is nonzero only in the velocity range

$$\left| \frac{v_{ph}}{\max\{v_j\}} \right| < 1. \quad (6)$$

Analogous considerations are also valid for isotropic electron distribution functions that are continuous in velocity. This situation can be analyzed by simply switching from summation to integration over v in expression (5), in the case of which we can also arrive at the situation with $\text{Im}(\epsilon) \neq 0$.

Hence, we started by developing a model of the multistream instability of a finite number of crossed beams and then developed successively more complicated models. In this way, we arrived at the conclusion that a plasma with an isotropic and continuous (e.g.,

Maxwellian) electron distribution function can be unstable against longitudinal waves.

ACKNOWLEDGMENTS

I am grateful to A.A. Rukhadze for valuable advice and to Yu.B. Kudasov for help in this work.

APPENDIX

Here, we present the exact values of the integrals I_j from Eq. (3); the integrals are numbered by j in the same manner as in the table:

$$I_1 = -2\pi i v_{ph} \frac{C}{R^{3/2}}, \quad (A.1)$$

$$I_2 = -\frac{2}{R}, \quad \theta \in [0; \pi], \quad (A.2)$$

$$I_3 = 2 \frac{\{R^{1/2}(-2v_0^4 - 8v_{ph}^2 + 8v_0^2 v_{ph}^2) + \sqrt{2}[A(v_0^5 - 4v_0^3 v_{ph}^2 + 4v_0 v_{ph}^4)]\}}{R^{3/2}(2v_0^2 - 4v_{ph}^2)[2(v_0 + v_{ph})^2 - (v_0 - 2v_{ph})^2]}, \quad (A.3)$$

$$\theta \in \left[-\frac{\pi}{4}; \frac{3\pi}{4}\right],$$

$$I_4 = 2 \frac{2\arctanh\left(\frac{v_0 + v_{ph}}{R^{1/2}}\right) - R^{1/2}}{R^{3/2}}, \quad \theta \in \left[-\frac{\pi}{2}; \frac{\pi}{2}\right], \quad (A.4)$$

$$I_5 = 2 \frac{\pi v_{ph} C - R^{1/2}}{v_0^2 R^{1/2}}, \quad (A.5)$$

$$I_6 = \frac{\pi i v_{ph} C}{R^{3/2}}, \quad (A.6)$$

$$I_7 = -2\pi \frac{iC(v_0^2 v_{ph} - v_{ph}^3) - R^{3/2}}{v_0^2 R^{3/2}}, \quad (A.7)$$

$$I_8 = -8\pi \frac{Z_1}{v_0^8 R^{1/2}}, \quad (A.8)$$

$$I_9 = 2\pi \frac{Z_2}{v_0^8 R^{3/2}}. \quad (A.9)$$

We direct the attention of the reader to the integration limits in formulas (A.2)–(A.4). In formulas (A.1) and (A.5)–(A.9), integration is carried out over the entire azimuthal angle, as is the case in Eq. (3).

In these formulas, we use the following notation: i is the imaginary unit;

$$R = v_0^2 - v_{ph}^2, \quad C = \text{csgn}[i(R^{1/2})^*(v_0 + v_{ph})],$$

the asterisk denoting the complex conjugate; and $\text{csgn}(\dots)$ is a complex sign function defined as

$$\text{csgn}(z) = \begin{cases} 1 & \text{at } (\text{Re}(z) > 0) \cup [(\text{Re}(z) = 0) \cap (\text{Im}(z) \geq 0)] \\ -1 & \text{at } (\text{Re}(z) < 0) \cup [(\text{Re}(z) = 0) \cap (\text{Im}(z) < 0)] \end{cases}$$

and

$$A = \arctanh \frac{(\sqrt{2} - 1)(v_0 - v_{ph})}{R^{1/2}} + \arctanh \frac{(\sqrt{2} + 1)(v_0 + v_{ph})}{R^{1/2}},$$

$$Z_1 = -v_0^6 R^{1/2} + 30v_0^4 v_{ph}^2 R^{1/2} - 120v_0^2 v_{ph}^2 R^{1/2} + 112v_{ph}^6 R^{1/2} + 8iv_0^6 v_{ph} C - 76iv_0^4 v_{ph}^3 C + 176iv_0^2 v_{ph}^5 C - 112iv_{ph}^7 C,$$

$$Z_2 = -4v_0^8 R^{1/2} + 124v_0^6 v_{ph}^2 R^{1/2} - 600v_0^4 v_{ph}^4 R^{1/2} + 928v_0^2 v_{ph}^6 R^{1/2} - 448v_{ph}^8 R^{1/2} + 31iv_0^8 v_{ph} C - 336iv_0^6 v_{ph}^3 C + 1008iv_0^4 v_{ph}^5 C - 1152iv_0^2 v_{ph}^7 C + 448iv_{ph}^9 C.$$

REFERENCES

1. R. Briggs, in *Advances in Plasma Physics*, Ed. by A. Simon and W. B. Thompson (Wiley, New York, 1969, 1971; Mir, Moscow, 1974), Vols. 3, 4.
2. J. M. Dawson, *Phys. Rev.* **118**, 381 (1960).
3. T. H. Stix, *The Theory of Plasma Waves* (McGraw-Hill, New York, 1962; Atomizdat, Moscow, 1965).
4. V. D. Fedorchenko, Yu. P. Mazalov, A. S. Bakaĭ, and B. N. Rutkevich, *Zh. Éksp. Teor. Fiz.* **65**, 2225 (1973) [*Sov. Phys. JETP* **38**, 1111 (1974)].
5. E. S. Weibel, *Phys. Rev. Lett.* **2**, 83 (1959).
6. A. F. Alexandrov, L. S. Bogdankevich, and A. A. Rukhadze, *Principles of Plasma Electrodynamics* (Vysshaya Shkola, Moscow, 1978; Springer-Verlag, Berlin, 1984).
7. S. G. Arutyunyan, A. V. Ignat'ev, and A. A. Rukhadze, *Fiz. Plazmy* **7**, 604 (1981) [*Sov. J. Plasma Phys.* **7**, 331 (1981)].
8. V. V. Korobkin and V. V. Romanovskĭĭ, *Zh. Éksp. Teor. Fiz.* **103**, 1221 (1993) [*JETP* **76**, 597 (1993)].
9. V. M. Smirnov, *Plasma Physics* (Atomizdat, Moscow, 1969), Vol. 2, p. 79.

Translated by G. Shepekina

Gas-Filled Electron Diode Based on a Glow Discharge

V. N. Devyatkov, N. N. Koval, and P. M. Schanin

High Current Electronics Institute, Siberian Division, Russian Academy of Sciences,
Akademicheskii pr. 4, Tomsk, 634055 Russia

e-mail: koval@opee.hcei.tsc.ru

Received June 14, 2000

Abstract—Stable ignition and sustention of a pulsed discharge with a current of up to 180 A and duration of 12 μ s at a pressure of 10^{-1} – 10^{-2} Pa are achieved in a glow-discharge plasma cathode with the help of an auxiliary initiating discharge. An electron emission current density of up to 100 A/cm² and accelerating voltage of 15 kV are obtained in a gas-filled diode based on this type of a plasma cathode. An electron beam with a neutralized space charge can be transported almost without losses in a weak axial magnetic field along a plasma channel formed due to the gas ionization by the accelerated electrons over a distance of up to 30 cm.
© 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

The perveance of gas-filled and plasma-filled diodes is higher than that of vacuum diodes due to the acceleration of electrons in a sheath between the cathode and the anode plasma, which is created either via gas ionization by a beam or using an auxiliary gas discharge. In an electron source with a plasma emitter [1] in which the cathode plasma is produced by an arc discharge, currents of up to 600 A with a pulse duration of 30 μ s at an accelerating voltage of 20 kV and a mean electron emission current density of 12 A/cm² are obtained. As in electron sources with cold cathodes, which provide current densities of up to several kA/cm², the service life of a source with an arc discharge is determined by the cathode lifetime. The use of a glow discharge with a hollow cathode in the plasma emitter lifts these restrictions and substantially increases the service life of the electron source due to the large surface area and weak erosion of the cathode. As was shown in [2], a glow-discharge plasma cathode can provide an emission current density of up to 60 A/cm².

The main disadvantage of a glow discharge is the high pressure required for igniting and sustaining the discharge; this circumstance substantially limits the electric strength of the accelerating gap. The use of an auxiliary initiating discharge makes it possible to significantly decrease the operating pressure and achieve the high electric strength of the accelerating gap [3, 4].

GAS-FILLED DIODE AND ELECTRON-BEAM DIAGNOSTICS

A schematic of the gas-filled diode is shown in Fig. 1. The gas-discharge system of the plasma emitter consists of two glow discharge systems. The first system consists of the cathode 1 and anode 3 and acts as a system for initiating the discharge, which occurs at a

relatively high pressure for a short time (2–3 μ s). The hollow cathode 1 is a 50-mm-long cylinder with an inner diameter of 10 mm. In order to decrease the ignition voltage of the initiating discharge, the cathode 1 is embedded in the 0.1-T magnetic field, which is created

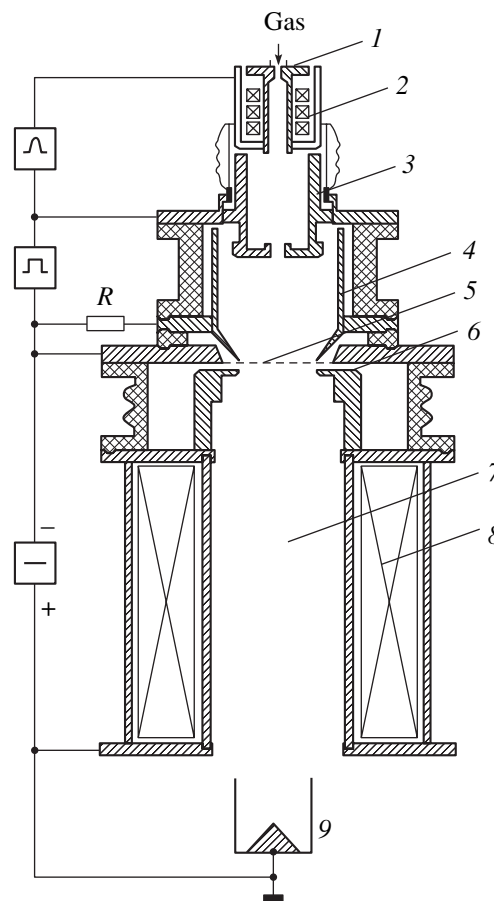


Fig. 1. Schematic of the gas-filled diode.

by ring Sm–Co permanent magnets 2. The main glow discharge with a duration of 12 μs occurs between a 40-mm-diameter and 50-mm-long cylindrical electrode 3, which acts as a hollow cathode in this discharge, and a hollow anode consisting of an 80-mm-diameter and 100-mm-long cylindrical electrode 4 and a grid electrode 5, placed at the end of the anode. To ensure the pressure difference between these discharge systems, the coupling between them occurs through a small contracting hole 6 mm in diameter. In spite of the small diameter of the contracting hole, the hollow cathode effect is ensured by the penetration of the anode potential because the thickness of the space charge sheath created by the plasma of the initiating discharge is fairly small. The presence of the hollow cathode effect in the discharge system consisting of electrodes 3 and 4 is confirmed by the sputtering of the wall of electrode 3. We note that the copper electrode is sputtered more intensely than one of stainless steel.

The electrons are extracted and accelerated by a dc voltage applied between an emissive grid electrode 5 and a cylindrical accelerating electrode 6. In experiments, it was possible to vary the diameters of the openings in the emissive and accelerating electrodes from 10 to 50 mm, the diameter of the latter electrode always being larger than that of the former. The distance between the grid emissive electrode 5 and accelerating electrode 6 was 5 mm.

The parameters of the electron beam were measured after its passage through a 300-mm-long and 80-mm-diameter drift channel 7. The main components of the diagnostic system were a collector 9, a set of Faraday cups with an aperture diameter of 3 mm, and a water-cooled calorimeter. The calorimeter was placed on a movable rod and could be moved along the diode axis without breaking the system sealing. To study the influence of the magnetic field on the transportation of the neutralized electron beam, a 0.015-T axial magnetic field was produced in the drift channel 7 by a solenoid 8.

To ignite the initiating discharge, a 10-kV pulse was applied from the secondary winding of a pulsed transformer; a TGI 500/16 thyatron was used to switch a 6000-pF storage capacitance. The current pulse of the initiating discharge had a half-sinusoidal shape with an amplitude of 20–30 A and a full width at a half-maximum of 3 μs . The main discharge was powered from a pulse-forming line. A voltage of 5–10 kV was switched on by a RU62 controlled discharger at the instant when the initiating discharge current reached its maximum. The duration of the main discharge current pulse was 12 μs , the current amplitude was up to 180 A, and the main discharge voltage was ≤ 1 kV. The pulse repetition rate, which ranged from single pulses to 25 Hz, was determined by the power supply capability. The accelerating gap was powered from a dc power supply with an output voltage of 0–15 kV and a storage capacitance of 5 μF , which enabled the operation with a partial dis-

charge of the capacitance and low variations in the voltage during the beam current pulses.

EXPERIMENTAL RESULTS

After puffing the working gas (air) into the discharge system and applying the corresponding voltages, the auxiliary glow discharge between electrodes 1 and 3 is first ignited. According to both the experiment and calculations, when the working gas is supplied through the hollow cathode 1 at a flow rate of $Q = 5\text{--}20$ mPa $\text{m}^3 \text{s}^{-1}$, the pressure in the discharge gap ranges within several pascals. The presence of the magnetic field in the cathode region of the auxiliary discharge allows us to lower the working pressure required for the auxiliary discharge ignition to 5×10^{-3} Pa. Within the operating pressure range of the gas-filled diode ($10^{-1}\text{--}10^{-2}$ Pa), the auxiliary discharge is ignited at a voltage of 5–10 kV. Then, in a time of ~ 3 μs , this voltage linearly decreases to zero. Depending on the pressure, the main glow discharge between electrodes 3 and 4 is ignited with a delay of 1–2 μs ; then, in 2–3 μs , the discharge current switches to the grid electrode 5 and its amplitude reaches 180 A, the pulse duration being 12 μs . To ensure the discharge switching to the grid electrode 5, this electrode is directly connected to the power supply, whereas it is connected to the cylindrical hollow anode 4 through a resistor ($R_0 = 100$ Ω). After switching, the current to electrode 4 is about several amperes. The main glow discharge voltage depends on the pressure only slightly. Thus, changing the pressure in the system by one order of magnitude (from 10^{-1} to 2×10^{-2} Pa) leads to an increase in the voltage from 800 to 900 V.

The plasma parameters near the grid electrode at the main discharge current of 100 A were measured using a small cylindrical probe for two values of the grid electrode diameter: $d_1 = 5$ cm and $d_2 = 1$ cm. The probe measurements showed that a 20-fold decrease in the anode surface area leads to an increase in the plasma electron density n_e by one order of magnitude (from 5×10^{11} to 10^{13} cm^{-3}) and a decrease in the electron temperature T_e from ~ 15 to ~ 10 eV, which is accompanied by a reversion of the negative anode potential drop to a positive one (from -8 to $+1$ V). At these plasma parameters, according to the formula

$$j_e \approx en_e \left(\frac{kT_e}{2\pi m_e} \right)^{1/2}, \quad (1)$$

where e and m_e are electron charge and mass, n_e is the plasma electron density, and k is the Boltzmann constant, the plasma cathode can provide an emission current density of ~ 12 and ~ 80 A/cm², respectively.

Recall that the extraction and acceleration of electrons is enabled by a dc voltage applied to the 5-mm-long accelerating gap between the 50-mm-diameter

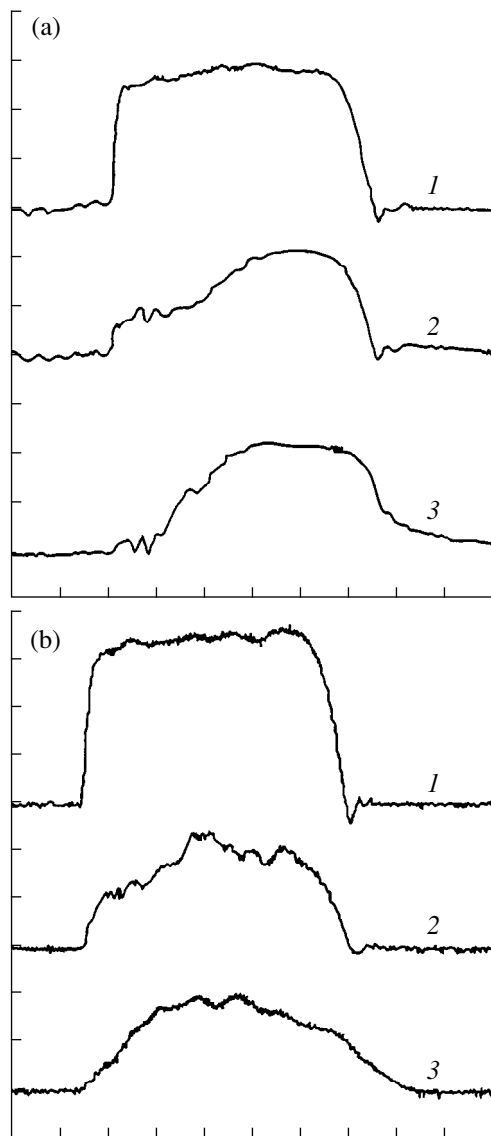


Fig. 2. Time evolution of (1) discharge current, (2) emission current, and (3) collector current in the gas-filled diode for an emissive electrode diameter of (a) 50 and (b) 10 mm. Scales are 40 A/division and 2.5 μ s/division.

electrode 5 and the accelerating electrode 6 with a 55-mm-diameter opening.

Figure 2a presents characteristic waveforms of the gas-filled diode current at an accelerating voltage of 15 kV, a pressure of 2×10^{-2} Pa, and a solenoid magnetic field of 0.015 T. At a discharge current of 120 A, the current in the accelerating gap was ~ 100 A. The beam was transported almost without losses to the collector, which was placed at a distance of 35 cm from the emissive grid. When the axial magnetic field created by solenoid 8 is higher than the azimuthal magnetic self-field of the electron beam (~ 0.01 T), the beam transportation through the drift channel is substantially improved. Although the optical transparency of the grid electrode with a 0.4×0.4 -mm mesh is 40%, the elec-

tron extraction efficiency α , which is defined as the ratio of the beam current I_b to the discharge current I_d , attains $\alpha = I_b/I_d = 0.8$. In the gas-filled diode, the emission current from the grid plasma cathode depends slightly on the accelerating voltage. Thus, the increase in the accelerating voltage from 2 to 15 kV increases the emission current by only 15%. As the pressure is decreased to 10^{-2} Pa, the extraction of electrons and the transportation of the beam become unstable. Both the emission and collector currents significantly decrease, and RF oscillations of the current occur at the leading edge of the pulse. Such an effect of the pressure decrease on the diode operation is explained by the deficit of ions, which are produced due to gas ionization by the electron beam. The deficit of ions impedes the formation of the plasma anode in the anode opening and results in the incomplete neutralization of the beam space charge in the drift channel.

The beam current was measured by a resistive shunt in the collector circuit and by a Rogowski coil. It was found that the measured current depends on the shunt resistance. As the resistance increased, the measured current sharply decreased. When the current flows through the shunt, a reverse-bias voltage arises at the collector; the plasma electron current, which is oppositely directed to the high-energy electron beam current, appears between the collector and the drift chamber. Since the density of the plasma electrons may be an order of magnitude higher than that of the beam electrons [5], the plasma electrons, in spite of these low velocities, can strongly affect the beam current measurements. A similar effect was observed in [6], where the current from the collector flew through the plasma created by a high-current beam near the collector. The measurements by a Rogowski coil show a low beam current after the discharge current pulse has come to an end (Fig. 2, curve 3). To evaluate the energy delivered by the beam to the collector in the gas-filled diode and to estimate the beam energy losses, we carried out calorimetric measurements of the beam energy. These measurements showed that the energy of the beam does not depend on the presence of a reverse-bias voltage at the calorimeter although the current in its circuit varies substantially. Thus, we can conclude that the calorimetric measurements more adequately describe the processes of electron extraction from the plasma cathode and transportation of the electron beam in a gas-filled diode.

The influence of a magnetic field on the efficiency of the transportation of an intense electron beam is illustrated in Fig. 3, which presents the electron beam energy E_b as a function of the distance L between the emissive electrode and calorimeter. It is seen that the axial magnetic field stabilizes the beam and that the electron motion in a magnetic field is nearly loss-free (curve 1). Curve 2 was obtained in the absence of a magnetic field. Figure 4 shows the distribution of the current density over the beam cross section measured

by Faraday cups placed behind the solenoid at a distance of 350 mm from the plasma cathode at a pressure of $P = 3 \times 10^{-2}$ Pa with (curve 1) and without (curve 2) imposing the magnetic field. In the presence of the axial magnetic field, the distribution shows a pronounced maximum at the beam axis with wings spreading beyond the initial beam diameter, which is determined by the diameter of the emissive opening.

In our experiments, the densities of the emission and beam currents were limited by the power of the pulse-forming line that feeds the discharge in the repetitive mode. To obtain high emission current density, we modified the configuration of the main discharge system and accelerating electrode. The diameters of the emissive opening in the grid plasma cathode and the accelerating electrode opening were reduced to 10 and 15 mm, respectively. In such a system, an emission current density of up to 100 A/cm^2 (Fig. 2b) was obtained at a discharge current of $\sim 140 \text{ A}$, an emission current of $\sim 80 \text{ A}$, and a collector current of $\sim 70 \text{ A}$. An extraction efficiency of $\alpha \approx 0.5\text{--}0.6$ corresponds to the transparency of the grid electrode with a $0.3 \times 0.3\text{-mm}$ mesh, which was used in this case.

An analysis of the beam cross-section profile measured by the burning of a system of thin Al foils showed that, at the entrance to the solenoid, the beam diameter was equal to the diameter of the accelerating electrode opening; then, the beam diameter decreased to 12 mm and remained almost unchanged until leaving the solenoid magnetic field at a distance of 30 cm from the emissive opening.

ANALYSIS OF THE RESULTS

(1) In contrast to the shape of the discharge current pulse, which remains unchanged as the diameter of the grid emissive electrode varies (Fig. 2), both the emission and collector currents depend on the grid electrode diameter. In particular, at a larger diameter, the emission current grows more slowly and the collector current is delayed for $\sim 2 \mu\text{s}$ with respect to the discharge current. Presumably, this is related to the different densities of the electron current extracted from the plasma. At higher current and electron beam densities, the formation of a plasma anode in the accelerating gap and the neutralization of the beam space charge in the drift channel proceed at a higher rate.

(2) The smearing of the beam current density profile at the collector may be related to various factors such as the thermal spreading of the electron beam and the scattering of electrons by the gas molecules during beam propagation along the gas-filled drift channel. At an energy of 10 keV and a gas pressure of $10^{-2}\text{--}10^{-1}$ Pa, the latter factor can be neglected because of the small differential scattering cross section $\sigma = 5 \times 10^{-18} \text{ cm}^2$.

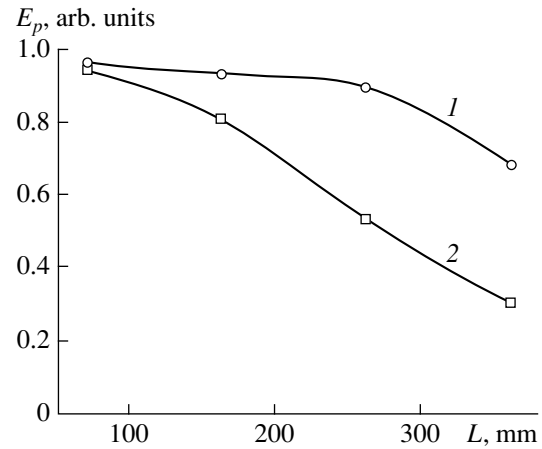


Fig. 3. Evolution of the beam energy along the drift channel.

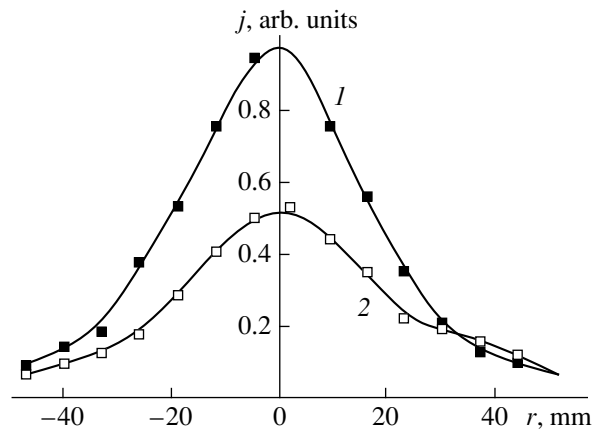


Fig. 4. Profile of the beam current density at the end of the solenoid.

The thermal spreading of the beam depends on the standard deviation [7]

$$\delta \approx t \left(\frac{kT_k}{m} \right)^{1/2}, \quad (2)$$

where $t = \int dz / (2eU/m)^{1/2}$ is the electron transit time through the drift space.

After the electron beam has passed through the accelerating electrode opening, the electron temperature increases to $T_a = T_k (r_a/r_k)^2$ [8], where T_k is the electron temperature at the cathode, r_a is the radius of the accelerating electrode opening, and r_k is the beam radius at the cathode. When the radii differ by 50%, the temperature increases by a factor of no more than 1.5 and, according to calculations, the spreading of the beam does not exceed 0.8 cm. Hence, neither electron thermal motion nor electron scattering due to elastic collisions can be responsible for the measured current density profile. The spreading of the beam can be explained by the fact that the electrons produced due to

gas ionization move toward the collector; as was mentioned above, their density may be one order of magnitude higher than the beam electron density. The electron motion toward the cathode is also responsible for the tail in the collector current after the discharge current pulse has come to an end. At a large grid electrode diameter, the peak of the collector current density at the beam axis is due to the focusing of the beam by the magnetic self-field in the region where the longitudinal magnetic field has not yet been formed.

(3) At a small grid electrode diameter, the moderate electron extraction efficiency, which is approximately equal to the grid transparency, is explained by the decrease or even vanishing of the negative anode potential drop near the grid electrode.

CONCLUSION

The use of a glow discharge in high-current electron sources increases their service life. Practical implementation of the discharge in the sources of charged particles is limited by the high pressure required for igniting and sustaining the discharge. This restriction can be lifted by using an auxiliary pulsed glow discharge (which can be ignited at higher pressures) to initiate the main discharge. In a plasma cathode based on this type of discharge, we obtained emission current densities of up to 100 A/cm^2 , which are comparable to the current densities in plasma emitters based on arc discharges.

The gas-filled diode enables the obtaining of electron beams with a higher perveance than in the vacuum diode. The very efficient transportation of intense electron beams generated in a gas-filled diode stems from

neutralizing the beam space charge by the ions produced due to gas ionization by the beam. Note that at currents of $\sim 100 \text{ A}$ and within the pressures range under investigation (10^{-2} – 10^{-1} Pa), applying an axial magnetic field stabilizes and improves the beam transportation.

REFERENCES

1. V. N. Devyatkov, N. N. Koval, and P. M. Schanin, *Zh. Tekh. Fiz.* **68** (1), 44 (1998) [*Tech. Phys.* **43**, 39 (1998)].
2. A. V. Zharinov, Yu. A. Kovalenko, I. S. Roganov, and P. M. Tyuryukanov, *Zh. Tekh. Fiz.* **56**, 687 (1986) [*Sov. Phys. Tech. Phys.* **31**, 413 (1986)].
3. A. V. Vizir', E. M. Oks, P. M. Schanin, and G. Yu. Yushkov, *Zh. Tekh. Fiz.* **67** (6), 27 (1997) [*Tech. Phys.* **42**, 611 (1997)].
4. V. I. Gushenets, N. N. Koval, P. M. Schanin, and V. S. Tolkachev, *IEEE Trans. Plasma Sci.* **27**, 1055 (1999).
5. Yu. R. Alanakyan and N. P. Shternov, *Fiz. Plazmy* **19**, 134 (1993) [*Plasma Phys. Rep.* **19**, 70 (1993)].
6. F. V. Lazarenko, E. S. Chebukov, and V. I. Éngel'ko, *Zh. Tekh. Fiz.* **59** (7), 159 (1989) [*Sov. Phys. Tech. Phys.* **34**, 806 (1989)].
7. S. I. Molokovskii and A. D. Sushkov, *High-Intensity Electron and Ion Beams* (Énergoatomizdat, Moscow, 1991).
8. P. T. Kirstein, G. S. Kino, and W. E. Waters, *Space-Charge Flow* (McGraw-Hill, New York, 1967; Mir, Moscow, 1970).

Translated by N. Ustinovskii

Determination of the Population of the $A^3\Sigma_u$ Electronic State of Nitrogen Molecules in a Glow-Discharge Plasma

B. T. Baïsova, S. L. Dolganeva, V. I. Strunin, N. N. Strunina, and I. A. Tikhomirov

Omsk State University, pr. Mira 55, Omsk, 644077 Russia

Received June 14, 2000; in final form, November 10, 2000

Abstract—Results on determining the population of the $A^3\Sigma_u$ electronic state of N_2 molecules in a gas discharge using a relatively simple spectroscopic technique are reported. The technique proposed can be used not only in pure N_2 , but also in nitrogen-containing gas mixtures. The populations of the metastable and nonequilibrium $B^3\Pi_g$ states of nitrogen molecules are determined. It is shown that the electron-impact excitation is a dominant mechanism for populating the $B^3\Pi_g$ state. © 2001 MAIK “Nauka/Interperiodica”.

In recent years, low-temperature plasma has found wide applications in chemical technology and material working. The characteristic feature of plasmachemical methods for synthesizing materials is that long-lived excited atomic and molecular states are used as reagents governing the efficiency of these processes. For example, when synthesizing aluminum nitride films at various substrates, the threshold reaction energy decreases if aluminum atoms or nitrogen molecules are in excited states. One way of decreasing the reaction threshold is to excite electronic levels of nitrogen molecules in a glow-discharge plasma. Molecular nitrogen has the $A^3\Sigma_u$ metastable state with an excitation energy of 6.7 eV and a lifetime of 1 s [1].

The aim of this paper is to determine the absolute population of the $A^3\Sigma_u$ state of molecular nitrogen and its main excitation mechanisms.

The glow discharge under study was excited between aluminum electrodes in a quartz tube 25 mm in diameter. The discharge-gap length was 32 cm. The discharge was powered by a dc power source providing a discharge current of up to 150 mA. The discharge was initiated in nitrogen at pressures from 0.3 to 1 torr.

Radiation passed through the side surface of the tube was focused with a condenser system into a narrow beam 6–8 mm long and was directed onto a slit of an STE-1 spectrograph with a 300-line/mm grating and glass prism, the slit width being 8 μm . The system measured the radiation spectrum within the wavelength range 4500–9000 Å. The dispersion in the wavelength region near 5500 Å was ~ 9.4 Å/mm. To obtain the absolute intensity of molecular bands, the system was calibrated with an Si-6-200 reference lamp. The population of the $B^3\Pi_g$ state of molecular nitrogen was determined from the intensity of the electronic-vibrational band sequence $\Delta v = 4$ of the first positive system

of nitrogen (the transition $B^3\Pi_g \rightarrow A^3\Sigma_u$) [2]. The intensity of the electronic-vibrational transition was found from the area $S_{v,v'}$ under the curve $S(x)$ describing the variation in the observed blackening of the given band along the spectrograph coordinate. The main processes resulting in the excitation of the $B^3\Pi_g$ state are the ion–electron recombination $N_2^+ + e + e \rightarrow N_2(B^3\Pi_g) + e$ and the electron-impact excitation $N_2(X) + e \rightarrow N_2(B^3\Pi_g) + e$.

In order to reveal the mechanism for populating the $B^3\Pi_g$ state, we measured the dependences of the band intensities of the first positive system on the discharge current 50–150 mA at various nitrogen pressures. Figure 1 shows the dependence of the intensity of the 5959.0-Å band of the first positive system of nitrogen for the same value of the gas pressure (0.5 torr), but for different values of the discharge current (50–150 mA).

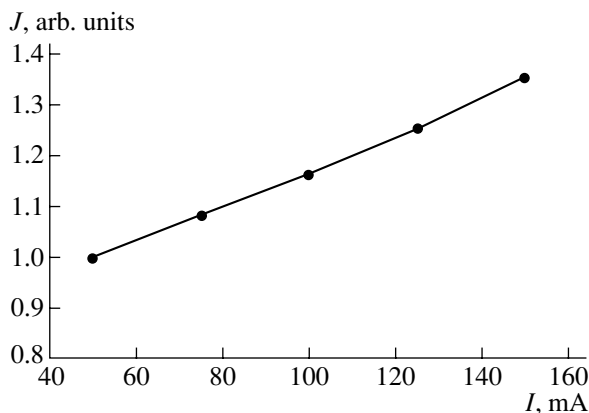


Fig. 1. Intensity of the 5959.0-Å band of the first positive system of nitrogen as a function of the discharge current at a pressure of 0.5 torr.

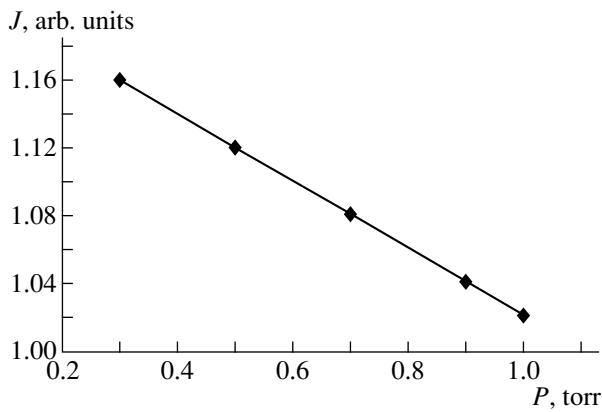


Fig. 2. Intensity of the 5959.0-Å band of the first positive system of nitrogen as a function of the nitrogen pressure for a discharge current of 100 mA.

It is seen that, as the discharge current increases, the intensity of the band grows, which is explained by the increase in the electron density [3]. The increase in the intensity may be due to the increase in both the rate of recombination processes and the electron–molecule collision frequency, which results in the excitation of the $B^3\Pi_g$ state of molecular nitrogen.

Figure 2 shows the dependence of the intensity of the 5959.0-Å band of the first positive system of nitrogen on the pressure (0.3–1 torr) for the same value of the discharge current (100 mA). The other bands of the system behave in a similar manner. As the pressure increases, the intensity of the band decreases. The decrease in the intensity of the bands with increasing gas pressure is due to the fact that T_e drops because the electron mean free path decreases and the number of events of stepwise ionization and vibrational excitation of molecules increases, which leads to a decrease in the mean electron energy [3]. As a result, the excitation efficiency of the $B^3\Pi_g$ state decreases, whereas the rates

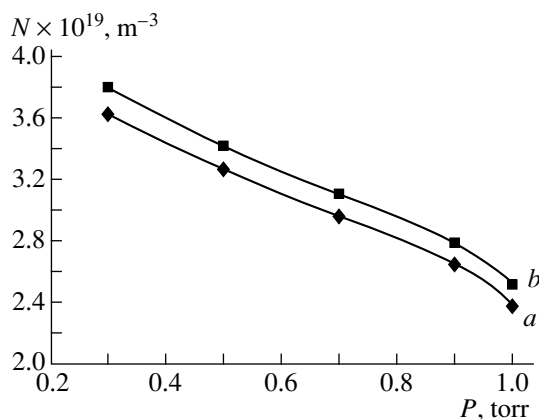


Fig. 3. Populations of (a) $B^3\Pi_g$ and (b) $A^3\Sigma_u$ states of nitrogen as functions of pressure for a discharge current of 100 mA.

of ion-recombination processes remains unchanged (provided that the densities of charged particles in the plasma remain constant). It follows from the above considerations that the direct electron-impact excitation of the $B^3\Pi_g$ state is dominant.

The population of the vibrational levels was calculated from the intensities of radiation bands by the formula [4]

$$N_{v'} = \frac{\frac{d\lambda}{dx} K(\lambda, T_k) S_{v'v''}}{lh\nu_{v'v''} A_{v'v''}}, \quad (1)$$

where $\frac{d\lambda}{dx}$ is the inverse dispersion determined from the spectrogram, l is the length of the viewed part of the positive discharge column, $K(\lambda, T_k)$ is the calibration factor, and $A_{v'v''}$ is the transition probability.

Summing the determined values of the population of individual vibrational levels, we obtain the population of the $B^3\Pi_g$ electronic state of molecular nitrogen.

To determine the absolute population of the $B^3\Pi_g$ state, it is necessary to take into account the radial distribution of molecules in the discharge tube. Since the electron impact is the main mechanism for the excitation of the $B^3\Pi_g$ state, the radial distribution of molecules exactly coincides with the radial distribution of electrons. The radial distribution of electrons in the cylindrical tube is given by the expression [5]

$$N(r) = C J_0\left(2.405 \frac{r}{r_0}\right), \quad (2)$$

where r_0 is the discharge-tube radius, C the constant determined from the boundary conditions, and J_0 is the zero-order Bessel function of the first kind.

From the integral population, taking into account the Bessel radial distribution of the electron density (2),

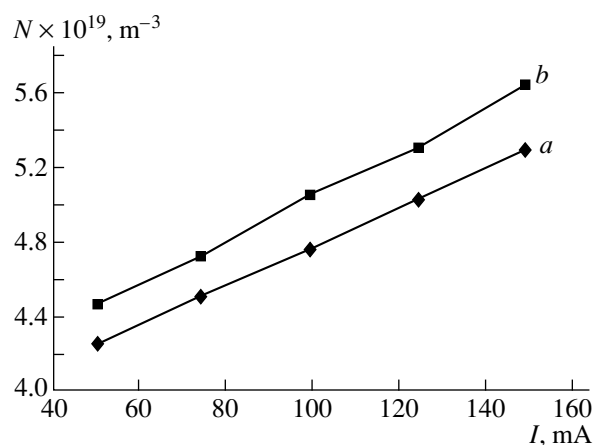


Fig. 4. Populations of (a) $B^3\Pi_g$ and (b) $A^3\Sigma_u$ states of nitrogen as functions of the discharge current at a pressure of 0.5 torr.

The main processes of populating and depopulating the $A^3\Sigma_u$ state in a discharge [6]

Process no.	Process	Rate coefficient, $\text{cm}^3 \text{s}^{-1}$
1	$\text{N}_2(X, \nu') + e \longrightarrow \text{N}_2(A, \nu) + e$	$0.1\text{--}2.3 \times 10^{-11}$
2	$\text{N}_2(B, \nu') \longrightarrow \text{N}_2(A, \nu) + h\nu(1^+)$	$1\text{--}2 \times 10^5 \text{ s}^{-1}$
3	$\text{N}_2(B, \nu') + \text{N}_2(X, \nu'') \longrightarrow \text{N}_2(A, \nu) + \text{N}_2(X, \nu')$	$1.3\text{--}7 \times 10^{-11}$
4	$\text{N}_2(X, \nu') + \text{N}_2(X, \nu') \longrightarrow \text{N}_2(A, \nu) + \text{N}_2(X, \nu' - \Delta\nu)$	$< 10^{-16}$
5	$\text{N}_2(A, \nu) + \text{N}_2(X, \nu') \longrightarrow \text{N}_2(B, \nu') + \text{N}_2(X, \nu' - \Delta\nu)$	$\sim 10^{-10}$
6	$\text{N}_2(A, \nu) \longrightarrow \text{N}_2(X, \dots)$	$\left(\frac{815}{p(T_0)}\right)\left(\frac{T_r}{300}\right)^{3/2} \text{ s}^{-1}$

we calculated the population of the $B^3\Pi_g$ state (at the discharge axis) under different glow-discharge conditions (Figs. 3, 4, curves *a*). The obtained values of the density of the $B^3\Pi_g$ state can be used to determine the population of the $A^3\Sigma_u$ state.

The main processes of populating and depopulating this state in a glow discharge are listed in the table. The contribution from process 4 to the population of the $A^3\Sigma_u$ state is small because of the low rate coefficient of this process and the low density of particles in the ground state ($\sim 10^{18} \text{ m}^{-3}$).

The main contribution to the population of the $A^3\Sigma_u$ state in the discharge comes from the direct electron-impact excitation of molecules (1), the deactivation of the $B^3\Pi_g$ state through radiative transitions (2), and quenching this state by nitrogen molecules (3). The processes of the excitation of the $B^3\Pi_g$ state with the participation of metastable molecules (5) and its quenching by nitrogen molecules are not reversible in detail. However, they almost balance each other. The deactivation of the $A^3\Sigma_u$ state occurs at the wall (6); the rate of this process depends on the density of the diffusion flow of metastables.

The metastable density was evaluated from the balance equation

$$D_m \Delta N_m = AN_B + k_{1,1} N_0 n_e, \quad (3)$$

where D_m is the diffusion coefficient of metastables ($\sim 10^{13} \text{ cm}^2 \text{ s}^{-1}$) [7], N_m is the population of the $A^3\Sigma_u$ metastable state, N_B is the population of the $B^3\Pi_g$ state, $k_{1,1}$ is the rate coefficient of the electron-impact excita-

tion of molecules, N_0 is the population of the $X^1\Sigma_g$ state, A is the probability of the $B^3\Pi_g \longrightarrow A^3\Sigma_u$ radiative transition, and n_e is the electron density.

Curves *b* in Figs. 3, 4 show the calculated values of the population of the $A^3\Sigma_u$ state at the discharge axis under different discharge conditions. It is seen that the population of the $A^3\Sigma_u$ state increases as the current increases or the pressure decreases. This is evidence that the $A^3\Sigma_u$ state is populated through the direct electron-impact excitation of molecules from the ground state and deexcitation of the $B^3\Pi_g$ state.

REFERENCES

1. B. M. Smirnov, *Excited Atoms* (Énergoizdat, Moscow, 1982).
2. R. W. B. Pearse and A. G. Gaydon, *The Identification of Molecular Spectra* (Chapman and Hall, London, 1941; Inostrannaya Literatura, Moscow, 1949).
3. V. L. Granovskii, *Electric Current in a Gas: Steady-state Current* (Nauka, Moscow, 1971).
4. L. S. Polak, D. I. Slovetskii, and A. S. Sokolov, *Opt. Spektrosk.* **32**, 472 (1979).
5. E. W. McDaniel and E. A. Mason, *The Mobility and Diffusion of Ions in Gases* (Wiley, New York, 1973; Mir, Moscow, 1976).
6. D. I. Slovetskii, *Mechanisms for Chemical Reactions in Nonequilibrium Plasmas* (Nauka, Moscow, 1980).
7. *Handbook of Physical Quantities*, Ed. by I. S. Grigor'ev and E. Z. Meĭlikhov (Énergoatomizdat, Moscow, 1991).

Translated by N. Larionova

Surface Waves at the Boundaries of Relativistic Plasma Streams

I. L. Sheinman

St. Petersburg State Electrotechnical University, St. Petersburg, 197376 Russia

Received June 23, 1999

Abstract—A study is made of stable and unstable electromagnetic surface waves at the boundaries of the plane and cylindrical relativistic plasma streams in the frequency range corresponding to positive values of the plasma permittivity. It is demonstrated that there are critical parameters for the transition from slow to fast waves, namely, the angle between the velocity and the wave vector in plane geometry and the smallest mode number in cylindrical geometry. It is shown that the critical parameter for the onset of the firehose instability of an electron stream is the transverse size of the stream. Higher firehose modes of the stream are shown to be suppressed by applying a strong longitudinal magnetic field. © 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

A number of phenomena revealed when studying the emission of electromagnetic waves and their propagation in moving beam–plasma media are important not only for the development of fundamental problems that link plasma physics with electrodynamics of moving media but also for some practical applications, such as plasma microwave electronics and experimental physics of linear accelerators. An analysis of the conditions for the generation of electromagnetic waves is also of interest for studying some phenomena in moving plasmas in the Earth’s ionosphere and in space.

Real plasma streams and charged particle beams are bounded. As a result, the dispersion properties of surface waves that arise at their boundaries may differ greatly from those of waves in unbounded media. Surface waves belong to a large class of waves that propagate along a certain surface and whose field decreases sharply either in one direction away from this surface (one-sided surface waves) or in both directions (true surface waves). Thus, a one-sided surface wave can be excited when an electromagnetic wave is totally reflected from the interface between two dielectric media, provided that it is incident from the side of a medium with a larger refractive index. The existence of true surface waves is associated with the nonzero imaginary part of the surface impedance [1], in the case of which the refractive index of one of the dielectric media should be complex [2]. In particular, in the absence of losses, this situation takes place in a plasma with a negative permittivity ($\epsilon < 0$). Surface waves cannot exist at the interface between two immobile media with positive permittivities. However, Barsukov and Kanarëkin [3] showed that the relativistic motion of one of the media can give rise to an electromagnetic surface wave even at the tangential discontinuity in the velocity of a homogeneous medium with $\epsilon > 0$. Analogously, a surface wave can be excited at the tangential discontinuity in the velocity of a compressed gas [4, 5]. In both cases,

a necessary condition for the existence of surface waves is that the medium be anisotropic because of the appearance of a preferential direction of motion. This conclusion is confirmed by the existence of surface waves at the interface between an isotropic medium and a crystal and at a crystal–crystal interface [6, 7].

The instability of beams and bunches in such systems has not yet been analyzed. On the one hand, the electromagnetic instability of a beam can be used to convert the beam kinetic energy into field energy in a medium, which is the desired effect. On the other hand, the parasitic electromagnetic instability of a beam is a serious obstacle to achieving high operation efficiency of wakefield accelerators [8] and devices in relativistic plasma microwave electronics [9] because it may lead to the decay of the beam. The goal of this paper is to study this problem. In particular, we consider one of the physical mechanisms for hydrodynamic instabilities of relativistic electron streams, specifically, the mechanism associated with the onset and growth of electromagnetic surface waves at the beam–plasma interface. The most interesting case (which has been studied very little) is that with $\epsilon > 0$, i.e., with the positive permittivities of both of the contacting media (the immobile plasma and the stream). In this case, the instability can be associated with the generation of both the longitudinal space-charge waves of the stream [10] and electromagnetic surface waves at the interface between moving and immobile plasmas. We show that, in the latter case, the instability occurs when the wave vector of the surface wave is noncollinear with the velocity vector in a medium. In cylindrical geometry, this instability is associated with the excitation of asymmetric surface waves. When the electron beam is sufficiently curved, the first mode of this instability is the so-called firehose instability of an electron beam, which was studied in [11–13].

BASIC EQUATIONS FOR A RELATIVISTIC ELECTRON BEAM IN A PLASMA

The waves excited in a relativistic beam–plasma system can be studied using a phenomenological approach in which a relativistic electron beam is treated as a moving continuous medium. This approach is equivalent to a hydrodynamic description of the beam–plasma system, which, unlike a more exact kinetic description, does not take into account some important features of the excited waves (in particular, their damping due to the thermal motion of the charged particles) [14]. However, in the frequency range in which the phase velocity of the excited waves is comparable with the velocity of the relativistic beam and is substantially higher than the thermal velocity of the charged particles, the main properties of the wave processes in a beam–plasma system can be described using the hydrodynamic approach.

In the phenomenological approach, an electron beam is treated as a plasma stream (with permittivity ϵ), which moves with the velocity $\mathbf{V} = \beta c$ (where c is the speed of light) with respect to an immobile background plasma. Such a system is described by Maxwell's equations

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}, \quad (1)$$

$$\nabla \times \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t}, \quad (2)$$

$$\nabla \mathbf{B} = 0, \quad (3)$$

$$\nabla \mathbf{D} = 0 \quad (4)$$

supplemented with the Minkowski constitutive relations for a moving medium [15]:

$$\begin{aligned} \mathbf{B} + \mathbf{E} \times \boldsymbol{\beta} &= \mathbf{H} + \mathbf{D} \times \boldsymbol{\beta}, \\ \mathbf{D} + \boldsymbol{\beta} \times \mathbf{H} &= \hat{\epsilon}(\mathbf{E} + \boldsymbol{\beta} \times \mathbf{B}), \end{aligned} \quad (5)$$

where \mathbf{E} and \mathbf{H} are the electric and magnetic field strengths and \mathbf{D} and \mathbf{B} are the electric and magnetic inductions, respectively.

Equations (5) apply to an arbitrary anisotropic medium with a complex permittivity tensor $\hat{\epsilon}$. In particular, the permittivity tensor of a gyrotropic medium has the form [15]

$$\hat{\epsilon} = \begin{vmatrix} \epsilon_{\perp} & ig & 0 \\ -ig & \epsilon_{\perp} & 0 \\ 0 & 0 & \epsilon_{\parallel} \end{vmatrix}, \quad (6)$$

where ϵ_{\perp} and ϵ_{\parallel} are the transverse and longitudinal permittivities, respectively, and g is the gyration coefficient.

On the other hand, we can switch from the constitutive relations (5) to the equations

$$\mathbf{B} = \mathbf{H}, \quad D_i = \hat{\epsilon}_{ij} E_j.$$

However, in this way, we are faced with the problem of calculating the permittivity tensor $\hat{\epsilon}$ using a particular model of a medium (in [16], the corresponding calculations were carried out for a relativistic electron beam in an immobile magnetized plasma).

In order to apply the phenomenological approach, it is necessary to determine the permittivity of the moving medium using the equations for particle motion and also to impose the conditions on the equilibrium states of the beam and plasma. In view of this, we use the phenomenological approach to study surface waves at the boundary of a neutralized relativistic flow (plasma stream) in an immobile plasma.

SURFACE WAVES AT THE BOUNDARY OF A MOVING PLANE PLASMA SLAB

It is well known that, in order for a true electromagnetic surface wave to exist at the planar interface between two media, the permittivities of the media should be opposite in sign: $\epsilon_1 > 0$ and $\epsilon_2 < 0$ [15]. The relativistic motion of one of the media (e.g., the motion of a high-current beam in a plasma) can give rise to qualitatively new effects. Thus, Pikulin and Stepanov [17] showed that, at the tangential discontinuity in the velocity of a nondispersive homogeneous medium, there are no surface waves whose transverse wave vector \mathbf{k}_{\perp} is collinear with the velocity \mathbf{V} of the motion of the medium. On the other hand, if the vectors \mathbf{k}_{\perp} and \mathbf{V} are noncollinear and the angle between them is φ , then there exists a critical angle φ_k such that, for $\varphi > \varphi_k$, surface waves can be excited at the tangential discontinuity in the velocity of a homogeneous medium whose permittivity on both sides of the discontinuity is positive [10, 18].

Along with the surface waves that are stable at the tangential discontinuity in the velocity, there exist growing electromagnetic surface waves that trigger a hydrodynamic instability of the plasma streams. The relativistic motion of a plane plasma slab gives rise to the Cherenkov instability (via the excitation of waves propagating away from the boundary in the external dielectric medium [10]) or to the instability of surface waves, provided that the plasma is immobile [18]. However, the analysis carried out in [10, 18] applies only to surface waves whose wave vector is collinear with the velocity of a medium.

Let a plane plasma slab with thickness $2h$ and permittivity ϵ_2 move with the velocity $\mathbf{V} = \beta c$ in an immobile medium with permittivity ϵ_1 . In Cartesian coordinates, the velocity has the form $\boldsymbol{\beta} = (0, \beta_y, \beta_z)$, and the transverse wave vector \mathbf{k}_{\perp} is directed along the z -axis. From Maxwell's equations (1)–(4), constitutive relations (5), and the boundary conditions, we obtain the following dispersion relation for surface waves in a planar waveguide filled with a moving plasma:

$$\begin{aligned} (\epsilon_2 \eta_1 S + \epsilon_1 \eta_2)(\eta_1 S^{-1} + \eta_2) \\ = (\epsilon_2 - 1)(\epsilon_1 - 1) \eta^2 \gamma^2 \beta_y^2, \end{aligned} \quad (7)$$

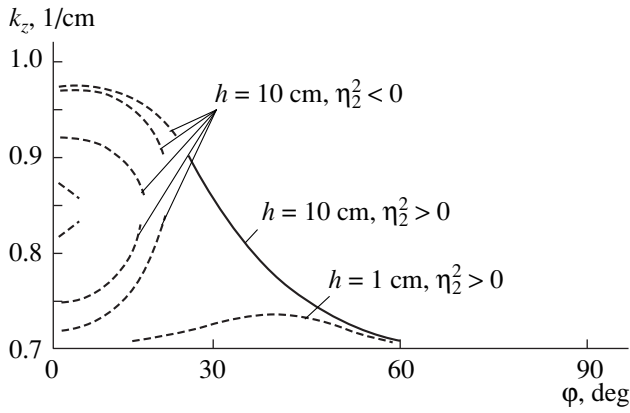


Fig. 1. Propagation constant k_z vs. the angle φ between the vectors \mathbf{k}_\perp and \mathbf{V} for the surface waves of a plane moving slab.

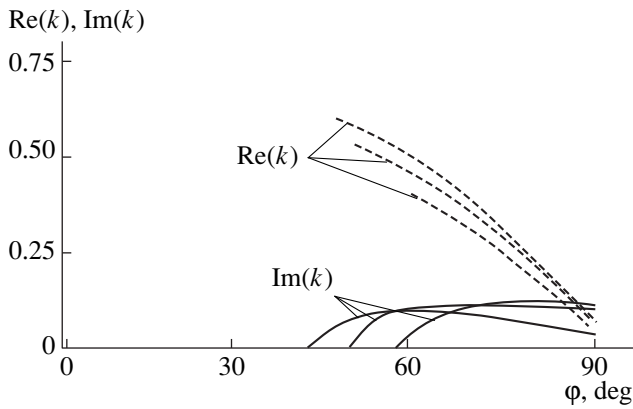


Fig. 2. Real and imaginary parts of $k = \omega/c$ vs. the angle φ between the vectors \mathbf{k}_\perp and \mathbf{V} for the surface waves at the tangential discontinuity in the velocity (light curves) and for the surface waves of a moving plane slab (heavy curves).

where $\eta_1^2 = \eta^2 - \epsilon_1$, $\eta_2^2 = \eta^2 - 1 - (\epsilon_2 - 1)\gamma^2(1 - \eta\beta_z)^2$, $\eta = k_z/k$, $k = \omega/c$, $k_z = |\mathbf{k}_\perp|$, and $\gamma = (1 - \beta^2)^{-1/2}$.

Equation (7) describes waves whose field component E_z is symmetric [$S = \coth(kh\eta_2)$] about the yz plane and waves with an antisymmetric [$S = \tanh(kh\eta_2)$] field component. For a tangential discontinuity in the velocity of a homogeneous medium (when $\epsilon_1 = \epsilon_2$), symmetric and antisymmetric waves are both described by Eq. (1). For $h \rightarrow +\infty$, the mutual influence of the boundaries is negligibly weak, so that Eq. (1) passes over to the following dispersion relation for surface waves at a plane boundary between two media moving with respect to one another [3]:

$$(\epsilon_2\eta_1 + \epsilon_1\eta_2)(\eta_1 + \eta_2) = (\epsilon_2 - 1)(\epsilon_1 - 1)\eta^2\gamma^2\beta_z^2. \quad (8)$$

Figure 1 illustrates the dependence of the propagation constant k_z on the angle φ between the vectors \mathbf{k}_\perp and \mathbf{V} for the surface waves of a moving plane plasma slab. The results were obtained for the parameter values $k = 1 \text{ cm}^{-1}$, $\beta = 0.941$, $\gamma = 2.96$, and $\epsilon_1 = \epsilon_2 = 0.5$. Outside the slab, the surface wave amplitude decreases

exponentially away from the boundaries. Inside the slab, surface waves can be divided into two groups: slow surface waves (for $\eta_2^2 > 0$) and fast internal waves (for $\eta_2^2 < 0$). In contrast to the case of an isolated tangential discontinuity in the velocity (when the surface waves cannot propagate at angles smaller than a certain critical angle φ_k between the vectors \mathbf{k}_\perp and \mathbf{V} [3]), the critical angle φ_k for the surface waves of a plane plasma slab corresponds to the transition from fast to slow waves. The critical angle is the largest for $S \approx \sqrt{\epsilon_1/\epsilon_2}$; at $\epsilon_1 = \epsilon_2$, this corresponds to $h \rightarrow +\infty$. The thinner the slab, the smaller the critical angle. For sufficiently small values of h , fast internal waves do not exist.

Along with stable solutions, Eq. (1) has solutions that grow in time or space and describe the onset of absolute or convective instabilities of a plane plasma slab, respectively.

Figure 2 shows the real and imaginary parts of k calculated as functions of the angle φ between the vectors \mathbf{k}_\perp and \mathbf{V} for surface waves at the tangential discontinuity in the relative velocity between two media and for the surface waves of a moving plasma slab. The profiles in Fig. 2 were obtained by solving Eqs. (7) and (8) numerically at a fixed real propagation constant k_z for the parameter values $k_z = 1 \text{ cm}^{-1}$, $n_1 = n_2 = 1.4 \times 10^{11} \text{ cm}^{-3}$, $\beta = 0.999$, $\gamma = 22.4$, and $h = 1 \text{ cm}$. In the ultrarelativistic limit, the tangential discontinuity in the velocity is stable against surface waves whose wave vector is parallel to the velocity of a medium, but it is unstable against surface waves propagating at angles larger than a certain critical angle φ . For a moving plane plasma slab (stream), the critical angles above which symmetric and antisymmetric (about the yz plane) surface waves start growing are different; the angular range in which the stream is stable is governed by the smaller of the critical angles.

SURFACE WAVES OF A CYLINDRICAL PLASMA STREAM

As has been noted previously, at a planar interface between two media, surface waves can exist only when the permittivities of the media are opposite in sign [15]. In cylindrical geometry, this condition on the permittivities is retained; however, for media moving at relativistic velocities, there exist new interesting effects that are associated with the multimode structure of the excited waves [19, 20] and have not yet been studied.

We consider a cylindrical plasma stream with radius R and permittivity ϵ_2 , moving in an immobile medium with permittivity ϵ_1 . Let the plasma in the stream move along its symmetry axis (the z -axis) at a constant velocity $\mathbf{V} = \beta\mathbf{c}$, where c is the speed of light.

From Maxwell's equations (1)–(4), constitutive relations (5), and the boundary conditions, we obtain

the dispersion relation

$$(\varepsilon_2 S_2 - \varepsilon_1 S_1)(S_2 - S_1) = \frac{v^2}{k^2 R^2} \left(\frac{\tilde{\eta}}{\eta_2^2} - \frac{\eta}{\eta_1^2} \right)^2, \quad (9)$$

where

$$S_1 = \frac{K'_v(kR\eta_1)}{\eta_1 K_v(kR\eta_1)}, \quad S_2 = \frac{I'_v(kR\eta_2)}{\eta_2 I_v(kR\eta_2)},$$

$I_v(x)$ and $K_v(x)$ are modified Bessel functions, $\eta = k_z/k$, $\tilde{\eta} = \gamma^2((1 - \varepsilon_2\beta^2)\eta + (\varepsilon_2 - 1)\beta)$, $\eta_1^2 = \eta^2 - \varepsilon_1$, $\eta_2^2 = \eta^2 - 1 - (\varepsilon_2 - 1)\gamma^2(1 - \eta\beta)^2$, and $k = \omega/c$.

Dispersion relation (9) describes surface waves in a waveguide formed by a relativistic beam propagating in a plasma. In the limit in which the waveguide radius is much larger than the characteristic transverse dimension of the surface wave field, Eq. (9) passes over to dispersion relation (8) for a plane tangential discontinuity in the velocity.

Figure 3 illustrates the propagation constant k_z as a function of the dimensionless waveguide radius kR for $\varepsilon_1 = \varepsilon_2 = 0.5$, $k = 3 \text{ cm}^{-1}$, and $\beta = 0.999$ ($\gamma = 22.4$). The heavy curves refer to the slow waves of the waveguide: the radial profiles of the fields E and H of these waves are described by modified Bessel functions both inside and outside the waveguide ($\eta_1^2 > 0$, $\eta_2^2 > 0$). The light curves refer to the fast waves of the waveguide: the radial profiles of the fields E and H of these waves are described by conventional Bessel functions inside the waveguide and by modified Bessel functions on the outside ($\eta_1^2 > 0$, $\eta_2^2 < 0$).

For $v = 0$, the dispersion relation has no roots, which, in planar geometry, corresponds to surface waves propagating in a direction parallel to the direction of motion of the medium. For a given radius of the stream, the slow waves with mode numbers larger than a certain critical mode number can only be excited; i.e., the dispersion relation has real roots when $v > v_{cr}$.

In contrast to slow waves that can be excited at the interface between two media and whose amplitude decreases exponentially with distance from the interface on both sides of it, fast waveguide waves can exist only when the waveguide is bounded in space. In plane geometry, fast waves are excited in a waveguide with two boundaries: inside the waveguide, the wave amplitude is described by the harmonic functions and, on the outside, it decreases exponentially away from the waveguide boundaries. In a cylindrical waveguide, fast waves can be excited at $v < v_{cr}$.

If the propagation constant k_z is treated as a root of dispersion relation (9) at a fixed real $k = \omega/c$, then we obtain waves whose amplitude increases with z when the imaginary part of the solution is negative. In contrast, if we treat k_z as a real parameter and k as a root of dispersion relation (9), then the positive imaginary part

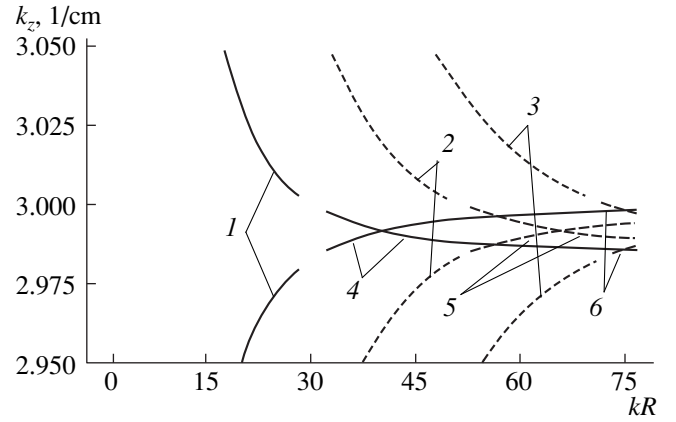


Fig. 3. Propagation constant k_z vs. the radius R of the plasma stream for $v = (1, 4)$ 1, $(2, 5)$ 2, and $(3, 6)$ 3. Curves 1–3 correspond to slow waves ($\eta_2^2 > 0$), and curves 4–6 refer to fast waves ($\eta_2^2 < 0$).

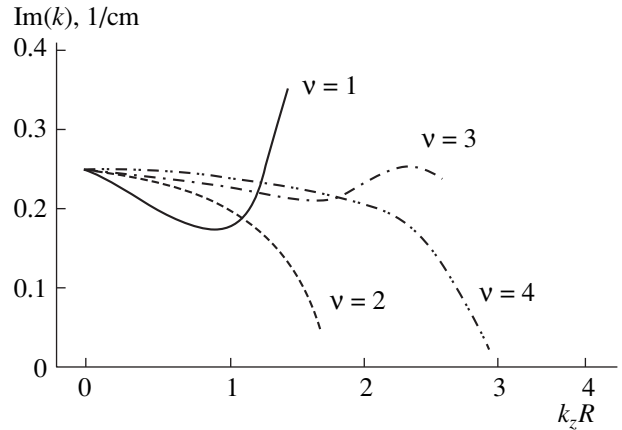


Fig. 4. Growth rate $\text{Im}(k)$ of the nonresonant instability vs. the dimensionless waveguide radius $k_z R$.

of the frequency $\omega = kc$ is the temporal growth rate $\delta = \text{Im}(\omega)$.

Let us analyze the instability related to the onset of the first mode of the surface wave in more detail. Substituting the relationship $\omega = \omega_0 + i\delta$ into the expression for the perturbed electron density \tilde{n} and representing the resulting formula in the real form, we obtain

$$\tilde{n}(r, \theta, z, t) = \tilde{n}(r) \exp(\delta t) \cos(\Theta + k_z z - \omega_0 t).$$

We can see that, for $k_z z - \omega_0 t = \text{const}$, a growing helical wave causes the beam particles to be displaced in the same direction, thereby forcing the beam to bend. For $\tilde{n} \sim n$, the model of a rigid cylindrical stream fails to hold and the instability associated with the excitation of the first mode of asymmetric surface waves is the firehose instability of an electron beam, which was studied in [11–13].

Figure 4 shows the growth rate (divided by the speed of light) of the nonresonant instability associated

with the excitation of slow surface waves as a function of the dimensionless waveguide radius $k_z R$ for $k = 1 \text{ cm}^{-1}$, $\beta = 0.999$ ($\gamma = 22.4$), $n_1 = 1.41 \times 10^{11} \text{ cm}^{-3}$, $n_2 = 3.12 \times 10^{12} \text{ cm}^{-3}$ (in which case we have $\epsilon_1 = 1 - k_{p1}^2/k^2 = 0.5$, $\epsilon_{2\perp} = 1 - k_{p2}^2/\gamma k^2 = 0.5$, and $k_{p1,2}^2 = 4\pi e^2 n_{1,2}/mc^2$).

Note that, for $v = 0$, dispersion relation (9) has no unstable solutions. In plane geometry, this corresponds to surface waves propagating in the direction parallel to the velocity of the medium. For $v \neq 0$ and the fixed stream radius, the surface waves with the mode numbers $v > v_{cr}$ can only be excited. For a sufficiently large stream radius, the moving-plasma-immobile-plasma system is stable against the excitation of the lowest modes. Therefore, we can eliminate, e.g., the $v = 1$ mode, which grows at the fastest rate and can give rise to a firehose instability of the stream. In other words, for the unstable solutions (in contrast to the stable ones), the critical parameter is the radius of the cylindrical stream below which the first mode of the surface waves can be excited.

In this section, we have shown that, in a system formed by a relativistic plasma stream and an immobile plasma, both stable and growing non-axisymmetric electromagnetic surface waves can exist in the frequency range in which the permittivity is positive on both sides of the tangential discontinuity in the velocity. We have found that, for the given parameters of the stream and the plasma, the critical parameter for the existence of a solution to the dispersion relation is the lowest mode number below which the fast waves with a spatially oscillating field can only exist inside the waveguide. In plane geometry, an analogous critical parameter is the critical angle between the vectors β and \mathbf{k}_\perp . The growth of the first mode of unstable surface waves can be interpreted as the initial phase of the firehose instability of an electron stream. For given values of β , ϵ , and ω , the critical parameters that govern the onset of instability in a moving-plasma-immobile-plasma system are the number of the lowest allowed mode and the transverse size of the stream.

SURFACE WAVES OF A CYLINDRICAL RELATIVISTIC MAGNETIZED PLASMA STREAM

Let us consider a nonrotating cylindrical relativistic plasma stream of radius R , moving with velocity \mathbf{V} in an immobile plasma. The spatial spreading of the stream can be prevented by applying a sufficiently strong longitudinal magnetic field \mathbf{B}_0 [21], so that both the moving stream and immobile plasma are gyrotropic. When the magnetic field energy is much higher than the energy of the stream, $B_0^2/8\pi \gg n_b W$ (where W is the electron energy in the beam), the stability problem is treated in the electrostatic approximation, without allowance for magnetic perturbations in the system

[21–23]. However, the densities and energies of the particle beams in modern-day experimental devices are so high that the conditions for the validity of the electrostatic approximation may be difficult to satisfy. On the other hand, in addition to potential electromagnetic waves, whose phase velocities are much lower than the speed of light and which can be described in the electrostatic approximation, there can also be nonpotential surface waves, whose phase velocities are comparable with the speed of light and which, under certain conditions, may cause the stream to become unstable [24]. It is this type of waves that are used for wakefield acceleration of charged particles. In this connection, it is of interest to solve the problem of the surface waves of a relativistic plasma stream by exactly analyzing Maxwell's equations for a moving gyrotropic medium.

For an immobile gyrotropic medium, nonpotential surface waves at the boundary of the plasma column that partially fills a cylindrical metal waveguide was considered by Kondratenko [25]. The emission of electromagnetic radiation during the excitation of surface waves by an electron beam in a bounded magnetized plasma was studied in [24, 26, 27]. It is worth noting that, in all of the cited papers (except for [21, 28]), an analysis was made of the axisymmetric surface waves of a cylindrical plasma stream. However, in the parameter range in which no axisymmetric surface waves can exist, the higher unstable modes of the plasma-filled waveguide can occur. Non-axisymmetric modes in metal waveguides filled completely with a plasma were investigated in [21, 24, 28].

We substitute the Minkowski constitutive relations (5) with permittivity tensor (6) into Maxwell's equations (1)–(4) in order to obtain the dispersion relation

$$Q_{E1} Q_{H2} = Q_{E2} Q_{H1}, \quad (10)$$

where

$$Q_{Hj} = \frac{v}{kR} (\tau - \varphi_{2j} \zeta_*) + S_{2j} \varphi_{2j} - \frac{\varphi_{2j} (S_{12} \varphi_{12} - S_{11} \varphi_{11}) - \varphi_{11} \varphi_{12} (S_{12} - S_{11})}{\varphi_{12} - \varphi_{11}},$$

$$Q_{Ej} = \frac{v}{kR} (\varphi_{2j} \tau - \zeta^*) + \epsilon_{\parallel 2} S_{2j} - \frac{\epsilon_{\parallel 1} (\varphi_{2j} (S_{12} - S_{11}) - (S_{12} \varphi_{11} - S_{11} \varphi_{12}))}{\varphi_{12} - \varphi_{11}},$$

$$\zeta_* = \left(\frac{g_2 \eta_{*2}^2}{w_2} - \frac{g_1 \eta_{*1}^2}{w_1} \right), \quad \zeta^* = \left(\frac{g_2 \eta_2^{*2}}{w_2} - \frac{g_1 \eta_1^{*2}}{w_1} \right),$$

$$\tau = \left(\frac{u_2}{w_2} - \frac{u_1}{w_1} \right), \quad \varphi_{ij} = \frac{\eta_{ij}^2 - \eta_{Ei}^2}{\sigma_i^2},$$

$$\eta_{i[1,2]}^2 = \frac{1}{2}(\eta_{Hi}^2 + \eta_{Ei}^2 \pm \sqrt{(\eta_{Hi}^2 + \eta_{Ei}^2)^2 + 4\varepsilon_{\parallel i}\sigma_i^4}),$$

$$\eta_{Hi}^2 = \frac{\gamma_i^2}{\varepsilon_{\perp 2}}(\eta_i^{*2}\varepsilon_{\perp i} + \eta_{*i}^2(\varepsilon_{\perp i}^2 - g_i^2)),$$

$$\eta_{Ei}^2 = \frac{\varepsilon_{\parallel i}\gamma_i^2}{\varepsilon_{\perp i}}(\eta_i^{*2} - \eta_{*i}^2\varepsilon_{\perp i}), \quad \sigma_i^2 = \frac{\gamma_i^2}{\varepsilon_{\perp i}}g_i\eta_{*i}\eta_i^*,$$

$$S_{1j} = \frac{K'_v(kR\eta_{1j})}{\eta_{1j}K_v(kR\eta_{1j})}, \quad S_{2j} = \frac{I'_v(kR\eta_{2j})}{\eta_{2j}I_v(kR\eta_{2j})},$$

$$u_i = \gamma_i^2((\eta_i^{*2} - \varepsilon_{\perp i}\eta_{*i}^2)(\eta_i^* + \varepsilon_{\perp i}\beta_i\eta_{*i}) + \eta_{*i}^3\beta_i g_i^2),$$

$$w_i = \gamma_i^2((\eta_i^{*2} - \varepsilon_{\perp i}\eta_{*i}^2)^2 - \eta_{*i}^4 g_i^2),$$

$$\eta_{*i} = 1 - \eta\beta_i, \quad \eta_i^* = \eta - \beta_i, \quad \gamma_i = (1 - \beta_i^2)^{-1/2},$$

$$\eta = k_z/k, \quad k = \omega/c, \quad \beta_1 = 0, \quad \beta_2 = V/c.$$

For a sufficiently strong longitudinal magnetic field such that $k_c \gg k\eta_{*i}\gamma_i$ and $k_c \gg k_{pi}^2$, where $k_{pi} = \omega_{pi}/c$, $k_c = \omega_c/c$, $\omega_{pi}^2 = 4\pi e^2 N_i/m$ is the squared plasma frequency, and $\omega_c = eB_0/mc$ is the cyclotron frequency, dispersion relation (10) can be simplified to

$$(\varepsilon_{\parallel 2}S_{22} - \varepsilon_{\parallel 1}S_{12})(S_{21} - S_{11}) = \frac{V}{kR}(\zeta_*^*(S_{21} - S_{11}) + \zeta_*(\varepsilon_{\parallel 2}S_{22} - \varepsilon_{\parallel 1}S_{12})) - \frac{V^2}{k^2 R^2}(\zeta_*\zeta_*^* - \tau^2). \quad (11)$$

Now, we determine the growth rates of the instability of a relativistic magnetized plasma stream associated with the excitation surface waves. For a cold magnetized plasma stream, the condition for the surface waves to be in resonance with the longitudinal space-charge waves, $k = k_0 + k_*$ (where $k_0 = \beta k_z$ and $|k_*| \ll k_0$), splits into the following resonance conditions:

$$k = k_0 + sk_c/\gamma + k_s, \quad |k_s| \ll k_0 + sk_c/\gamma, \quad (12)$$

$$s = 0, \pm 1.$$

The $s = 0$ resonance is commonly referred to as a Cherenkov resonance, and the $s = 1$ and $s = -1$ resonances are called cyclotron resonances associated with the normal and anomalous Doppler effect, respectively. Among all these resonances, the $s = 0$ resonance is the only one that can give rise to the instability of a nonrotating stream. The other resonances cannot result in an instability, because the energy of the transverse motion of the beam electrons in a magnetic field is equal to zero and, thus, cannot decrease during the excitation of

surface waves, as they should for $s = \pm 1$. For this reason, waves with phase velocities above the speed of light cannot be excited, because condition (12) fails to hold for $n \leq 0$ and $\omega > |k_z|c$ [29].

We consider the Cherenkov resonance of surface waves with longitudinal beam oscillations: $k = k_0 + k_*$, where $k_0 = \beta k_z$ and $|k_*| \ll k_0$. Using the above expressions for the elements of the permittivity tensor of a moving cold plasma, we obtain the following two weakly coupled equations for the E and H waves: $Q_{E2} \approx 0$ and $Q_{H1} \approx 0$. The equation for the E waves can be solved analytically:

$$k_* = \frac{k_0}{\beta^2 \gamma^2} \left(1 \pm i \frac{V}{R} \sqrt{\frac{k_{p2}^2 \beta^2}{k_c^2 k_0^2 \varepsilon_{\perp 2} \gamma^3}} \right). \quad (13)$$

The condition $|k_*| \ll k_0$ implies that, for solution (13) to be valid, it is necessary that $\gamma \gg 1$. For nonzero modes, solution (13) is complex, which indicates that the stream is unstable against non-axisymmetric surface waves.

For an arbitrarily strong longitudinal magnetic field and $|k_*| \ll k_0$, dispersion relation (11) splits into the following dispersion relations for the E and H waves:

$$\varepsilon_{\parallel 2}S_{22} - \varepsilon_{\parallel 1}S_{12} = \frac{V}{kR}\zeta_*^*,$$

$$S_{21} - S_{11} = 0,$$

where

$$S_{22} = \frac{kI'_v(T_{E2}R)}{T_{E2}I_v(T_{E2}R)}, \quad S_{12} = \frac{kK'_v(T_{E1}R)}{T_{E1}K_v(T_{E1}R)},$$

$$S_{21} = \frac{kI'_v(T_{H2}R)}{T_{H2}I_v(T_{H2}R)}, \quad S_{11} = \frac{kK'_v(T_{H1}R)}{T_{H1}K_v(T_{H1}R)},$$

$$T_{H2}^2 = T_{H1}^2 = \frac{k_0^2}{\beta^2 \gamma^2}, \quad T_{E1,2}^2 = \varepsilon_{\parallel 1,2} \frac{k_0^2}{\beta^2 \gamma^2}.$$

For $\varepsilon_{\parallel 1} > 0$, the stream can be unstable against E-type surface waves:

$$k_{*1,2} = \frac{\varepsilon_{\parallel 1}S_{12}V k_{p2}^2 \beta^2 \pm k_{p2}\beta \sqrt{V^2 \beta^4 k_{p2}^2 + (\beta^2 - \varepsilon_{\parallel 1}^2 S_{12}^2 / \gamma^2) k_0 R k_c}}{k_0 R k_c (\varepsilon_{\parallel 1}^2 S_{12}^2 - \gamma^2 \beta^2)}.$$

For the onset of the Cherenkov instability associated with the excitation of the zeroth mode of a plasma waveguide, it is necessary to satisfy the condition

$\epsilon_{\parallel 1}^2 S_{12}^2 - \gamma^2 \beta^2 > 0$, in the case of which the instability growth rate is equal to

$$\delta = k_{p2} \beta \gamma^{-1} (\epsilon_{\parallel 1}^2 S_{12}^2 - \gamma^2 \beta^2)^{-1/2}.$$

The condition for the excitation of non-axisymmetric surface waves has the form

$$\epsilon_{\parallel 1}^2 S_{12}^2 - \gamma^2 \beta^2 - \frac{v^2 \gamma^2 \beta^4 k_{p2}^4}{k_0^2 R^2 k_c^2} > 0. \quad (14)$$

Recall that, in an unmagnetized plasma, a nonresonant instability occurs at $v \geq v_{cr}$. In contrast, in a strongly magnetized plasma, the modes with numbers $v < v_{cr}$ are unstable; the larger the ratio of the mode number to the stream radius, the lower the instability growth rate. In the ultrarelativistic limit, condition (14) for the onset of a resonant instability fails to hold and the application of a strong longitudinal magnetic field suppresses instabilities associated with the transverse motion of the plasma electrons [30].

CONCLUSION

We have considered model problems in order to analyze the characteristic features of electromagnetic surface waves that can be excited at the interface between a relativistic plasma stream and an immobile plasma and that give rise to a nonresonant hydrodynamic instability of the stream in the frequency range in which the permittivity is positive on both sides of the discontinuity.

For the stable surface waves of a plane plasma slab moving at a relativistic velocity in an immobile plasma, there is a critical angle between the wave vector of the surface wave and the direction of the slab motion; this angle corresponds to the transition from slow surface waves to fast waves. In cylindrical geometry, the role of the critical parameter for the existence of stable slow surface waves is played by the lowest mode number below which the fast waves can only occur.

The growing surface waves give rise to the instability of a relativistic plasma slab; the instability growth rate depends on the angle between the wave vector of the surface waves and the direction of the slab motion. The critical angles above which symmetric and anti-symmetric (about the median plane of the slab) surface waves start growing are different; the angular range in which the surface waves are stable is governed by the smaller of the critical angles. In a cylindrical plasma waveguide with given values of the velocity and density of the plasma stream, the critical parameter for the instability is the transverse size of the waveguide at which the lowest mode is excited. Applying a strong

longitudinal magnetic field suppresses higher unstable modes that are associated with the transverse displacement of plasma particles.

In relativistic plasma microwave electronics, the surface waves under consideration can be used to generate electromagnetic waves in directions different from the direction of the plasma stream motion. The calculations carried out with allowance for the parasitic character of higher unstable modes in wakefield accelerators show that an electron beam should be additionally focused, e.g., by applying a strong external longitudinal magnetic field.

REFERENCES

1. G. T. Markov and A. F. Chaplin, *Excitation of Electromagnetic Waves* (Énergiya, Moscow, 1967).
2. S. Solimeno, B. Crosignani, and P. Di Porto, *Diffraction and Confinement of Optical Radiation* (Academic, Orlando, 1986; Mir, Moscow, 1989).
3. K. A. Barsukov and A. D. Kanareikin, *Zh. Tekh. Fiz.* **55**, 1847 (1985) [*Sov. Phys. Tech. Phys.* **30**, 1082 (1985)].
4. L. D. Landau, *Dokl. Akad. Nauk SSSR* **44**, 151 (1944).
5. S. I. Syrovatskiĭ, *Zh. Éksp. Teor. Fiz.* **27**, 121 (1954).
6. M. I. D'yakonov, *Zh. Éksp. Teor. Fiz.* **94** (4), 119 (1988) [*Sov. Phys. JETP* **67**, 714 (1988)].
7. N. S. Averkiev and M. I. D'yakonov, *Opt. Spektrosk.* **68**, 1118 (1990) [*Opt. Spectrosc.* **68**, 653 (1990)].
8. W. Gai, A. D. Kanareykin, A. L. Kustov, and J. Simpson, in *Proceedings of the 6th Workshop on Advanced Accelerator Concepts, Fontana, WI, 1994*; *AIP Conf. Proc.* **335**, 463 (1994).
9. A. A. Rukhadze, P. V. Rybak, and V. P. Tarakanov, *Relativistic Plasma Microwave Electronics* (Nauka, Moscow, 1994) [*Tr. Inst. Obshch. Fiz. Ross. Akad. Nauk* **45** (1994)].
10. A. N. Kondratenko and V. M. Kuklin, *Foundations of Plasma Electronics* (Énergoatomizdat, Moscow, 1988).
11. E. P. Lee, *Phys. Fluids* **21**, 1327 (1978).
12. M. Lampe, G. Joyce, S. P. Slinker, and D. H. Wittum, *Phys. Fluids B* **5**, 1888 (1993).
13. D. H. Wittum, *Phys. Fluids B* **5**, 4432 (1993).
14. V. D. Shafranov, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Gosatomizdat, Moscow, 1963; Consultants Bureau, New York, 1967), Vol. 3.
15. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).
16. L. S. Bogdankevich, M. V. Kuzelev, and A. A. Rukhadze, *Usp. Fiz. Nauk* **133**, 3 (1981) [*Sov. Phys. Usp.* **24**, 1 (1981)].
17. V. D. Pikulin and N. S. Stepanov, *Zh. Tekh. Fiz.* **45**, 2288 (1975) [*Sov. Phys. Tech. Phys.* **20**, 1429 (1975)].
18. V. G. Gavrilenko, G. A. Lupanov, and N. S. Stepanov, *Izv. Vyssh. Uchebn. Zaved., Radiofiz.* **13**, 700 (1970).

19. A. D. Kanareikin and I. L. Sheinman, *Pis'ma Zh. Tekh. Fiz.* **22** (2), 61 (1996) [Tech. Phys. Lett. **22**, 74 (1996)].
20. A. D. Kanareikin and I. L. Sheinman, *Pis'ma Zh. Tekh. Fiz.* **23** (5), 76 (1997) [Tech. Phys. Lett. **23**, 201 (1997)].
21. L. S. Bogdankevich and A. A. Rukhadze, *Usp. Fiz. Nauk* **103**, 609 (1971) [Sov. Phys. Usp. **14**, 1631 (1971)].
22. A. A. Rukhadze, L. S. Bogdankevich, S. E. Rosinskii, and V. G. Rukhlin, *Physics of High-Current Relativistic Electron Beams* (Atomizdat, Moscow, 1980).
23. V. L. Ginzburg and A. A. Rukhadze, *Waves in Magnetized Plasmas* (Nauka, Moscow, 1975).
24. A. F. Aleksandrov, L. S. Bogdankevich, and A. A. Rukhadze, *Oscillations and Waves in Plasma Media* (Mosk. Gos. Univ., Moscow, 1990).
25. A. N. Kondratenko, *Surface and Internal Waves in Bounded Plasmas* (Énergoatomizdat, Moscow, 1988).
26. B. A. Aronov, L. S. Bogdankevich, and A. A. Rukhadze, *Zh. Tekh. Fiz.* **43**, 2493 (1973) [Sov. Phys. Tech. Phys. **18**, 1569 (1973)].
27. M. V. Kuzelev, F. Kh. Mukhametzyanov, and A. G. Shkvarunets, *Fiz. Plazmy* **9**, 1137 (1983) [Sov. J. Plasma Phys. **9**, 655 (1983)].
28. M. V. Kuzelev, A. A. Rukhadze, P. S. Strelkov, and A. G. Shkvarunets, *Usp. Fiz. Nauk* **146**, 709 (1985) [Sov. Phys. Usp. **28**, 724 (1985)].
29. B. N. Breizman, in *Reviews of Plasma Physics*, Ed. by B. B. Kadomtsev (Énergoizdat, Moscow, 1987; Consultants Bureau, New York, 1990), Vol. 15.
30. R. Keinings and M. E. Jones, *Phys. Fluids* **30**, 252 (1987).

Translated by O. Khadin

A Trapped Wave in a Ferroelectric with a Moving Strip Domain

O. Yu. El'meshkin and N. S. Shevyakhov

*Institute of Radio Engineering and Electronics, Ul'yanovsk Branch, Russian Academy of Sciences,
Ul'yanovsk, 432011 Russia*

e-mail: ufire@mv.ru

Received January 10, 2000

Abstract—The existence of a shear homogeneous plane wave trapped by a strip domain moving uniformly in a tetragonal ferroelectric crystal is demonstrated. Using the nonrelativistic quasi-static approximation and a change to the moving frame of reference, the solution is obtained and the dispersion relation is derived for the spatial spectrum of the trapped wave. The uniqueness of the solution is established in the case of the multiple degeneracy of the roots of the characteristic equation when the trapped wave contains an attached wave in the strip domain. The trapped wave is shown to correspond to the antisymmetric mode of the electroacoustic wave of the strip domain at the initial point of the spectrum branch. The specific features of the spatial resonance of the trapped wave with a shear wave obliquely incident on the strip domain are considered. © 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

From the viewpoint of acoustic spectroscopy of polydomain crystals [1], the acoustodomain interaction in ferroelectrics becomes increasingly attractive as its mechanisms become better understood. For example, on the basis of controlled domain structures, a variety of tunable data-processing devices operating on acoustic waves were proposed [2–4]. When such devices are switched or tuned, the domain boundaries (DB) in them can move, which should be taken into account in considering the effects of the acoustodomain interaction.

The acoustodomain interaction that occurs under the conditions of a given uniform motion of domain boundaries had been studied earlier [5] for the case of the shear wave refraction by a 180° domain wall in a BaTiO₃-type ferroelectric. The results of this study were used to evaluate the efficiency of the parametric transformation of a tone signal by the transverse motion of DB [6]. In particular, it was shown that a frequency (Doppler) transformation of acoustic pumping with a rotation of the wave normal in the direction of the DB motion is possible. This effect manifests itself as the shear wave being engaged in the transverse propagation by the moving DB because of the coupling through the piezoelectric polarization charges. Therefore, a concept was put forward [7] that ferroelectrics with moving DB exhibit a special kind of natural wave motion, which was called the trapped wave. Being similar to a homogeneous harmonic wave propagating at an angle to the trapping domain walls in the direction of their motion, the trapped wave is characterized by the transverse propagation determined solely by the DB motion and manifests itself as a transferred disturbance.

Since the attempts to represent the carrier of the trapped wave as a single moving DB failed, it was

assumed [7] that this role is played by a strip domain (SD) formed by two parallel 180° domain walls. The theoretical results obtained with this model [7] were not quite faultless because of an accidental error. The aim of the study described in this paper is to settle the doubts upon the existence of the trapped wave and to describe the properties of the latter and its relationship with other types of natural waves. To make the results more evident, we consider the wave motion not in the laboratory frame of reference as in [7], but in the moving frame of reference connected with the moving SD (below, we call it the SD rest frame for brevity). However, in some cases, we compare the results obtained in the two frames of reference when such a comparison is worthwhile.

BASIC EQUATIONS AND BOUNDARY CONDITIONS

We consider a ferroelectric of class 4 mm with the fourfold axis being parallel to the z -axis of the laboratory coordinate system $xOyz$. The ferroelectric contains an SD with two 180° DB that are parallel to the (010) plane and are thin as compared to the wavelength: $k\Delta \ll 1$, where k is the wave number of shear waves and Δ is the geometric thickness of the DB. In order to eliminate any noticeable structural sensitivity of the DB to the motion conditions because of the appearance of not only spontaneous polarization but also spontaneous deformation in the domains [8, 9], we restrict our consideration to the SD velocities $\mathbf{V}_D \parallel y \parallel [010]$ that are not too close to the Cerenkov sound limit c_t , i.e., to the velocity of shear waves in a single-domain crystal. We also assume that the ferroelectric is far from the phase transition. In the conditions specified above, the motion of the SD can be considered as preset with $V_D = \text{const}$,

and the DB can be considered as thin, structureless 180° domain walls with the current coordinates

$$y_1 = V_D t, \quad y_2 = V_D t + d, \quad (1)$$

where t is time and d is the thickness of the SD.

The latter assumption is known [9] to be appropriate in the case of DB in BaTiO₃-type ferroelectrics within the entire ultrasonic frequency range. By convention [1, 10–13], a ferroelectric with an SD can be likened to a piezoelectric crystal twinned along the planes $y = y_{1,2}$. The antiparallel orientation of the polar fourfold directions relative to the z -axis within the SD and outside it will be expressed in this case by a change of sign of the acting piezoelectric modulus e_{15} : $e_{15} = e$ for $y_1 < y < y_2$ and $e_{15} = -e$ for $y < y_1$ and $y > y_2$, where e is the magnitude of the piezoelectric modulus. We assume that all other properties of the specified regions of the crystal are identical.

For constructing the solution in the form of a homogeneous plane wave that is characterized by the displacements \mathbf{u} , $\mathbf{u}_j \parallel z$ and trapped by the moving SD, it is convenient to use the SD rest frame $\tilde{x}0\tilde{y}\tilde{z}$. In this frame of reference, the SD does not move, while the ferroelectric represents a medium continuously moving as a whole (drifting in the transverse direction) with the velocity V_D in opposition to the real motion.

Generally speaking, the “motion” of the medium means the relativistic nature of the solution. However, in view of the evident limitations $V_D/c \ll 1$ and $c_l/c \ll 1$ (c is the velocity of light), the relativistic corrections involved in the electrodynamics equations [14] and in the equations of the theory of elasticity [15] (including the Minkowski constitutive relations for a piezoelectric medium) will be no greater than the usually neglected corrections for the electric field delays [10–13]. Therefore, in the adopted nonrelativistic quasi-static approximation, the relations between the coordinates of the laboratory frame of reference and those of the SD rest frame can be expressed through the Galilean transformation:

$$\tilde{x} = x, \quad \tilde{y} = y - V_D t, \quad \tilde{z} = z, \quad \tilde{t} = t. \quad (2)$$

In the case of the propagation of shear waves with a horizontal polarization in the (001) plane of the ferroelectric, we can use the following equations in the laboratory frame of reference [5, 10–13]:

$$\left(\frac{1}{c_t^2} \frac{\partial^2}{\partial t^2} - \nabla^2 \right) \begin{pmatrix} u \\ u_j \end{pmatrix} = 0, \quad \nabla^2 \begin{pmatrix} \Phi \\ \Phi_j \end{pmatrix} = 0. \quad (3)$$

Here, u and u_j ($j = 1$ for $y < y_1$ and $j = 2$ for $y > y_2$) are the shear displacements within the SD and outside it; Φ and Φ_j are the components of the electric potentials

$$\varphi = \frac{4\pi e}{\varepsilon} u + \Phi, \quad \varphi_j = -\frac{4\pi e}{\varepsilon} u_j + \Phi_j, \quad (4)$$

these components representing the statistical fields induced in the SD and outside it by the piezoelectric polarization charges located at the boundaries $y = y_{1,2}$; $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2$; and ε is the dielectric constant.

In the nonrelativistic quasi-static approximation in the laboratory frame of reference, it is appropriate to apply the conventional continuity requirements [5, 10–13] at the DB for the shear displacements, the potentials given by Eqs. (4), the stress tensor components

$$T_{yz} = \lambda \frac{\partial u}{\partial y} + e \frac{\partial \Phi}{\partial y}, \quad T_{yz}^{(j)} = \lambda \frac{\partial u_j}{\partial y} - e \frac{\partial \Phi_j}{\partial y}, \quad (5)$$

and the y -components of the electric inductions

$$D_y = -e \frac{\partial \Phi}{\partial y} + 4\pi e \frac{\partial u}{\partial y}, \quad D_y^{(j)} = -e \frac{\partial \Phi_j}{\partial y} - 4\pi e \frac{\partial u_j}{\partial y}. \quad (6)$$

The superscripts $j = 1, 2$ in parentheses indicate the regions outside the SD for the quantities from Eqs. (5) and (6), and λ represents the elastic modulus of the crystal. The quantities involved in the boundary conditions contain no derivatives with respect to time. Equations (2) lead to the following change of differential operators:

$$\frac{\partial}{\partial x} \rightarrow \frac{\partial}{\partial \tilde{x}}, \quad \frac{\partial}{\partial y} \rightarrow \frac{\partial}{\partial \tilde{y}}, \quad \frac{\partial}{\partial t} \rightarrow \frac{\partial}{\partial \tilde{t}} - V_D \frac{\partial}{\partial \tilde{y}}. \quad (7)$$

Hence, if we change to the SD rest frame, the form of the boundary conditions will remain the same. It should only be remembered that the boundary conditions are transferred to the planes $\tilde{y}_j = (j - 1)d$, where $j = 1, 2$. Taking into account Eqs. (4)–(6), we obtain

$$\begin{aligned} (u - u_j)|_{\tilde{y}=\tilde{y}_j} &= 0, \quad \left(\frac{\partial \Phi}{\partial \tilde{y}} - \frac{\partial \Phi_j}{\partial \tilde{y}} \right) \Big|_{\tilde{y}=\tilde{y}_j} = 0, \\ \left(\frac{4\pi e}{\varepsilon} u + \Phi \right) \Big|_{\tilde{y}=\tilde{y}_j} &= \left(-\frac{4\pi e}{\varepsilon} u_j + \Phi_j \right) \Big|_{\tilde{y}=\tilde{y}_j}, \quad (8) \\ \left(\lambda^* \frac{\partial u}{\partial \tilde{y}} + e \frac{\partial \Phi}{\partial \tilde{y}} \right) \Big|_{\tilde{y}=\tilde{y}_j} &= \left(\lambda^* \frac{\partial u_j}{\partial \tilde{y}} - e \frac{\partial \Phi_j}{\partial \tilde{y}} \right) \Big|_{\tilde{y}=\tilde{y}_j}, \end{aligned}$$

where $\lambda^* = \lambda + 4\pi e^2/\varepsilon$.

In view of expressions (7), Eqs. (3) will take the following form in the SD rest frame:

$$\left[\frac{1}{c_t^2} \left(\frac{\partial}{\partial \tilde{t}} - V_D \frac{\partial}{\partial \tilde{y}} \right)^2 - \tilde{\nabla}^2 \right] \begin{pmatrix} u \\ u_j \end{pmatrix} = 0, \quad \tilde{\nabla}^2 \begin{pmatrix} \Phi \\ \Phi_j \end{pmatrix} = 0. \quad (9)$$

Here, $\tilde{\nabla}^2 = \partial^2/\partial \tilde{x}^2 + \partial^2/\partial \tilde{y}^2$. Formally, the problem under consideration consists in solving Eqs. (9) with the boundary conditions given by Eqs. (8).

A TRAPPED WAVE IN THE SD REST FRAME

The advantage of using the SD rest frame consists in the possibility of constructing the solution without "guessing" at the behavior of its time dependences in different parts of the crystal. By virtue of the temporal homogeneity of the boundary conditions (8), all these dependences will be identical, and, under the assumption that the oscillations are harmonic, they will be governed by the common phase factor $\exp(-i\Omega\tilde{t})$, where Ω is the oscillation frequency in the SD rest frame. Adopting the general coordinate dependence of the solution to Eqs. (9) in the form of $\exp(ik_{\parallel}\tilde{x})$ and taking into account that the trapped wave must have the structure of an ordinary homogeneous plane wave [7], we can write

$$\begin{aligned}\Phi, \Phi_j &\sim \exp[i(k_{\parallel}\tilde{x} - \Omega\tilde{t})] \exp(qy), \\ u, u_j &\sim \exp[i(k_{\parallel}\tilde{x} - \Omega\tilde{t})] \exp(ik_{\perp}\tilde{y}).\end{aligned}$$

In this case, from Eqs. (9), we obtain

$$q^2 - k_{\parallel}^2 = 0, \quad c_t^2(k_{\parallel}^2 + k_{\perp}^2) = (\Omega + k_{\perp}V_D)^2. \quad (10)$$

The quantities k_{\parallel} and k_{\perp} have the meaning of the longitudinal and transverse components of the wave vector \mathbf{k} of the trapped wave, respectively.

From the first of Eqs. (10), we obtain $q = \pm k_{\parallel}$. Therefore, with allowance for the requirement that the potentials Φ and Φ_j be limited in their domains of definition, we write

$$\begin{aligned}\Phi &= \exp(i\xi)[A \exp(-k_{\parallel}\tilde{y}) + B \exp(k_{\parallel}\tilde{y})], \quad 0 < \tilde{y} < d, \\ \Phi_1 &= F_1 \exp(i\xi) \exp(k_{\parallel}\tilde{y}), \quad \tilde{y} < 0, \\ \Phi_2 &= F_2 \exp(i\xi) \exp(-k_{\parallel}\tilde{y}), \quad \tilde{y} > d.\end{aligned} \quad (11)$$

In Eqs. (11), $\xi = k_{\parallel}\tilde{x} - \Omega\tilde{t}$ is the phase of oscillations of the trapped wave, this phase being measured along the SD; A , B , and $F_{1,2}$ are the amplitude constants.

With the given k_{\parallel} and Ω and with the limitation $\beta = V_D/c_t < 1$ introduced above, the second of Eqs. (10) yields the following representation of k_{\perp} :

$$k_{\perp} = \frac{\beta\Omega \pm \sqrt{\Omega^2 - k_{\parallel}^2 c_t^2 (1 - \beta^2)}}{c_t(1 - \beta^2)}. \quad (12)$$

Note that, in constructing the solution, one should take into account both roots given by Eq. (12), because neglecting one of them overdetermines the problem by boundary conditions (8). However, the choice of the values of k_{\perp} and Eq. (12) is not arbitrary but is subject to the causality condition: the transverse propagation of the trapped wave is caused by the SD motion. This, firstly, means that the condition $k_{\perp} > 0$, which provides the coincidence of the transverse propagation direction of the trapped wave with the direction of the SD

motion, is satisfied and, secondly, leads to the limiting condition

$$\lim_{\beta \rightarrow 0} k_{\perp} = 0, \quad (13)$$

which corresponds to the termination of the transverse propagation of the trapped wave with the termination of the SD motion.

In selecting the roots k_{\perp} in [7], the stronger requirement (13) was not considered. However, this fact should not affect the final result obtained there, because the structure of the solution (see Eq. (4) in [7]) obtained for the trapped wave was correct. However, because of the slip in writing the sign of one of the terms with the phase exponent, the dispersion relation for the trapped wave was written in [7] with an error, which resulted in a wrong conclusion about the multibranch structure of the solution and the existence of the corresponding discrete set of modes of trapped waves refracted by the moving SD. The correction of the results obtained in [7] inevitably leads to condition (13).

It can be easily established that the condition $k_{\perp} > 0$ ($V_D > 0$) is satisfied for both roots (12) in the frequency range $\Omega \in [k_{\parallel}c_t(1 - \beta^2)^{1/2}, k_{\parallel}c_t]$ with its lower bound included. The latter fact is of crucial importance, because, according to Eq. (12), just at this lower bound, i.e., at

$$\Omega = k_{\parallel}c_t(1 - \beta^2)^{1/2}, \quad (14)$$

simultaneous fulfillment of the limiting condition (13) is possible. Relationship (14), which determines the existence of only multiple roots (12) with the magnitude

$$k_{\perp} = \frac{\beta\Omega}{c_t(1 - \beta^2)} \quad (15)$$

can be considered as the form of the relation that occurs between the main spectral parameters k_{\parallel} and Ω at a current value of β and is peculiar to the trapped wave.

Taking into account the twofold degeneracy of the roots k_{\perp} given by Eq. (12) and the required boundedness of the solution u_j to the first of Eqs. (9) in the outer regions, we obtain

$$\begin{aligned}u_1 &= U_1 \exp i(\xi + k_{\perp}\tilde{y}), \quad \tilde{y} < 0, \\ u_2 &= U_2 \exp i(\xi + k_{\perp}\tilde{y}), \quad \tilde{y} > d, \\ u &= (V\tilde{y} + W) \exp i(\xi + k_{\perp}\tilde{y}), \quad 0 < \tilde{y} < d.\end{aligned} \quad (16)$$

Expressions (11) and (16), where W is the amplitude of displacements and V is the amplitude of the shear strain $2u_{yz}$ in the SD, completely describe the structure of the trapped wave. We note that, according to Eq. (16), the part of the trapped wave with the amplitude coefficient V can be interpreted as an attached wave, i.e., a degenerate waveguide mode [8], but only within the SD. An extension of this interpretation to the

whole wave (as in [7]) is obviously inappropriate. This fact caused the introduction of the term “trapped wave,” which is more suitable for describing the phenomenon under study.

DISPERSION RELATION AND THE SPECTRUM OF THE TRAPPED WAVE

To derive the dispersion relation for the trapped wave, we substitute Eqs. (11) and (16) into the boundary conditions (8). The resulting system of homogeneous algebraic equations in the amplitudes of partial waves u_j and u and accompanying electric oscillations (11) must satisfy the solvability condition $\text{Det} = 0$, which is equivalent to satisfying the following relationships between the amplitudes U_j of displacements at the SD boundaries:

$$\frac{U_1}{U_2} = \frac{1 - 2\mathcal{H}^2 k_{\parallel} d \exp(ik_{\perp} d - k_{\parallel} d)}{1 - 2\mathcal{H}^2 k_{\parallel} d} \quad (17)$$

$$= \frac{\exp(ik_{\perp} d)(1 - 2\mathcal{H}^2 k_{\parallel} d)}{\exp(ik_{\perp} d) - 2\mathcal{H}^2 k_{\parallel} d \exp(-k_{\parallel} d)}.$$

After some transformations, from Eq. (17) we obtain the desired dispersion relation in the form of a transcendental equation

$$\mathcal{H}^2 k_{\parallel} d [1 - \exp(-2k_{\parallel} d)] + \exp(-k_{\parallel} d) \cos(k_{\perp} d) - 1 = 0, \quad (18)$$

which, in addition to Eqs. (14) and (15), determines the relationship between the components of the wave vector of the trapped wave.

At first glance, the imposition of relation (18) in addition to Eqs. (14) and (15) leads to a contradiction: the substitution of Eq. (15) into Eq. (18) yields a non-linear dependence $k_{\parallel}(\Omega)$, whereas, in essence, expressions Eqs. (14) and (15) point to the linear character of the frequency dependences of the wave vector components of the trapped wave. The contradiction is eliminated if we assume that β does not remain constant when we move along the linear frequency spectrum of the trapped wave, i.e., along the dispersion line for shear bulk waves. In other words, there exist some parametric dependences $\beta(k_{\parallel})$ and $\beta(k_{\perp})$, or vice versa, that provide the tuning of solution (18) to the form of the linear frequency dependences determined by Eqs. (14) and (15) in the situation under study. Namely, from Eqs. (14) and (15), by eliminating Ω , we obtain

$$k_{\perp} = k_{\parallel} \frac{\beta}{\sqrt{1 - \beta^2}}, \quad k = \sqrt{k_{\parallel}^2 + k_{\perp}^2} = \frac{k_{\parallel}}{\sqrt{1 - \beta^2}}. \quad (19)$$

The inverse parametric relationship $k_{\parallel}(\beta)$ for the trapped wave can be considered in this case as the

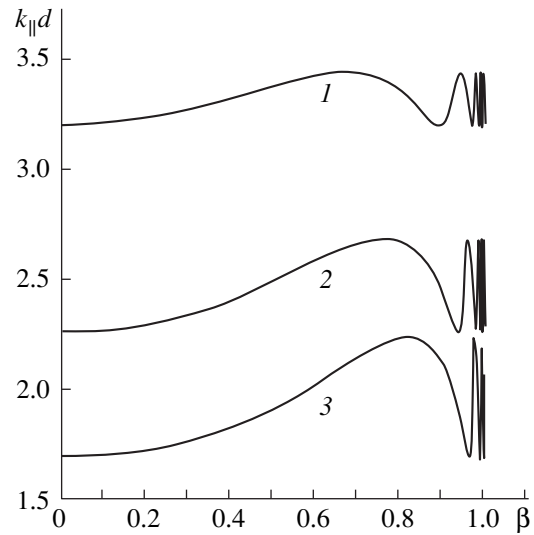


Fig. 1. Dependence of the normalized longitudinal component of the wave vector $k_{\parallel}d$ of the trapped wave on β for $\mathcal{H}^2 = (1) 0.3, (2) 0.4, \text{ and } (3) 0.5$.

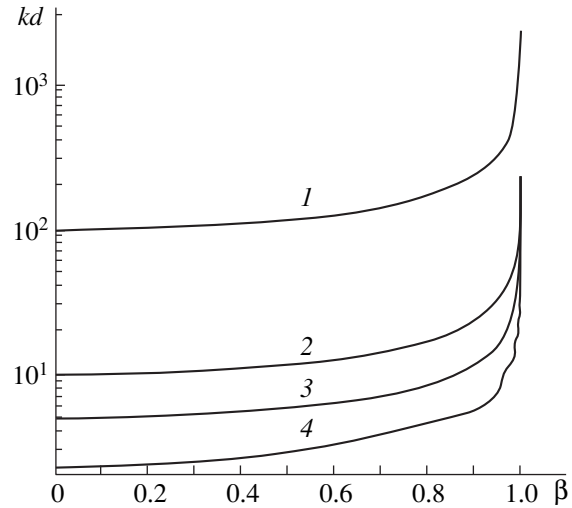


Fig. 2. Dependence of the normalized wave number kd of the trapped wave on β for $\mathcal{H}^2 = (1) 0.01, (2) 0.1, (3) 0.2, \text{ and } (4) 0.4$.

solution to the equation

$$\mathcal{H}^2 k_{\parallel} d [1 - \exp(-2k_{\parallel} d)] + \cos\left(k_{\parallel} d \frac{\beta}{\sqrt{1 - \beta^2}}\right) \exp(-k_{\parallel} d) - 1 = 0, \quad (20)$$

which follows from Eq. (18). The typical forms of the parametric dependences $k_{\parallel}(\beta)$ and $k(\beta)$ obtained by numerically solving Eq. (20) with allowance for Eq. (19) is shown in Figs. 1 and 2, respectively, for different values of \mathcal{H}^2 .

The calculation of the characteristics of the trapped wave in the SD rest frame can be performed in the

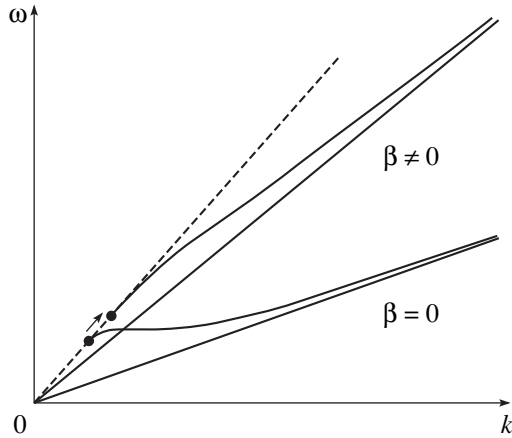


Fig. 3. Qualitative representation of the dispersion branches of the antisymmetric mode of the electroacoustic wave trapped by a moving strip domain and by a stationary one.

reverse sequence: setting β and determining k_{\parallel} as a root of the transcendental equation (20), we calculate k and k_{\perp} by Eqs. (19), and the frequency Ω is obtained by any of Eqs. (14) or (15). The latter option allows us to check the results of our calculations by comparing the values of Ω obtained by the two formulas. To determine the characteristics of the trapped wave in the laboratory frame of reference, we can use the relationship

$$\omega = \Omega + \mathbf{kV}_D, \tag{21}$$

which follows from Eq. (16) when we change from the SD rest frame to the laboratory frame of reference according to Eqs. (2). We note that, in Eq. (21), ω has the meaning of the oscillation frequency of the trapped wave in the laboratory frame of reference.

In law (21) governing the transformation of the spectrum of the trapped wave, we have $\mathbf{kV}_D > 0$ as a consequence of the coincidence of the transverse propagation direction of the trapped wave with the direction of the SD motion ($k_{\perp} > 0, V_D > 0$), and, therefore, we will always have $\omega > \Omega$. This result can be considered as a positive Doppler frequency shift that manifests itself locally but identically at every point of the medium when oscillations arriving from the oncoming source (the trapped wave) are received by a stationary receiver (the crystal). In fact, in the problems on the natural wave motion of the type under study, the source of oscillations and the receiver are not explicitly described, and they can be put in correspondence with nothing but the wave itself and the crystal. Thus, the Doppler source and receiver prove to be continuously distributed over the same spatial region and are interpreted as such just in the local sense. The origin of the frequency shift in the conditions under study is evident: because of the coupling with the SD through the polarization charges at the boundaries, the trapped wave is engaged in the motion along the y -axis everywhere in

the same way, this motion having the form of a translational transfer of the wave relative to the crystal.

Recasting the quantity $\mathbf{kV}_D = k_{\perp}V_D$ in Eq. (21) with allowance for expression (15) for k_{\perp} , we obtain a relation between the frequencies:

$$\omega = \frac{\Omega}{\sqrt{1 - \beta^2}}. \tag{22}$$

Then, instead of Eqs. (14) and (15), we will have $k_{\parallel} = \omega(1 - \beta^2)^{1/2}/c_t$ and $k_{\perp} = \omega\beta/c_t$, and, hence, in the laboratory frame of reference, we obtain $k = \omega/c_t$. As one would expect and as was mentioned above, the latter equality represents a linear spectrum of shear bulk waves in the ferroelectric. This result can be obtained directly from the second equality of Eqs. (10) by taking into account Eq. (21). In closing the description of the trapped wave in the laboratory frame of reference, we note that, when we change to this coordinate system, Eqs. (18)–(20) remain unchanged. Hence, Figs. 1 and 2 are suitable for discussing the properties of the trapped wave in both the SD rest frame and the laboratory frame of reference.

To determine the place of the trapped wave among the known types of natural waves, we consider the case of a static SD, $\beta = 0$, when the trapped wave front is orthogonal to the trapping boundaries ($k_{\perp} = 0, k_{\parallel} = k$) and the spectral behavior of the wave obeys (according to Eq. (20)) the equation

$$\mathcal{H}^2 k_{\parallel} d [1 + \exp(-k_{\parallel} d)] - 1 = 0. \tag{23}$$

From the type of deformation and the type of relation to the electrostatic fields, as well as from the noticeable effect of \mathcal{H}^2 on the downward displacement of the parametric dependences (which can be seen from Figs. 1 and 2), one can conclude that the trapped wave is in some sense similar to an electroacoustic wave trapped by the SD. According to the results reported in [12], one can see that Eq. (23) represents not only the dispersion relation for the trapped wave at $\beta = 0$, but also the condition of a complete delocalization for the oscillations of the antisymmetric mode of the electroacoustic wave at the initial point of the spectrum (see the dots in Fig. 3), where this point moves along the straight line representing the spectrum of the shear bulk waves (the dashed line in Fig. 3).

A similar interpretation can be applied to Eq. (20) by generalizing the above conclusion, i.e., the conclusion that the trapped wave is identical to the antisymmetric mode of the electroacoustic wave at the initial point of the spectrum, to the case $\beta \neq 0$. The qualitative representation of the changes that occur in the spectrum of the antisymmetric mode because of the SD motion is presented in Fig. 3. It illustrates the parametric dependence $k(\beta)$ (Fig. 2) for the trapped wave as a consequence of the motion of the initial point of the antisymmetric mode spectrum in the direction shown by the arrow along the linear spectrum of shear bulk waves. As

for the high-frequency asymptotes of the spectra of the antisymmetric mode, which are shown by solid straight lines, they represent the dispersion law for an electroacoustic wave at a single DB (the Maerfeld–Tournois mode [10]).

The asymmetry of the profile of shear displacements with respect to the middle plane $\tilde{y} = d/2$ is peculiar to the antisymmetric mode for a static SD [12], and it will manifest itself in the trapped wave by the antiphase relationship of the oscillation amplitudes in the shear waves at the SD boundaries: $U_1 = -U_2$. Using Eqs. (17) and (18), it is also possible to show that the equality $|U_1| = |U_2|$ is satisfied at any $\beta \neq 0$. However, the antiphase character of the shear displacement oscillations that occur in the trapped wave at the SD boundaries at $\beta \neq 0$ will be violated. The measure of the deviation from antiphase is the quantity $\varphi_u = \text{Arg}(U_1/U_2)$, which is estimated in the first period of the 2π -periodic phase functions of the trapped wave solution by the formula

$$\varphi_u = \pi - \arctan \frac{\sin(k_{\perp}d)2\mathcal{K}^2k_{\parallel}d \exp(-k_{\parallel}d)}{(1 - \mathcal{K}^2k_{\parallel}d)^2 + (\mathcal{K}^2k_{\parallel}d)^2 [1 - 2 \exp(-2k_{\parallel}d)]} \quad (24)$$

Figure 4 presents the typical dependences of φ_u on β that are calculated by Eqs. (20) and (24) for different values of \mathcal{K}^2 . These dependences show that, in the case of an acceptable level of electromechanical coupling in the crystal, the deviations φ_u from π are not too large, and, in total, the oscillations of shear displacements that occur in the trapped wave on different sides of the moving SD are close to antiphase oscillations.

SPATIAL RESONANCE WITH AN INCIDENT SHEAR WAVE

Both the fact that the wave vector of the trapped wave is noncollinear with the SD boundaries and the delocalization of oscillations form prerequisites for the spatial resonance of the trapped wave with a shear homogeneous plane wave incident at the angle $\Theta = \arctan(k_{\parallel}/k_{\perp})$ on a receding SD. In this case, the motion of the SD is a crucial factor, because, in the static case corresponding to $\beta = 0$, the spatial resonance with the trapped wave is only formally possible in the grazing incidence limit $\Theta \rightarrow \pi/2$ but is unrealizable in practice.

In considering the spatial resonance, it is natural to preset the characteristics of the incident shear wave u_1 from Eqs. (16) in the laboratory frame of reference and then to change to the SD rest frame. Thus, taking the wave vector of the incident wave in the form $\mathbf{k} = (k_x, k_y)$, where $k_x = k \sin \Theta$, $k_y = k \cos \Theta$, $k = \omega/c_t$, and Θ is the angle of incidence, we obtain an expression for the fre-

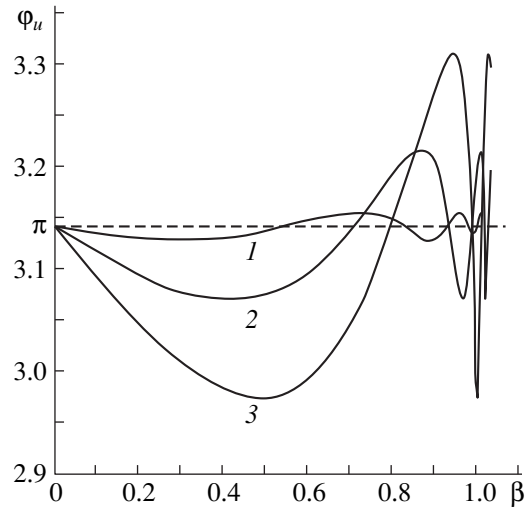


Fig. 4. Effect of the motion of the strip domain on the departure of the boundary values of the shear displacement amplitudes in the trapped wave from the antiphase state for $\mathcal{K}^2 =$ (1) 0.2, (2) 0.3, and (3) 0.4. The dashed line corresponds to the purely antiphase state.

quency Ω in the SD rest frame in accordance with Eq. (21):

$$\Omega = \omega(1 - \beta \cos \Theta). \quad (25)$$

In the problems of wave refraction by moving boundaries, the main requirement imposed on any refracted wave consists in its synchronism with the incident wave at the points of the moving boundary [16]. Therefore, if $\mathbf{k}' = (k'_x, k'_y)$ is the wave vector and Ω' is the frequency of the refracted wave in the SD rest frame, we have $\Omega' \equiv \Omega$ and $k'_x \equiv k_x$. The components k'_y can be determined from Eq. (12) by replacing k_{\parallel} and k_{\perp} by k_x and k'_y , respectively. Simple transformations yield

$$k'_y = \frac{\omega}{c_t} \cos \Theta, \quad (26)$$

$$k'_y = \frac{\omega}{c_t(1 - \beta^2)} [2\beta - \cos \Theta(1 + \beta^2)].$$

According to Eqs. (26), on each side of the moving SD boundary, $\tilde{y} = 0$ or $\tilde{y} = d$, a pair of differently refracted waves is formed in the general case. However, in the semibounded regions $\tilde{y} < 0$ and $\tilde{y} > d$, we should leave only one wave in each region by choosing the waves on the basis of the Mandel'shtam radiation principle for a nonequilibrium moving (in the SD rest frame) medium [17]. Without going into details, we note that the first of the refracted waves given by Eq. (26) should be eliminated from the region $\tilde{y} < 0$

and the second wave should be eliminated from the region $\tilde{y} > d$. Then, instead of Eq. (16), we write

$$\begin{aligned} u_1 &= U \exp i \xi [\exp(i k_y \tilde{y}) + R \exp(i k'_y \tilde{y})], \quad \tilde{y} < 0, \\ u_2 &= U T \exp i (\xi + k_y \tilde{y}), \quad \tilde{y} > d, \end{aligned} \quad (27)$$

$$u = U [V \exp(i k_y \tilde{y}) + W \exp(i k'_y \tilde{y})] \exp i \xi, \quad 0 < \tilde{y} < d.$$

Here, the primes marking the y -component of the wave vector of the first of the refracted waves (26) are omitted, because, when propagating in the regions $0 < y < d$ and $\tilde{y} > d$, this wave is identical to the incident wave ($k'_y = k_y$, $k'_x = k_x$). Correspondingly, the quantity k'_y involved in Eqs. (27) will be determined only by the second of Eqs. (26). Other quantities involved in Eqs. (27) are as follows: $\xi = k_x \tilde{x} - \Omega \tilde{t}$ is the amplitude of the incident wave, which is assumed to be known; and the quantities R and T have the meaning of the reflection and transmission coefficients for the shear wave in the case of a receding SD.

Substituting Eqs. (27) in combination with Eqs. (11), where k_{\parallel} should be replaced by k_x , into the boundary conditions (8) that remain unchanged, we obtain a system of inhomogeneous algebraic equations that allow us to determine the amplitude coefficients of the potentials of electrostatic oscillations and those of the shear refracted waves. Below, we restrict our consideration to the expression for R , which, with allowance for Eqs. (25) and (26), can be represented in the form

$$R = -2w \mathcal{H}^2 i \zeta [1 + \exp(i \zeta)] - Z \quad (28)$$

$$2w \mathcal{H}^2 (i 2 \zeta - Z) - \zeta^2.$$

Here, $\kappa = kd$, $w = \kappa \sin \Theta$, $\zeta = 2\kappa(\cos \Theta - \beta)/(1 - \beta^2)$, and

$$\begin{aligned} Z &= 2w \mathcal{H}^2 (1 - e^{-2w})(e^{i \zeta} - 1) \\ &+ i \zeta \cos(w - \zeta/2) e^{-w} e^{-i \zeta/2}. \end{aligned}$$

The angle of incidence $\Theta = \Theta^{**}$ that corresponds to the spatial resonance of the incident and trapped waves is determined by the expression

$$\cos \Theta^{**} = \beta, \quad (29)$$

which follows from Eq. (19), if we take into account that $k_{\perp} = k \cos \Theta^{**}$ and $k_{\parallel} = k \sin \Theta^{**}$. According to Eq. (29), in the spatial resonance conditions, the frequency (25) of the refracted waves is converted (as one would expect) to the trapped wave frequency given by Eq. (22). We note that, when $\Theta \rightarrow \Theta^{**}$, the transverse components k'_y of the wave vectors of both types of refracted waves (26) become identical: $k'_y \rightarrow k_y$. This fact reflects the multiple degeneracy of k_{\perp} from Eq. (12), this degeneracy being peculiar to the trapped wave.

For the following analysis, it is important to note that, in the limit $\Theta \rightarrow \Theta^{**}$, the quantity k'_y given by the second of Eqs. (26) changes its sign at $\Theta = \Theta^*$ ($\Theta^* < \Theta^{**}$, $\cos \Theta^* = 2\beta(1 + \beta^2)^{-1}$). If we denote $k'_y = k' \cos \Theta'$, where $k' = \omega'/c$, ω' is the frequency of the refracted wave in the laboratory frame of reference, and Θ' is the refraction angle, we can determine Θ^* as the angle separating the normal refraction region ($\Theta' < \pi/2$, $k'_y < 0$) from the region of refraction at obtuse angles ($\Theta' > \pi/2$, $k'_y > 0$). The possibility of the manifestation of the latter is the specific feature of the solutions to the problems of refraction of waves by receding boundaries [16].

The quintessence of the problem of refraction at obtuse angles, including the spatial resonance angle Θ^{**} , consists in the representation and interpretation of the refracted wave, which usually ($\Theta' < \pi/2$) proves to be (see the second term in Eq. (27) for u_1) an ordinary reflected wave. For the classical Einstein problem of a moving mirror, such a representation and interpretation was performed in [16] where the refracted wave at $\Theta' > \pi/2$ was interpreted as an infinitely long (because of its monochromatic nature) stub of radiation from secondary (Huygens) sources located on the mirror surface, and this stub of radiation following the mirror progressively lagged behind it. In this case, the refracted wave plays the role of a wave complementing the incident one. The corresponding refraction scheme is represented in terms of wave vectors in Fig. 5a.

The study of the refracted wave in [16] was facilitated by the preset uniqueness of the choice of the region of its propagation. Owing to the perfection (opacity) of the mirror, the refracted wave could propagate only on the same side of the mirror as the incident wave. Another situation was considered in [5] where, because of the identical properties of the ferroelectric on both sides of the moving DB, the region of propagation for the wave refracted at $\Theta' > \pi/2$ could coincide with both the region of propagation of the incident wave ($\tilde{y} < 0$) and the region of propagation of the directly transmitted shear wave ($\tilde{y} > 0$). With few insignificant exceptions, the first case corresponds to Fig. 5a. The second variant of the solution (Fig. 5b), which was considered in [5], corresponds to the birefringence of the shear wave by a receding DB in a ferroelectric.

Such a choice made in [5] was based on the fact that the solution was constructed directly in the laboratory frame of reference, where the dispersion law for shear waves,

$$\omega' = \mathbf{n}' \mathbf{k}' c_t, \quad \omega = \mathbf{n} \mathbf{k} c_t, \quad (30)$$

did not indicate any spectral anomalies, and, in estimating the direction of the energy transfer from the boundary, it seemed to be sufficient to proceed exclusively from the choice of the direction of wave propagation, i.e., from the wave normal vector \mathbf{n}' or \mathbf{n} . This approach

is suitable with the only addition that it is necessary to take into account the position of the moving domain boundary. The possibility of one or another propagation direction for the refracted wave can be estimated only relative to this boundary, which means that one has to change to the frame of reference fixed to the boundary and, finally, to take into account the changes in the dispersion properties of the medium in the DB frame of reference.

In view of the aforesaid, in constructing the solution corresponding to Eqs. (27), we reject the scheme of refraction shown in Fig. 5b and proposed in [5] and consider the type of refraction presented in Fig. 5a. Here, the refracting boundary is the backside of the receding SD (i.e., the DB at $\tilde{y} = 0$). Line 1 (in Fig. 5, it represents the incident shear wave) should be continued in this case beyond the boundary to represent the directly transmitted wave. By analogy with Eq. (21), the dispersion law for the refracted wave in the SD rest frame has the form $\Omega' = \omega' - \mathbf{k}'\mathbf{V}_D$. Expressing ω' and ω through Eqs. (30) and using the definition of the group velocity, we arrive at the equalities $V_{g_y} = n_y c_t - V_D$ and $V'_{g_y} = n'_y c_t - V_D$, where $n_y = \cos\Theta$ and $n'_y = k'_y/k'$. Taking into account that $k'_x = \omega \sin\Theta / c_t$ and k'_y is determined by the second of Eqs. (26), we obtain

$$V_{g_y} = c_t(\cos\Theta - \beta),$$

$$V'_{g_y} = -V_{g_y} \frac{1 - \beta^2}{1 + \beta^2 - 2\beta \cos\Theta}. \quad (31)$$

For $\Theta < \Theta^{**}$ ($\cos\Theta > \beta$), according to Eq. (31), we have $V_{g_y} > 0$ and $V'_{g_y} < 0$. According to the radiation principle [17], to provide the energy transfer from the boundary, it is necessary to exclude the possibility of the propagation of the refracted wave at $\tilde{y} > 0$ and the propagation of the incident-type wave at $\tilde{y} < 0$. The resulting refraction corresponds to the diagram shown in Fig. 5a. We note that Eq. (31) for V_{g_y} equally refers to the incident wave, which, in contrast to other waves, provides the energy transfer to the boundary $V_{g_y} > 0$ at $\tilde{y} < 0$. The termination of this process and, as a consequence, the termination of the interaction of the incident wave with the boundary at $\Theta \rightarrow \Theta^{**}$ ($V_{g_y} \rightarrow 0$, $V'_{g_y} \rightarrow 0$) are explained by the fact that, as is characteristic of the end point of the interval of obtuse refraction angles [5, 16], the wave in these conditions ceases catching up with the boundary.

Thus, returning to the object of our study, i.e., to the trapped wave, we can state that it cannot be excited by a shear wave incident on the SD at the angle Θ^{**} . The solution presented above (see Eqs. (27) and (28)) shows that, as we approach the point of the spatial resonance with the trapped wave ($\zeta \rightarrow 0$), the quantity R exhibits

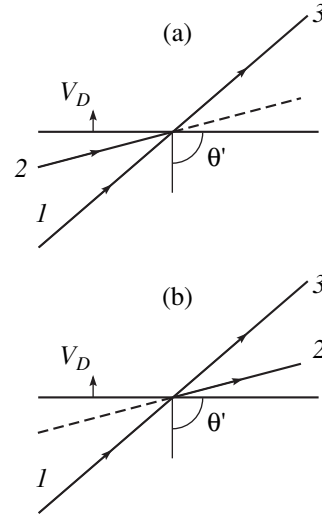


Fig. 5. Alternative variants of the interaction of a wave with a receding boundary at obtuse refraction angles $\Theta' > \pi/2$: (1) the incident wave, (2) the refracted wave, and (3) the directly transmitted wave.

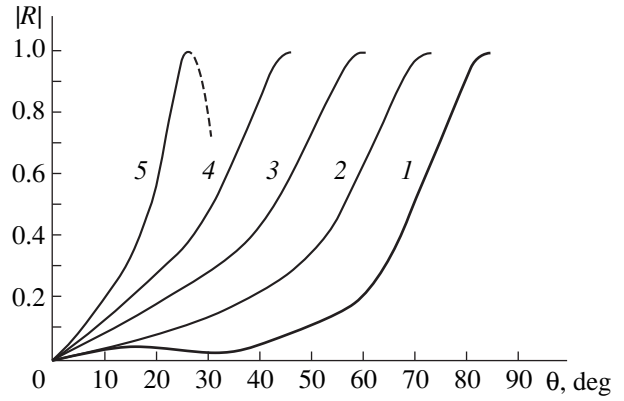


Fig. 6. Angular dependences of the absolute value of the reflection coefficient for the shear wave reflection from a receding strip domain at $\mathcal{H}^2 = 0.38$ and $kd = 2$ for $\beta =$ (1) 0.1, (2) 0.3, (3) 0.5, (4) 0.7, and (5) 0.9.

not a polar (as was expected) but a removable singularity of the 0/0 type: $R \rightarrow 1$ as $\zeta \rightarrow 0$. In this case, the degeneracy of the spatial resonance takes the form of a combination of the resonance decrease in the denominator of the quantity R and a simultaneous decrease that occurs in its numerator due to the termination of the interaction of the incident wave with the SD. The angular dependences $|R(\Theta)|$ calculated by Eq. (28) and shown in Fig. 6 demonstrate the latent (virtual) character of the “peaks” of the degenerate spatial resonance that occurs in this case (see the dashed extension of curve 5 in Fig. 6 to the region of nonphysical values of $\Theta > \Theta^{**}$). It is also important that, at the very point of the spatial resonance, the solution described by Eqs. (27) and (28) becomes meaningless. Its inadequacy is expressed as a paradoxical cancellation of the

incident wave by the refracted one: at $\Theta \rightarrow \Theta^{**}$, these two waves have the same direction of propagation (lines 1, 2 in Fig. 5a coincide) and opposite phases ($R \rightarrow -1$).

We have already noted that the point $\Theta = \Theta^{**}$ is characterized by the absence of energy transfer in the waves relative to the boundary. At this point, the selection of the refraction schemes (Fig. 5) on the basis of the radiation principle [17] is meaningless. This fact can account for the aforementioned failure of the attempt to represent the solution according to the refraction scheme shown in Fig. 5a precisely at the point $\Theta = \Theta^{**}$.

Now, let us discuss the possibility of using the scheme of refraction shown in Fig. 5b for constructing the solution at $\Theta = \Theta^{**}$. The required changes in its structure consist in moving the term describing the refracted wave from the expression for u_1 in Eqs. (27) to the expression for u_2 . The final result obtained with the selected solution at the point of the spatial resonance with the trapped wave does not depend on the presence or absence of the front boundary $\tilde{y} = d$ of the receding SD. Hence, one could directly use the solution [5] for the extreme right point of the angular interval of reflectionless birefringence. Its specific features are as follows: at $\Theta = \Theta^{**}$, the refracted and the directly transmitted waves propagate in the same direction (lines 2, 3 in Fig. 5b coincide), they have infinitely large amplitudes ($|R| \rightarrow \infty$, $|T| \rightarrow \infty$), and they are in antiphase ($R/T \rightarrow -1$). Such a solution provides a physically correct picture of the termination of interaction: the resulting field of shear displacements is mutually cancelled beyond the receding boundary ($u_2 \equiv 0$), whereas, before the boundary, the field has the form of an incident wave propagating so that it strictly follows the boundary and experiences no changes precisely because of the termination of the interaction. The typical resonance features of the solution ($|R| \rightarrow \infty$, $|T| \rightarrow \infty$) will be retained, but only for individually selected partial waves, which demonstrates their virtual character.

CONCLUSION

The results of this study show that a moving strip domain in a ferroelectric crystal can cause a translational transfer of a shear homogeneous plane wave with noncollinear orientation of the wave vectors with respect to the boundaries; this wave is identified as the antisymmetric mode of the electroacoustic wave of the

strip domain at the initial point of the spectral branch, i.e., the trapped wave. In particular, the developed approach is used to obtain a physically consistent interpretation of the resonance parametric effects of acoustodomain interaction in the conditions of extremely fast motion of domain boundaries.

ACKNOWLEDGMENTS

This work was supported by the State Program "Integratsiya" (project no. A 0066).

REFERENCES

1. G. P. Morozova and O. Yu. Serdobol'skaya, *Vestn. Mosk. Univ., Ser. 3: Fiz., Astron.* **35** (6), 42 (1994).
2. K. Nakamura and H. Shimizu, *IEEE Ultrason. Symp. Proc.* 527 (1983).
3. S. A. Zhgun, *Akust. Zh.* **33**, 645 (1987) [*Sov. Phys. Acoust.* **33**, 376 (1987)].
4. A. N. Alekseev, *Izv. Akad. Nauk, Ser. Fiz.* **57** (6), 92 (1993).
5. N. S. Shevyakhov, *Akust. Zh.* **36**, 160 (1990) [*Sov. Phys. Acoust.* **36**, 87 (1990)].
6. L. M. Lyamshev and N. S. Shevyakhov, *Akust. Zh.* **37**, 1170 (1991) [*Sov. Phys. Acoust.* **37**, 612 (1991)].
7. L. M. Lyamshev and N. S. Shevyakhov, *Pis'ma Zh. Tekh. Fiz.* **17** (17), 13 (1991) [*Sov. Tech. Phys. Lett.* **17**, 611 (1991)].
8. L. A. Vaĭnshteĭn, *Electromagnetic Waves* (Nauka, Moscow, 1988).
9. A. S. Sonin and B. A. Strukov, *An Introduction to Ferroelectricity* (Vysshaya Shkola, Moscow, 1970).
10. C. Maerfeld and P. Tournois, *Appl. Phys. Lett.* **19** (14), 117 (1971).
11. B. A. Auld, *Acoustic Fields and Waves in Solids* (Wiley, New York, 1973), Vol. 2.
12. V. N. Lyubimov and D. G. Sannikov, *Kristallografiya* **24**, 5 (1979) [*Sov. Phys. Crystallogr.* **24**, 1 (1979)].
13. Xingjiao Li, *J. Appl. Phys.* **61**, 2327 (1987).
14. W. Pauli, *The Theory of Relativity* (Dover, New York, 1981; Nauka, Moscow, 1983).
15. I. I. Gol'denblat, *Nonlinear Problems of the Theory of Elasticity* (Nauka, Moscow, 1969).
16. B. M. Bolotovskii and S. N. Stolyarov, *Usp. Fiz. Nauk* **159** (1), 155 (1989) [*Sov. Phys. Usp.* **32**, 813 (1989)].
17. B. M. Bolotovskii and S. N. Stolyarov, *Einsteinian Collection, 1974* (Nauka, Moscow, 1976), pp. 179–275.

Translated by E. Golyamina

Effect of the Loading Rate on the Mechanism of Plastic Deformation in Bismuth

O. M. Ostrikov* and S. N. Dub**

* Mozyr State Pedagogical Institute, ul. Studencheskaya 28, Mozyr, Gomel oblast, 247760 Belarus

** Institute for Superhard Materials, National Academy of Sciences of Ukraine, Avtozavodskaya ul. 2, Kiev, 04074 Ukraine
e-mail: root@pedinst.belpak.gomel.by

Received March 27, 2000; in final form, September 13, 2000

Abstract—Mechanisms of plastic deformation of bismuth single crystals that are operative upon the deformation of the (111) plane using a Berkovich indenter at a loading rate of 1–10 mN/s were studied. The plastic deformation in the range of the loading rates studied was established to occur mainly by pyramidal slip. © 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

Bismuth recommended itself as a good model material suitable for studying plastic deformation by twinning and slip [1–5]. The laws of plastic deformation revealed in this material can successfully be extended to other materials, since the perfect and partial twinning dislocations are identical in many parameters in all the crystals and often differ only in the magnitude and direction of their Burgers vectors [6]. In this work, bismuth was chosen as a model material to study plastic deformation of metals with a rhombohedral structure using the method of indenting the cleavage plane (111) with a concentrated load at small (below 10 mN/s) loading rates.

This problem is of interest since in the available literature there is virtually no information on the bismuth behavior at small loading rates, because the microhardness tests were mainly performed on a PMT-3 device with a manual loading. The loading rates studied are in the range corresponding to loading pulse durations reaching 10^{-5} s [7].

EXPERIMENTAL

Bismuth single crystals with dimensions $4 \times 5 \times 50$ mm were grown by the Bridgman technique from the initial metal of 99.999% purity. Samples $4 \times 5 \times 10$ mm in size were obtained by cleaving single crystals at a liquid-nitrogen temperature along the cleavage planes (111). The initial density of forest dislocations was determined by selective etching to be 10^5 cm^{-2} for basal dislocations and 10^3 cm^{-2} for pyramidal dislocations.

The tests were performed on a NANO INDENTER II (Nano Instrument Inc., USA) nanohardness meter using a Berkovich triangular pyramid at small (below 20 g) loads. During the tests, the displacement of the

indenter depending on the load $h = f(P)$ is detected to a high accuracy [8]. The accuracy of measuring the depth of the impression is ± 0.04 nm and that of the load at the indenter is ± 75 nN.

The tests were performed on a cleavage plane (111) of a bismuth single crystal at loads to 200 mN. The rate of loading was varied from 1 to 10 mN/s.

MODEL

The action of a concentrated load on the (111) surface of the bismuth single crystal activates slip in the system of $\{11\bar{1}\}$ planes that are arranged most favorably with respect to the action of the external load. In Fig. 1, the site of the action of the indenter is designated by the point 0. The traces from the intersection of planes $\{11\bar{1}\}$ with the plane (111) form a hexagon $A_1A_2A_3A_4A_5A_6$ (Fig. 1). Under the effect of the concentrated external load P , the $\{11\bar{1}\}$ dislocations begin their motion on the planes of a pyramid whose faces are formed by the planes $\{11\bar{1}\}$ (Fig. 2). After going a cer-

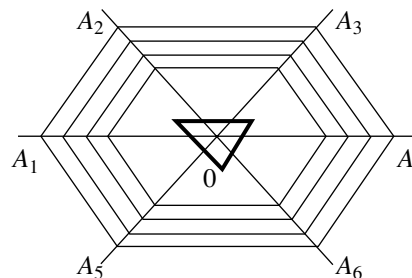


Fig. 1. Intersection of $\{11\bar{1}\}$ planes with the cleavage plane (111) (schematic). Point 0 is the site where the concentrated load was applied.

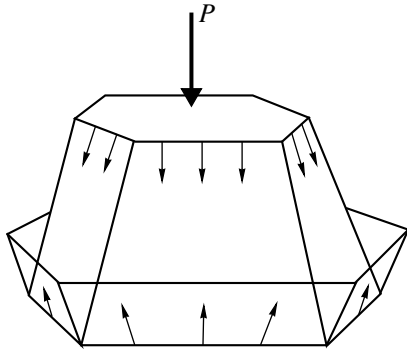


Fig. 2. Pyramidal faces (planes $\{11\bar{1}\}$) along which dislocation motion occurs.

tain distance into the depth of the crystal, the dislocations, owing to the specific features of the field of stresses produced by the Berkovich pyramid, change the direction of their motion to the opposite one, although on the planes of the same system $\{11\bar{1}\}$ inclined at a corresponding angle to the planes of the initial motion (Fig. 2). Part of these dislocation go onto the surface, forming steps on it, which can be detected metallographically. Around an indenter impression, such steps form hexagons, which are shown schematically in Fig. 1.

If we assume that under the action of the load P the dislocations began their motion on planes $\{11\bar{1}\}$ simultaneously, then their energy W_i at a certain fixed time instant t can be determined, proceeding from the concept that they form hexagons (Figs. 1, 2), as the energy of a hexagonal loop with a Burgers vector that is orthogonal to its plane. According to [9], the expression for W_i in this case has the form

$$W_i = 6L_i \frac{\mu b^2}{4\pi(1-\nu)} \left(\ln \frac{L_i}{\rho} + 0.16 \right), \quad (1)$$

where L_i is the length of the i th dislocation of the $\{11\bar{1}\}$ type, b is the modulus of the Burgers vector, ν is the Poisson ratio, and ρ is a small parameter

The total energy W_Σ of all $N/6$ dislocation loops under consideration is determined as

$$W_\Sigma = \sum_{i=1}^{N/6} W_i. \quad (2)$$

Calculated values of α depending on the loading rate V

$V, \text{ mN/s}$	α
1	0.8
3	1.0
10	1.5

For an ensemble of N dislocations $\{11\bar{1}\}$, it is suitable to use a quantity \bar{L} , which is the average length of dislocation segments of the hexagonal loop determined as

$$\bar{L} = \frac{1}{N} \sum_{i=1}^N L_i. \quad (3)$$

The energy of a hexagonal loop with a segment length \bar{L} is

$$\bar{W} = 6\bar{L} \frac{\mu b^2}{4\pi(1-\nu)} \left(\ln \frac{\bar{L}}{\rho} + 0.16 \right). \quad (4)$$

Then, in the approximation of the smallness of the interaction of hexagonal loops with one another [see Eq. (2)], we can write

$$W_\Sigma \approx \frac{N}{6} \bar{W} \quad (5)$$

or

$$\bar{W}_\Sigma \approx N\bar{L} \frac{\mu b^2}{4\pi(1-\nu)} \left(\ln \frac{\bar{L}}{\rho} + 0.16 \right). \quad (6)$$

Expression (6) can be used when the dislocation mobility is high and the processes of their generation are hindered. In the case of active dislocation generation and their small mobility, when the distance R between neighboring dislocations of different hexagonal loops closest to one another is so small that we cannot neglect dislocation interaction, the energy W'_Σ of this interaction in the approximation where each loop interacts only with two neighboring loops can be estimated by the formula [9]

$$W'_\Sigma \approx 12N\bar{L} \frac{\mu b^2}{2\pi} \ln \frac{2\bar{L}}{R}. \quad (7)$$

Estimation (7) was obtained in the approximation that all dislocations are of length \bar{L} . Summing Eqs. (6) and (7), we find the total energy W from the external forces spent on the realization of slip on the $\{11\bar{1}\}$ planes

$$W \approx N\bar{L} \frac{\mu b^2}{2\pi} \left(\frac{1+24(1-\nu)}{2(1-\nu)} \ln \left(\frac{\bar{L}}{R} \right) + 8.5 \right) \quad (8)$$

or, taking into account that $\nu \approx 0.33$,

$$W \approx N\bar{L} \frac{\mu b^2}{2\pi} \left(12.7 \ln \left(\frac{\bar{L}}{R} \right) + 8.5 \right). \quad (9)$$

Expression (9) can be simplified at $R \rightarrow 0$, since in this case $\ln(\bar{L}/R) \rightarrow \bar{L}/R$ and the second term can be

neglected. As a result, we obtain

$$W \approx 12.7N \frac{\mu b^2 \bar{L}^2}{2\pi R}. \quad (10)$$

This expression is more suitable than Eq. (9) for analyzing the physical picture of the phenomenon considered.

Korsunsky *et al.* [10] give the following relationship between the energy W and the depth h of the indenter penetration into the deformed material:

$$W = \frac{Hh^3}{3k}, \quad (11)$$

where H is the microhardness of the material and k is a geometrical parameter depending on the shape of the indenter. It follows from a comparison of Eqs. (10) and (11) that $h \sim N^{1/3}$, $h \sim \bar{L}^{2/3}$, and $h \sim R^{-1/3}$.

RESULTS AND DISCUSSION

Upon deformation of the (111) surface of bismuth single crystals with a concentrated load at large loading rates, an ensemble of wedge twins of the $\{110\}\langle 001 \rangle$ type are usually formed in the deformed region [3, 6, 11, 12]. However, no twins were revealed in this work near the indenter impression upon deformation of the cleavage surface of the bismuth single crystal by a Berkovich piramide at loading rates of 1–10 mN/s. This indicates that in the process of realization of twinning in bismuth an important role belongs to the rate of deformation, which is much higher than 10 mN/s when a PMT-3 device is used, as was the case in [3, 6, 11, 12].

All these facts indicate that the relaxation of external stresses produced in the crystal by a Berkovich indenter at loading rates of less than 10 mN/s mainly occurs by slip on $\{11\bar{1}\}$ planes. This makes it possible to use the above model for the case of deformation of single-crystal bismuth using small loading rates.

The $h(P)$ dependences at various loading rates are shown in Fig. 3, from which we see that the loading rate affects the $h(P)$ curve, although insignificantly. The higher the loading rate, the slower the growth of h with P . Using the formula [10]

$$P = \frac{Hh^2}{k}, \quad (12)$$

we plotted a theoretical dependence of P on h , assuming that $k = 0.03797$ [13] and $H = 152$ MPa. The microhardness was determined experimentally using the samples under study; it is in good agreement with the commonly accepted values (132 MPa [1]). The course of the theoretical curve coincides with the experimental curve corresponding to the loading rate of 3 mN/s. This indicates that formula (12) cannot be used for all load-

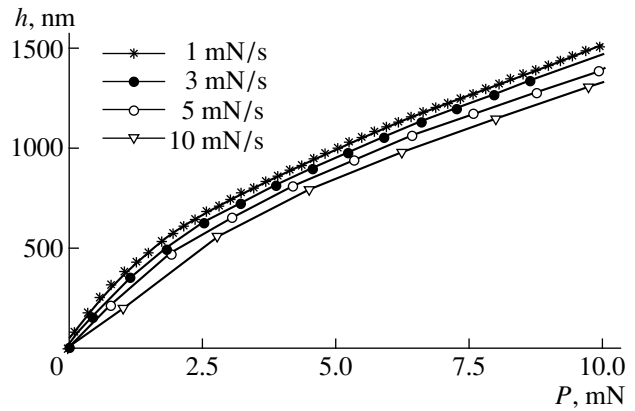


Fig. 3. Depth h of penetration of a Berkovich indenter into the material as a function of the load P .

ing rates. Therefore, it is expedient to introduce a relation

$$P = \alpha \frac{Hh^2}{k}, \quad (13)$$

where α is a certain dimensionless coefficient depending on V_p .

The calculated values of α are given in the table; it is seen that the coefficient α increases with increasing V_p . In the first approximation, the dependence of α on V_p can be written in the form

$$\alpha = \beta V_p, \quad (14)$$

where β is a quantity whose dimensionality is the inverse of that of V_p .

We can assume that α characterizes the measure of inertia of the process of the formation of the indenter impression caused by the inertia of the dislocation slip process. Similarly to Eq. (13), we can write Eq. (11) in the form

$$W = \alpha \frac{Hh^3}{3k}. \quad (15)$$

Comparing Eq. (15) with Eqs. (10) and (14), we obtain that $N \sim V_p$, $V_p \sim L^{-2}$, and $V_p \sim 1/R$, i.e., increasing the loading rate activates the processes of dislocation generation, but the dislocation length and spacing between dislocations decrease in this case. The activation of the process of dislocation generation gives grounds to believe that with the increasing loading rate, the probability of involving twinning and basal slip into the process of plastic deformation increases.

REFERENCES

1. D. Z. Grabko, Yu. S. Boyarskaya, and M. P. Dyntu, *Mechanical Properties of Semimetals of Bismuth Type* (Shtiintsa, Kishinev, 1982).

2. O. M. Ostrikov, Zh. Tekh. Fiz. **69** (5), 130 (1999) [Tech. Phys. **44**, 597 (1999)].
3. O. M. Ostrikov, Fiz. Met. Metalloved. **86** (6), 106 (1998).
4. V. I. Bashmakov and T. S. Chikova, Fiz. Met. Metalloved. **51**, 1066 (1981).
5. V. I. Bashmakov and T. S. Chikova, Inzh.-Fiz. Zh. **51** (5), 148 (1980).
6. O. M. Ostrikov, Inzh.-Fiz. Zh. **72** (5), 967 (1999).
7. V. I. Bashmakov and M. M. Brodskii, Fiz. Met. Metalloved. **35** (1), 163 (1973).
8. N. V. Novikov, S. N. Dub, Yu. V. Mil'man, *et al.*, Sverkhverd. Mater. **18** (3), 36 (1996).
9. J. P. Hirth and J. Lothe, *Theory of Dislocations* (McGraw-Hill, New York, 1967; Atomizdat, Moscow, 1972).
10. A. M. Korsunsky, M. R. McGurk, S. J. Bull, and T. F. Page, Surf. Coat. Technol. **99**, 171 (1998).
11. V. S. Savenko, V. V. Uglov, and O. M. Ostrikov, Kristallografiya **44**, 1100 (1999) [Crystallogr. Rep. **44**, 1030 (1999)].
12. O. M. Ostrikov, Fiz. Met. Metalloved. **87** (5), 78 (1999) [Phys. Met. Metallogr. **87** (5), 435 (1999)].
13. P. P. Kostin, *Physicomechanical Tests of Metals, Alloys and Nonmetallic Materials* (Mashinostroenie, Moscow, 1990).

Translated by S. Gorin

The Structure and Properties of Magnetic Inhomogeneities Arising in Nonuniform Magnetic Fields

R. M. Vakhitov and A. R. Yumaguzin

Bashkortostan State University, ul. Frunze 32, Ufa, 450074 Bashkortostan, Russia

e-mail: VakhitovRM@bsu.bashedu.ru

Received April 18, 2000

Abstract—Conditions at which magnetic inhomogeneities like 0° domain walls arise in (111) plates with combined anisotropy and the properties of these inhomogeneities are studied. In local magnetic fields, the inhomogeneities may behave as stable states. Their role in magnetization reversal of the crystals is demonstrated. © 2001 MAIK “Nauka/Interperiodica”.

It is known that different domain structures (DS) (strip domains, zigzag domains, bubble lattices, etc.) are observed in garnet ferrite crystals, which exhibit appreciable cubic anisotropy (CA) [1, 2]. As a rule, the DS configuration and parameters depend mainly on external factors (voltages, fields, etc.), which is of practical importance [1]. However, theoretical studies show that, under certain conditions, magnetic inhomogeneities like 0° domain walls (0° DW) may occur in those crystals where induced uniaxial anisotropy (IUA) and CA coexist [3]. The analysis of the stability of these inhomogeneities showed [4] that they arise as new-phase nuclei at lattice defects and play a significant part in phase transitions like spin flip.

In this work, we consider the possibility of stable 0° DW forming in nonuniform magnetic fields and investigate their properties that can help to clarify magnetization reversal in garnet ferrite films.

0° DOMAIN WALL MODEL

Let us consider a cubic ferromagnetic in the form of an infinitely long plane-parallel plate of thickness D wherein IUA is observed. We assume that the easy anisotropy axis coincides with the [111] direction and is perpendicular to the plate plane. The coordinate system is such that the axis $OZ \parallel [111]$ and the axis $OX \parallel [11\bar{2}]$. The axis OY lies in the (111) plane, making an angle φ_0 with the $[1\bar{1}0]$ direction, and coincides with the direction along which the magnet is inhomogeneous. Then, in this coordinate system, the energy of the magnetic inhomogeneities with regard for the energies of exchange interaction, combined anisotropy, and demagnetizing fields of space charges (localized on

DWs) is expressed [5] in the form (winter approximation)

$$E = \int dV \left\{ A \left[\left(\frac{\partial \Theta}{\partial y} \right)^2 + \sin^2 \Theta \left(\frac{\partial \varphi}{\partial y} \right)^2 \right] + K_u \sin^2 \Theta + K_1 \left[\frac{\sin^4 \Theta}{4} + \frac{\cos^4 \Theta}{3} + \frac{\sqrt{2}}{3} \sin^3 \Theta \cos \Theta \cos 3(\varphi - \varphi_0) \right] + 2\pi M_s^2 (\sin \Theta \sin \varphi - \sin \Theta_\infty \sin \varphi_\infty)^2 \right\}. \quad (1)$$

Here, A is an exchange factor; K_u and K_1 are the IUA and CA constants, respectively; M_s is the saturation magnetization; Θ and φ are polar and azimuth angles of the vector \mathbf{M} , respectively; Θ_∞ and φ_∞ define the direction of \mathbf{M} in the domains; and V is the plate volume. Here, we assume that the plate thickness is large, so that the contribution from the demagnetizing fields of surface charges to energy (1) can be neglected (idealized model).

The Euler equations minimizing energy (1) have the form

$$\frac{\delta E}{\delta \Theta} = 0, \quad \frac{\delta E}{\delta \varphi} = 0, \quad \frac{\delta E}{\delta \varphi_0} = 0 \quad (2)$$

if condition

$$\delta^2 E > 0 \quad (3)$$

is satisfied.

From these equations, it follows that, in the range $4/3 < \kappa < 3/2$ ($\kappa = K_1/K_u$), solutions like

$$\tan \Theta(y) = \frac{1}{a \cosh(by/\Delta_0) - c};$$

$$\begin{aligned} \varphi &= 0, \pi; \quad \varphi_0 = \pi k/3, \quad k \in Z, \\ a &= \frac{\sqrt{3\kappa-4}}{2(1-2\kappa/3)}; \quad b = \sqrt{1-2\kappa/3}; \\ c &= \frac{\pm\sqrt{2}\kappa\cos 3(\varphi-\varphi_0)}{6(1-2\kappa/3)} \end{aligned} \quad (4)$$

exist [6], along with 180° DWs with $\mathbf{M} \parallel [uvw]$. Here, $\Delta_0 = \sqrt{A/K_u}$.

Equation (4) implies that, in a (111) plate, the solitons (0° DWs) may be of two types: large-amplitude (LA) and small-amplitude (SA). They have different magnetization distributions in the inhomogeneity region [different signs of the parameter c in Eq. (4)], as well as different energies E , widths Δ_s , and amplitudes (i.e., maximum angles Θ_s of deviation of the magnetization vector \mathbf{M} from the uniform state). The width and amplitude of the solitons, which characterize their size, are given by

$$\begin{aligned} \tan \Theta_s &= 1/(a-c); \\ \Delta_s &= 2(\delta/b - \Theta(\delta)/\Theta'(\delta)), \end{aligned} \quad (5)$$

where $\delta = \ln(k + \sqrt{k^2 - 1})$ and k is the largest root of the cubic equation

$$\begin{aligned} k^3 + pk + q &= 0, \\ p &= -\left(2 + \frac{1+c^2}{a^2}\right), \quad q = 2\frac{c}{a}, \quad k > 1. \end{aligned} \quad (6)$$

For the LA solitons, we have $E/E_0 \approx 1.4$, where $E = 4\sqrt{AK_u}$ and $\Theta_s \approx 160^\circ-170^\circ$; for the SA ones, $E/E_0 \leq 10^{-2}$ and $\Theta_s \approx 10^\circ$. At $\kappa \rightarrow 3/2$, the width of the SA solitons goes to infinity, $\Delta_s \rightarrow \infty$ (the domain wall diffuses), whereas for the LA solitons, $\Delta_s \rightarrow 8.54\Delta_0$ [3]. As follows from stability condition (3), a 0° DW as a one-dimensional inhomogeneity is unstable in terms of the idealized model [7]. However, with the finiteness of the plate and the presence of lattice defects taken into consideration [4], DWs become stable and can exist near the defects in a wide range of the material parameters. This was confirmed by many experiments [1, 8, 9], from which it follows that solitary magnetic inhomogeneities do arise at defects in the absence of saturation fields. These magnetic inhomogeneities help to detect structural imperfections in of the crystals [10–12]. As a rule, the magnetization distribution near the defects is complicated [13]; however, in the one-dimensional approximation, it corresponds to a 0° DW.

It also follows from the results in [4] that the 0° DW stability is favored largely by the presence of defects, which cause a nonuniform distribution of the magnetic parameters in the sample. As such a defect, we consider a nonuniform magnetic field \mathbf{H} lying in the $(\bar{1}10)$ plane

and making an angle Ψ with the [111] direction. Let \mathbf{H} depend on the coordinate as

$$H(y) = \begin{cases} H, & |y| \leq \frac{L}{2} \\ 0, & |y| > \frac{L}{2}, \end{cases} \quad (7)$$

where L is the size of the part of the plate that is subjected to local field (7), i.e., the size of the field nonuniformity.

The energy of interaction between the defect and a 0° DW is given by the expression

$$\begin{aligned} E_d &= -\int_V (\mathbf{M} \cdot \mathbf{H}) dV \\ &= -M_s H L_x D \int_{-L/2}^{L/2} \cos(\Theta(y) - \Psi) dy. \end{aligned} \quad (8)$$

It is seen that the 0° DW–field interaction makes a negative contribution to (1). Its absolute value is maximum for $\Theta(y) = \Psi$. Thus, interaction (8) makes the 0° DW shape follow the profile of the field nonuniformity.

The inclusion of the plate finiteness requires that the contribution from the demagnetizing fields of nonuniformly distributed surface charges to energy (1),

$$\begin{aligned} E_{ms} &= M_s^2 L_x \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\cos \Theta(y) \cos \Theta(y') - 1) \\ &\quad \times \ln\left(1 + \frac{D^2}{(y-y')^2}\right) dy dy', \end{aligned} \quad (9)$$

be taken into account (L_x is the plate size in the OX direction). Space charges on Bloch DWs make no contribution.

It was shown [4] that the magnetostatic factor substantially narrows the stability range of 0° DWs; however, it should be taken into account in this case.

To quantitatively describe the effect of a nonuniform magnetic field on the structure and stability of 0° DWs, we will apply the variational method. As a test function, we consider the magnetization variation in the transition layer [Eq. (4), where a , b , and c are considered as variational parameters]. Obviously, these parameters can be determined by minimizing energy (1) in view of (8) and (9), i.e., by minimizing the energy

$$E_s = E + E_d + E_{ms}. \quad (10)$$

In this case, the variational approach is the most reasonable, since the factors considered have an insignificant effect on the structure of solitary magnetic inhomogeneities, only changing the structure parameters [4].

STATIC PROPERTIES OF 0° DWs

To find stable states of 0° DWs in nonuniform magnetic field (7), the corresponding variational problem was solved by numerically minimizing the reduced energy $\epsilon_s = E_s/M_s^2 L_x D \Delta_0$. In this model, we took into account that the centers of the wall and the field may be spaced at a value of ξ .

The results obtained are shown in Figs. 1–5, where all the values having the dimension of length are reduced to Δ_0 . We can conclude that the walls are stable in certain ranges of the sample parameters and nonuniform field. As follows from Fig. 1, the equilibrium position of the wall is defined by the center of the defect. When this position is shifted relative to the center, quasi-elastic forces for which the Hooke law is valid at small ξ arise. This means that, first, harmonic vibration of the wall about the equilibrium position is possible and, second, that the interaction forces in the region of wall localization on the defect stabilize the structure of the wall and to a great extent specify its parameters. In fact, from Fig. 2, the 0° DW width is proportional to that of the defect. As the latter increases, so does the width of the wall, and Θ_s asymptotically approaches the limit that coincides with the polar angle of the vector \mathbf{M} in a uniformly magnetized plate [14]. Obviously, such a situation will occur for $L \rightarrow \infty$. In this case, the magnetic field \mathbf{H} becomes uniform and the equilibrium orientation of the vector \mathbf{M} is defined by the combined effect of IUA, CA, and magnetic field. On the other hand, Fig. 2 shows that there is the minimum size of the nonuniform field region, at which 0° DWs become collapse-unstable; hence, the origination of 0° DWs at the center of the nonuniform field is of a threshold nature. This is in agreement with dependences of the DW parameters on h and Ψ (Figs. 3, 4). Moreover, we can argue that there exists the least energy of interaction of the 0° DW with the local field necessary for the formation of the stable wall structure. This energy depends on the size of the defect and on its other parameters (Figs. 3, 4). From the dependences obtained, it also follows that the 0° DW shape tends to follow the profile of the defect. This tendency becomes dominant with increasing L ; i.e., our previous contention is confirmed.

The dependence of the 0° DW stable states on κ (Fig. 5) has an interesting feature. It is seen that the walls remain stable in the wide range of κ , the stability range exceeding that predicted by the idealized model. Moreover, the wall exists even for $\kappa = 0$, i.e., when CA is absent. This is a nontrivial result, since it is combined anisotropy in the crystal that causes solutions like 0° DWs to appear in the idealized model.

On the other hand, as is seen in Fig. 5, the stability region is bounded both above and below: at the upper bound, the DW diffuses ($\Delta_s \rightarrow \infty, \Theta_s \rightarrow \Theta_m$), while at the lower bound, it collapses. In both cases, the plate becomes uniformly magnetized, but, in the first

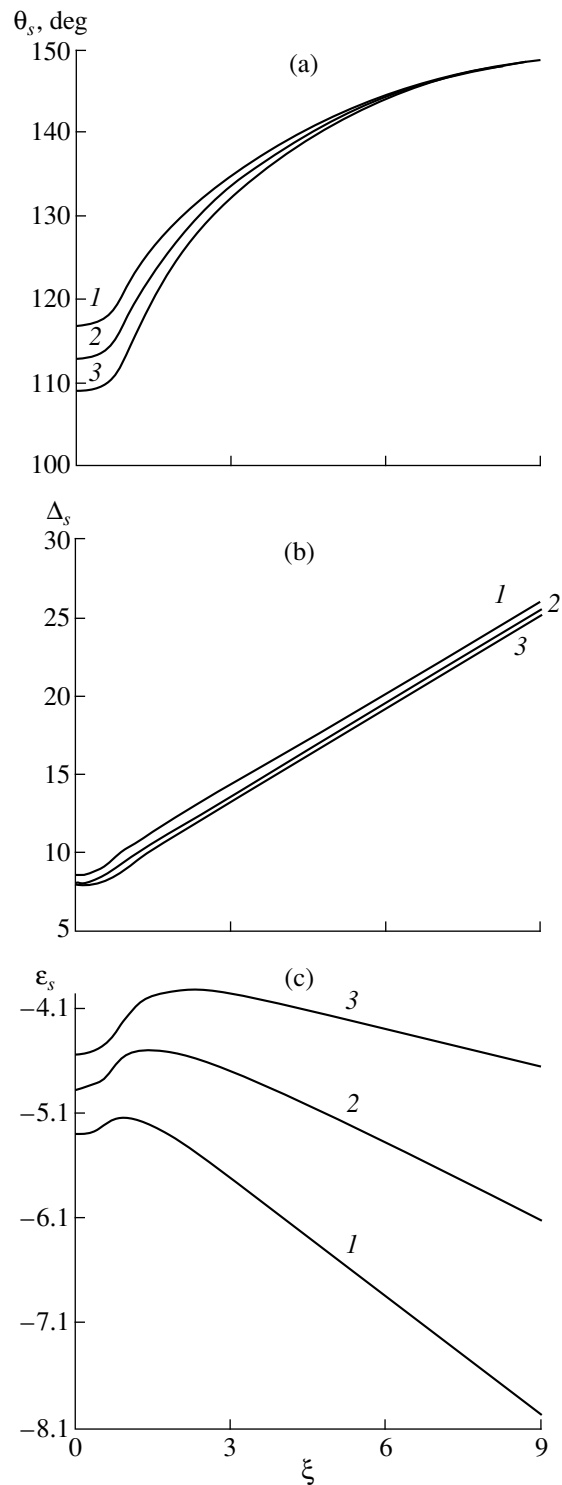


Fig. 1. 0° DW parameters vs. displacement ξ : (a) Θ_s , (b) Δ_s , and (c) ϵ_s . $\kappa = 1.42, D = 35, h = 3, \Psi = 90^\circ$, and $L = 5$ ($h = HM_s/K_u$). $Q = (1) 5, (2) 8, \text{ and } (3) 15$.

case, its state will correspond to the canted phase with $\mathbf{M} \parallel [uuw]$ and, in the second case, to the highly symmetrical phase with $\mathbf{M} \parallel [111]$. When the other parameters of the sample or defect vary (Figs. 3, 4), the wall

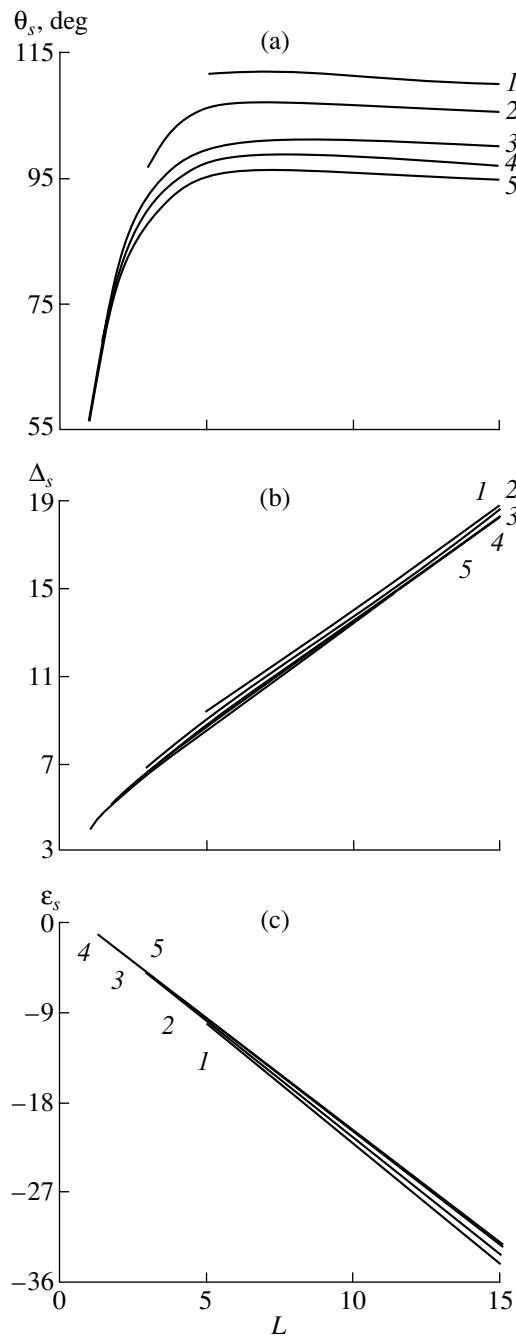


Fig. 2. The same as in Fig. 1 vs. defect width L . $Q = 3.5$, $D = 35$, $h = 5$, and $\Psi = 90^\circ$. $\kappa = (1) 1.42$, (2) 0.83, (3) 0.25, (4) 0.5, and (5) 0.25.

behaves in a similar way. This is an additional reason for the fact [4] that the occurrence of 0° DWs in the crystal is the intermediate stage in the phase transition of the magnet from one homogeneous state to another and that 0° DWs can be considered as new-phase nuclei localized at the defect, i.e., in the local-field region.

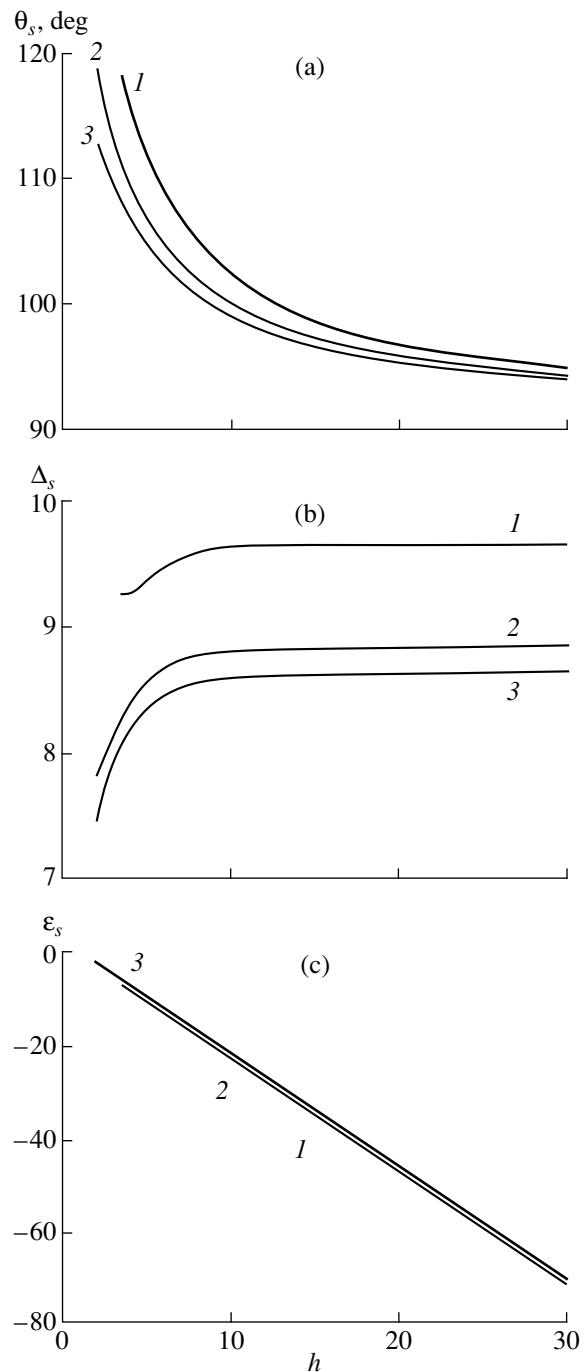


Fig. 3. The same as in Fig. 1 vs. h . $\kappa = 1.42$, $D = 35$, $\Psi = 90^\circ$, and $L = 5$. $Q = (1) 3.5$, (2) 8, and (3) 15.

DISCUSSION

From the aforesaid, it follows that, in (111) plates with combined anisotropy, local magnetic fields can give rise to inhomogeneities of specific configuration, i.e., 0° DWs. They have a number of intriguing properties; in particular, their size correlates with that of the field nonuniformity (Fig. 2). Our results may help to

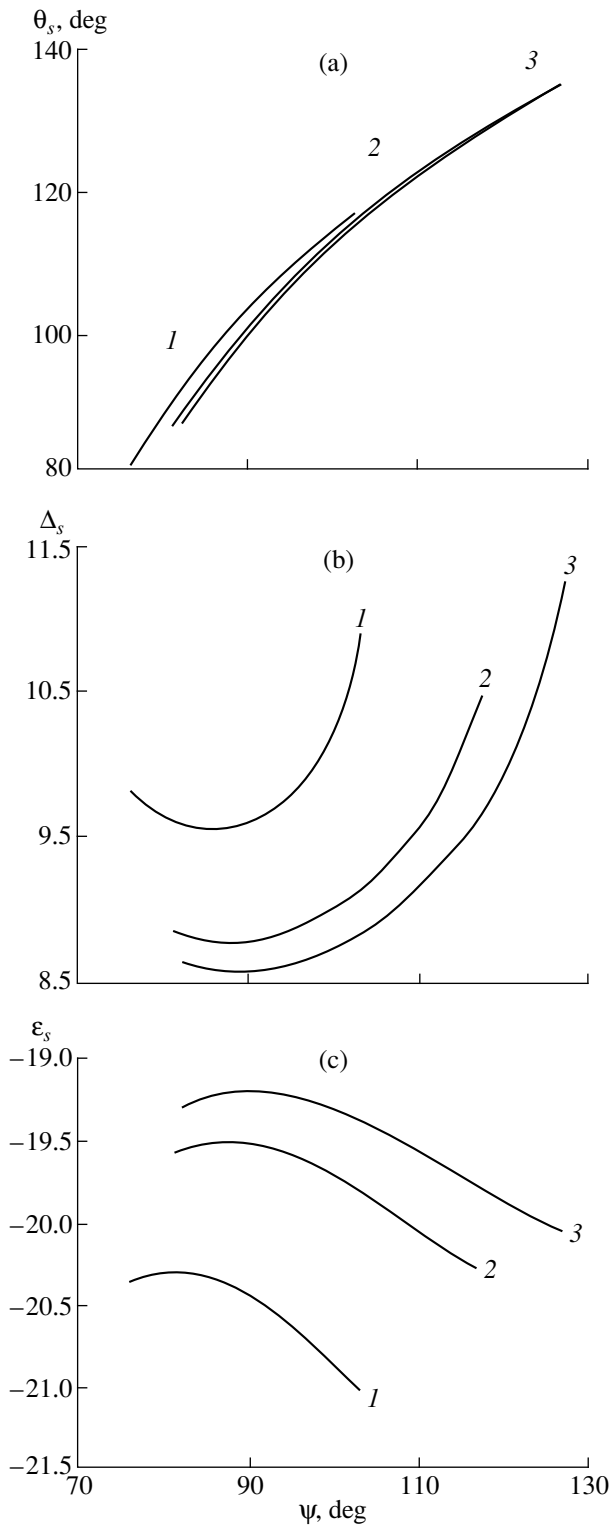


Fig. 4. The same as in Fig. 1 vs. angle Ψ for $h = 9$. The other material parameters are the same as in Fig. 3.

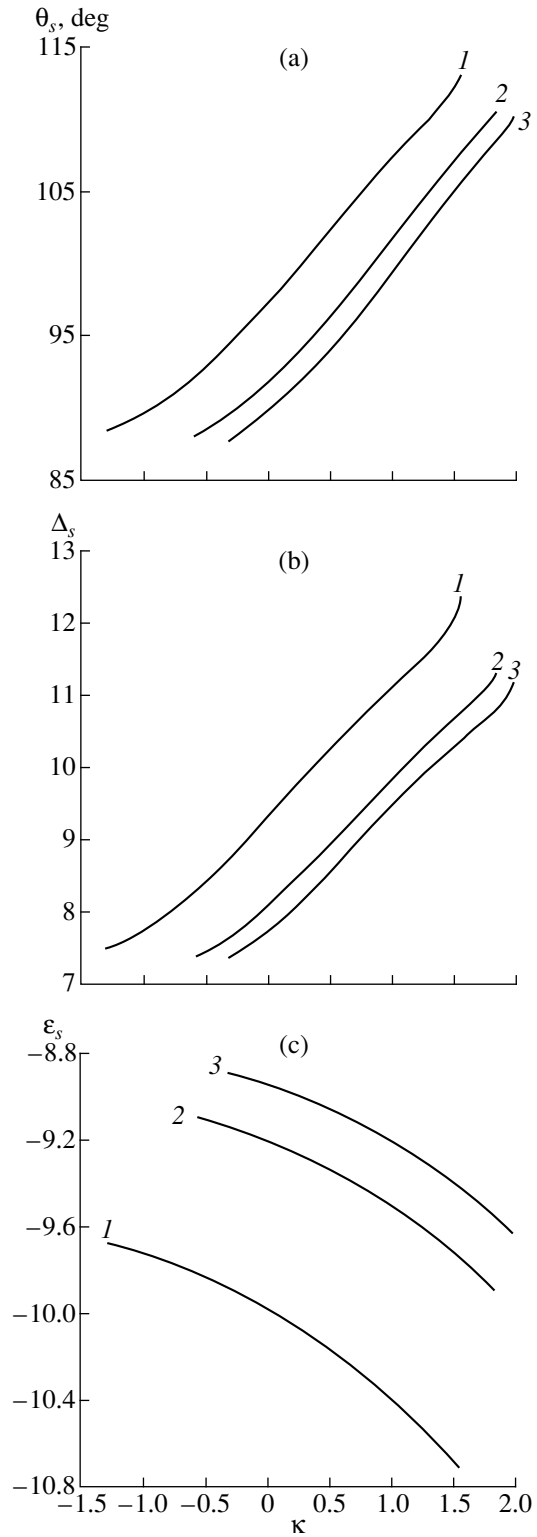


Fig. 5. The same as in Fig. 1 vs. the κ for $h = 5$. The other parameters are the same as in Fig. 3.

elucidate the magnetization reversal effect in (111)-oriented garnet ferrite films in pulsed fields [1, 15]. In the steady-state mode (at the end of the pulse, when magnetization reversal becomes), stationary stable triangular domains were observed at the center of the axisymmetric nonuniform field. A specific feature of these domains is that their size is comparable to that of the field nonuniformity and is much larger than that of bub-

lar domains were observed at the center of the axisymmetric nonuniform field. A specific feature of these domains is that their size is comparable to that of the field nonuniformity and is much larger than that of bub-

bles, which also have triangular shape [15–17]. Such a shape is explained by the symmetry of the CA easy axes [16, 17]. We considered the one-dimensional case of a solitary domain. However, our results can readily be generalized to the two-dimensional case. This is due to the fact that, first, the shape of the domains will also be triangular because of the symmetry and, second, their sizes will correlate with those of the field because the shape of the 0° DWs tends to follow the profile of the defect.

CONCLUSION

Thus, we have shown that stable 0° DWs can arise in local fields. Their stability domain (in terms of the sample and field parameters) is limited by two critical (diffusion and collapse) fields. Based on our data, 0° DWs can be thought of as nuclei of a new phase that arise during phase transitions like spin flip. In addition, results of this work may help in the interpretation of other experiments wherein nonuniform magnetic fields and different defects are the subject of investigation and can be used in applications.

REFERENCES

1. V. V. Randoshkin and A. Ya. Chervonenkis, *Applied Magneto-optics* (Énergoatomizdat, Moscow, 1990).
2. G. S. Kandaurova and L. A. Pamyatnykh, *Fiz. Tverd. Tela (Leningrad)* **31** (8), 132 (1989) [*Sov. Phys. Solid State* **31**, 1351 (1989)].
3. R. M. Sabitov and R. M. Vakhitov, *Izv. Vyssh. Uchebn. Zaved., Fiz.* **31**, 51 (1988).
4. R. M. Vakhitov and A. R. Yumaguzin, in *Proceedings of the Moscow International Symposium on Magnetism, Moscow, 1999*, Part II, p. 53.
5. A. Hubert, *Theorie der Domanenwände in Geordneten Medien* (Springer-Verlag, Berlin, 1974; Mir, Moscow, 1977), p. 306.
6. R. M. Vakhitov, R. M. Sabitov, and Z. V. Gabbasova, *Phys. Status Solidi B* **168**, K87 (1991).
7. W. F. Brown, *Micromagnetics* (Interscience, New York, 1963; Nauka, Moscow, 1979).
8. V. K. Vlasko-Vlasov and M. V. Indenbom, *Zh. Éksp. Teor. Fiz.* **86**, 1084 (1984) [*Sov. Phys. JETP* **59**, 633 (1984)].
9. Bruce E. Bernacki, Te-ho Wu, and M. J. Mansuripur, *J. Appl. Phys.* **73**, 6838 (1993).
10. A. F. Reïderman and Yu. G. Gobov, *Defektoskopiya*, No. 3, 70 (1989).
11. V. V. Randoshkin, A. F. Martynov, and L. N. Mikhaïlov, *Élektron. Tekh., Ser. Mater.*, No. 5, 77 (1983).
12. V. L. Dorman, A. V. Kovalev, I. V. Nikonets, *et al.*, *Mikroélektronika* **17** (2), 133 (1988).
13. A. B. Dichenko and V. V. Nicolaev, *J. Magn. Magn. Mater.* **53**, 71 (1985).
14. V. V. Grinevich and R. M. Vakhitov, *Fiz. Tverd. Tela (Leningrad)* **32**, 3409 (1995) [*Sov. Phys. Solid State* **32**, 1973 (1995)].
15. L. P. Ivanov, A. S. Logginov, A. T. Marchenko, and G. A. Nepokoïchitskiï, *Zh. Tekh. Fiz.* **52**, 1246 (1982) [*Sov. Phys. Tech. Phys.* **27**, 761 (1982)].
16. R. M. Sabitov and R. M. Vakhitov, *Fiz. Tverd. Tela (Leningrad)* **22**, 2523 (1980) [*Sov. Phys. Solid State* **22**, 1474 (1980)].
17. W. A. Bonner, R. C. LeCraw, R. D. Pierce, and L. G. Uitert, *J. Appl. Phys.* **49**, 1874 (1978).

Translated by M. Astrov

Crystal Chemical Simulation of Low-Permittivity Ferroelectrics

L. A. Reznichenko, E. M. Kuznetsova, O. N. Razumovskaya, and L. A. Shilkina

Research Institute of Physics, Rostov State University,
pr. Stachki 194, Rostov-on-Don, 344090 Russia

e-mail: klevtsov@iphys.rnd.runnet.ru

Received June 21, 2000

Abstract—It is shown that the uniform strain parameter of the unit cell of solid solutions of alkali metal niobates can be raised and their permittivity can drastically be decreased by isovalent superstoichiometric modification in the A subcell. The underlying mechanism is suggested. It is concluded that these findings are helpful in studying (simulating) piezoelectric ferroelectrics of practical value. © 2001 MAIK “Nauka/Interperiodica”.

Low-permittivity (ϵ) piezoelectric ferroelectrics (PFs) are promising in many areas of technology; in particular, they are candidate materials for high-frequency acoustoelectric transducers [1, 2]. Upon simulating PFs based on perovskite-like oxides (POs), it is necessary to relate ϵ with crystal chemical characteristics of PF cations and with crystal structure parameters. In insulators, $\epsilon \sim P/E$, where P is the polarization or the dipole moment per unit volume and E is the strength of an electric field applied to the crystal. Since $P = mn$ (m is the dipole moment per ion, and n is the volumetric density of ions), where $m \sim \alpha$ (α is the polarizability) and $n \sim 1/r^3$ (r is the ion radius), ϵ depends on the crystal chemical parameters of an ion as $\epsilon \sim \alpha/r^3$. Clearly, ions with small values of α/r^3 are best suited for simulating low- ϵ PFs. Such are ions whose electronic configuration is similar to that of noble gases [3]; the smallest values of α/r^3 are typical of ions of alkali metals (AMs): Li^+ (0.166), Na^+ (0.210), K^+ (0.382), as well as of Mg^{2+} (0.240) and Al^{3+} (0.360). Only for them, the ratio α/r^3 is much smaller than unity; for others, it is close to unity. Therefore, among known PFs, alkali metal niobates [4], including magnesium- and aluminum-substituted ones [5], have the least permittivities. Their permittivity can be decreased further by improving the crystal structure. In [6], a relationship between ϵ and the uniform strain coefficient δ , the basic parameter of the PO structure that characterizes the distortion of the cubic perovskite cell [7], was derived. It was established that $\epsilon \sim 1/\delta$. Thus, the larger δ , the smaller ϵ . The value of δ can be increased, in particular, by lowering the defect density in the solid solutions. In this case, the process of breaking “polarization circuits” by vacancies [8] will be suppressed, causing the spontaneous polarization P_s and, hence, δ to grow owing to the relationship $\delta \sim P_s^2$ [7].

The value of δ in niobate PFs can be raised by isovalent (in the A subcell) superstoichiometric modification. The underlying mechanism is as follows. AM niobates, like other Nb-containing complex oxides,

“inherit” the defect structure of the Nb_2O_5 high-temperature modification, which is used in material synthesis [9]. In this case, the anion-deficient nonstoichiometry of Nb_2O_5 , i.e., $\text{Nb}_2\text{O}_5: (\text{Nb}_{1-x}^{5+}\text{Nb}_x^{4+})_2\text{O}_{5-x}$, due to the ease of change of the oxidation degree ($\text{Nb}^{5+} \longleftrightarrow \text{Nb}^{4+}$), may be accompanied by a change in the point defect density because of crystallographic shear [10]. As a result, the number of both anion and A sites may decrease as compared with the initial structure [11]. Excessive A cations occupy newly formed (by crystallographic shear) defects: tetrahedral and octahedral voids appearing where regular blocks join together. In Nb-based compounds, the joints are due to crystallographic shears in two almost mutually orthogonal directions [10]. The vacancies (voids) do not all become occupied, since their concentration usually exceeds the concentration of excessive A cations. Clearly, the remaining vacancies serve to decrease δ and, accordingly, increase ϵ . The introduction of additional AM oxides would favor at least the partial occupation of the vacant sites in order to rise δ and lower ϵ .

This was done in Na-substituted LiNbO_3 solid solutions ($\text{Na}_{1-x}\text{Li}_x\text{NbO}_3$), which are of great practical interest [1, 5], and Nb-deficient solid solutions of formula $\text{Na}_{1-x}\text{Li}_x\text{Nb}_{1-y}\text{O}_3$. We studied solid solutions with $x = 0.08, 0.10, 0.12$, and 0.14 ; y was varied from 0.01 to 0.04. The modifiers (Na_2CO_3) were introduced superstoichiometrically in amounts of 2.0–8.0 at. % ($Z = 0.02$ – 0.08) at the stage of component mixing. The Na_2CO_3 and Li_2CO_3 components were of chemical grade, and Nb_2O_5 was of optical grade. The material was prepared by two-stage solid-phase synthesis: 800°C for 5 h and 850°C for 5 h. Subsequent sintering was performed by hot pressing [7] at a pressure of 200 kg/cm^2 with isothermal heat treatment (1000 – 1150°C depending on the composition) for 40 min. The solid solutions prepared were impurity-free, as indicated by X-ray diffraction analysis.

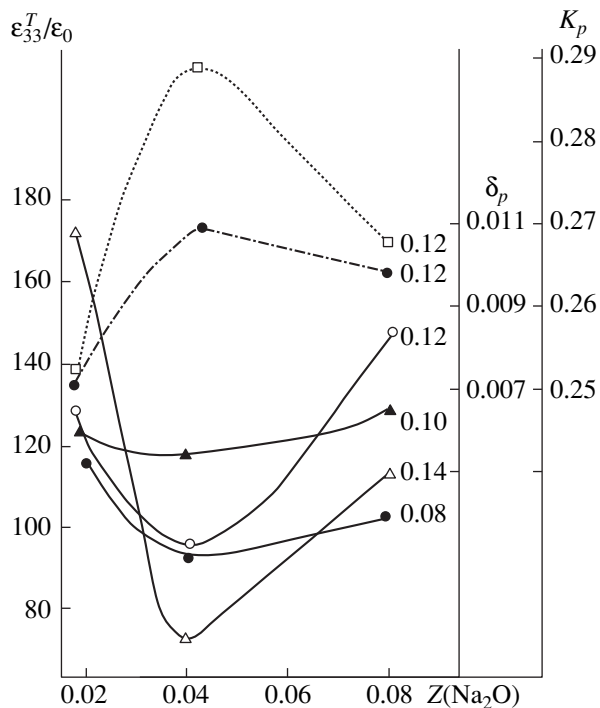


Fig. 1. $\epsilon_{33}^T/\epsilon_0$, K_p , and δ_p vs. $\text{Na}_{1-x}\text{Li}_x\text{NbO}_3 + Z\text{Na}_2\text{O}$ (Li_2O) solid solution composition. The figures at the curves indicate the value of x (the content of LiNbO_3). Continuous curves, $\epsilon_{33}^T/\epsilon_0$; dashed curves, K_p ; and dash-and-dot curves, δ_p .

Experimental results are shown in Fig. 1. The expected effect of increasing δ and decreasing ϵ (in more exact terms, δ_p , the uniform strain parameter for the rhombically distorted cell, and $\epsilon_{33}^T/\epsilon_0$, the relative permittivity after polarization) is observed in the solid solutions with y and Z not exceeding 0.02 and 0.04, respectively. The piezoelectric parameters (K_p , electro-mechanical coupling coefficient; d_{31} , piezoelectric modulus; and g_{31} , piezoelectric sensitivity) are considerably improved. As Z and y go beyond the above limits, ϵ grows (sometimes sharply) and K_p declines. This may be associated with the fact that when the concentration of the modifiers reaches some value, the concentration of excessive A cations begins to exceed the concentration of vacancies. As a result, the cations occupy interstices, introducing disorder (defects) into the structure.

Similar results of Japanese investigators (Fig. 2 in [12–14]), which studied substituted AM niobates of composition $\text{Na}_{1-x}\text{Li}_x\text{Nb}_{1-y}\text{O}_3 + Z\text{A}_2\text{O}$ ($A = \text{Na}, \text{Li}$), where $x = 0.01\text{--}0.3$, $y = 0.005\text{--}0.1$, and $Z = 0.001\text{--}0.1$, support our assumption.

Isovalent modification at small Z significantly activates the process of charge compacting during sintering: the rate and degree (β) of shrinkage increase (Fig. 3). Moreover, even the kinetics of sintering changes. If for pure AM niobates, the process usually

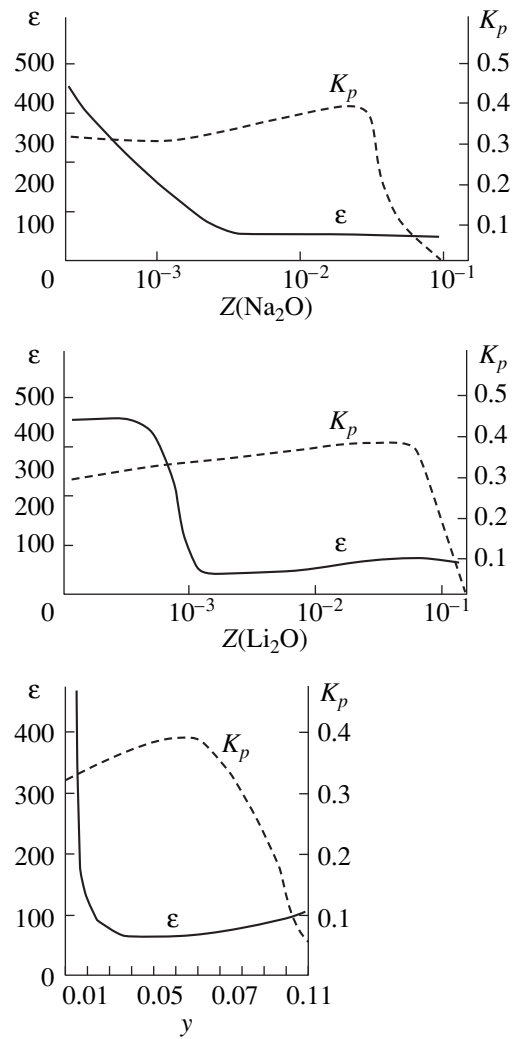


Fig. 2. ϵ and K_p vs. $\text{Na}_{1-x}\text{Li}_x\text{Nb}_{1-y}\text{O}_3 + Z\text{A}_2\text{O}$ ($A = \text{Na}, \text{Li}$) solid solution composition [12–14].

goes in the solid phase and involves a small amount of the liquid phase [15], at $Z = 0.04\text{--}0.08$, the liquid phase prevails and the ceramic formation temperature T_1 drops sharply (Fig. 3). The reason is that the solid-phase reaction is typically incomplete and the powders contain unreacted sodium carbonates and hydroxides, which form low-melting eutectics. The Li_2CO_3 and LiOH compounds are unstable, and Li_2O is a high-melting compound; therefore, lithium, as a rule, does not form the liquid phase and the kinetics does not change. If the liquid phase does form, its role in improving the solid solution structure is to occupy vacancies, to “refine” and bond grains, and eventually to increase δ .

It is noteworthy that at certain concentrations of additional Na, the solid solutions with the orthorhombic structure become two-phase, which means that they shift in the $(\text{NaLi})\text{NbO}_3$ state diagram [16] because of composition changes. This also may point to variations

Electrophysical parameters of the solid solutions

Composition			$T_k, ^\circ\text{C}$	$\epsilon_{33}^T/\epsilon_0$	K_p	$d_{31} \times 10^{12},$ C/N	$g_{31} \times 10^{-3},$ W/N	Q_m	$V_R \times 10^{-3},$ m/s	Fabrication technology
x	y	Z								
0.08	0	0.04	332	95	0.207	10	11.7	614	5.88	Laboratory
0.1	0	0.04	345	119	0.25	15	14	145	5.92	"
0.12	0	0.04	350	95	0.29	13	15.7	30	5.65	"
0.12	0	0.08	334	150	0.27	14	11.4	177	5.89	"
0.14	0	0.04	346	75	0.21	11	17.1	45	5.75	"
0.12	0.01	0	350	102	0.275	15	16.8	14	5.64	"
0.12	0.02	0	359	101	0.281	16	16.9	97	5.87	"
0.12	0.06	0	334	141	0.246	14	14	217	5.96	"
0.14	0	0.04	356	102	0.32	16	17.7	271	5.3	"
0.14	0	0.04	356	105	0.294	15	15.9	253	5.46	Industrial (variable hot-pressing conditions)
0.14	0	0.04	356	104	0.304	15	16.5	239	5.37	
0.14	0	0.04	362	108	0.281	14	15.1	252	5.42	

of the electrophysical parameters especially with regard for the behavior of the solid solutions in dependence of the stoichiometric ratio [17]. The table lists the dielectric and piezoelectric characteristics of some of the solid solutions obtained under both laboratory (hot-pressed samples measuring 10×1 mm) and industrial conditions (large blocks $110 \times 110 \times 25$ mm in size, which were cut into samples of appropriate dimensions). The solid solutions are seen to offer ultralow

$\epsilon_{33}^T/\epsilon_0$; relatively high T_k , K_p , d_{31} , and g_{31} ; a wide range of the mechanical figure of merit Q_m ; increased velocity of sound V_R ; as well as high reproducibility of the parameter from block to block within a lot and from lot to lot. Thus, the solid solutions are promising PF materials for various applications.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research (grant no. 99-02-17575).

REFERENCES

1. E. G. Fesenko, A. Ya. Dantsiger, L. A. Reznichenko, *et al.*, *Zh. Tekh. Fiz.* **52**, 2262 (1982) [*Sov. Phys. Tech. Phys.* **27**, 1389 (1982)].
2. A. Ya. Dantsiger, O. N. Razumovskaya, L. A. Reznichenko, *et al.*, *High-Efficiency Piezoceramic Materials: A Handbook* (Kniga, Rostov-on-Don, 1994).
3. G. I. Skanavi, *Physics of Dielectrics (Weak Field Region)* (Fizmatgiz, Moscow, 1958).
4. A. Ya. Dantsiger, O. N. Razumovskaya, L. A. Reznichenko, and S. I. Dudkina, *High-Efficiency Piezoceramic Materials: Optimization of Search* (Paik, Rostov-on-Don, 1995).
5. L. A. Reznichenko, Author's Abstract of Candidate's Dissertation (Rostov-on-Don, 1980).
6. E. G. Fesenko, A. Ya. Dantsiger, and O. N. Razumovskaya, *Novel Piezoelectric Ceramic Materials* (Rostov Gos. Univ., Rostov-on-Don, 1983).
7. E. G. Fesenko, *Perovskite Family and Ferroelectricity* (Atomizdat, Moscow, 1972).
8. H. Thomman, *Z. Angew. Phys.* **20**, 554 (1966).
9. L. A. Reznichenko, O. N. Razumovskaya, and L. A. Shilkinina, in *Proceedings of the International Scientific and*

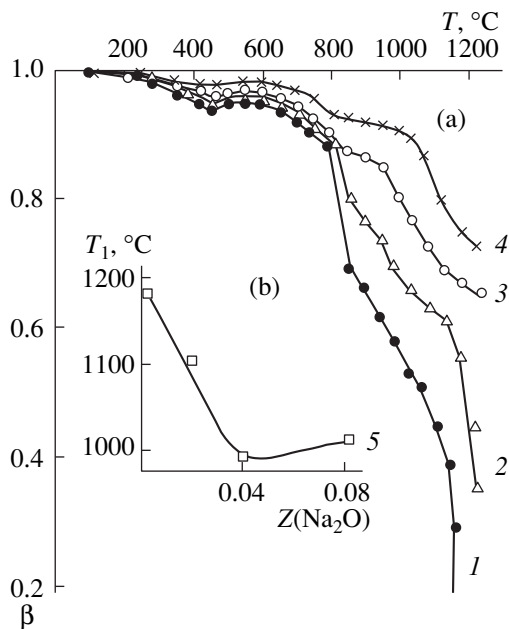


Fig. 3. (a) $\beta(T)$ and (b) $T_1(Z)$ for $\text{Na}_{1-x}\text{Li}_x\text{NbO}_3 + \text{ZA}_2\text{O}$ ($A = \text{Na, Li}$). (1) $x = 0.1, Z = 0.04$ (Na_2O); (2) $x = 0.12, Z = 0.04$ (Na_2O); (3) $x = 0.1, Z = 0$; (4) $x = 0.1, Z = 0.04$ (Li_2O); and (5) $x = 0.1$.

- Practical Conference "Piezotechnique-97," Obninsk, 1997, p. 191.*
10. C. N. R. Rao and J. Gopalakrishnan, *New Directions in Solid State Chemistry* (Cambridge Univ. Press, Cambridge, 1986; Nauka, Novosibirsk, 1990).
 11. A. G. Petrenko and V. V. Prisedskii, *Structural Defects in Ferroelectrics* (Kiev, 1989).
 12. Jpn. Patent No. 45-22269 on cl. 62C23 (1970).
 13. Jpn. Patent No. 45-22270 on cl. 62C23 (1970).
 14. Jpn. Patent No. 45-30150 on cl. 62C23 (1970).
 15. L. A. Reznichenko, O. N. Razumovskaya, L. A. Shilkina, and V. A. Aleshin, in *Proceedings of the 7th International Workshop on Physics of Ferroelectrics-Semiconductors* (Kniga, Rostov-on-Don, 1996), Vol. 6, p. 112.
 16. L. A. Shilkina, L. A. Reznichenko, M. F. Kupriyanov, and E. G. Fesenko, *Zh. Tekh. Fiz.* **47**, 2173 (1977) [*Sov. Phys. Tech. Phys.* **22**, 1262 (1977)].
 17. L. A. Reznichenko and L. A. Shilkina, *Izv. Akad. Nauk SSSR, Ser. Fiz.* **39**, 1118 (1975).
 18. *OST (Standard) 110444-87: Piezoceramic Materials. Technical Conditions*, 1988, Group É10.

Translated by V. Isaakyan

Reconstruction of the Autowave Structure upon Deformation of Polycrystalline Aluminum

L. B. Zuev, B. S. Semukhin, and N. V. Zarikovskaya

*Institute of Strength Physics and Materials Science, Siberian Division, Russian Academy of Sciences,
Akademicheskii pr. 2/1, Tomsk, 634021 Russia*

Received June 26, 2000

Abstract—The evolution of zones of localized plastic deformation in polycrystalline aluminum was investigated. At the stage of the linear strain hardening, such zones were established to move synchronously, whereas at the stage of parabolic strengthening they are stationary. The quantitative characteristics (wavelength, propagation velocity) of deformation waves that are formed at the stage of linear strengthening were determined. A relation between the quantitative characteristics of the process of deformation localization and the grain size was found. The distribution of local deformations upon transition from one stage of plastic flow to another was investigated. A model that explains the generation of coarse-scale structures of localized plastic deformation is suggested. © 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

The investigations of the kinetics of the development of deformation localization that were performed on a wide circle of materials using a high-precision technique [1–4] show that the process of plastic flow tends to localization during its entire course and that the forms of localization have a clearly pronounced relation to the multistage nature of the flow curve. The most interesting problem in this connection is the synchronous motion of plastic-flow zones observed at the stage of linear strengthening, at which $\sigma \sim \Theta \epsilon$, where σ is the stress in the sample, ϵ is the related total deformation, and $\Theta = d\sigma/d\epsilon$ is the strain-hardening coefficient. The detailed picture of flow localization in this case makes it possible to obtain the characteristics of the specific wave process, i.e., the wavelength λ and the velocity of its propagation V_w (which is inversely proportional to Θ) [5]. At the stage of parabolic strengthening, when $\sigma \sim \epsilon^m$ ($m < 1$) and the strain-hardening coefficient decreases with increasing deformation, the zones of plastic flow are stationary. To understand the origin of these differences, of significant interest is the evolution of the process of deformation localization in the material, which can exhibit a repeated change of the parabolic and linear stages of strengthening. It proved that polycrystalline aluminum is unique in this respect; its $\sigma(\epsilon)$ curve consists of three stages [6]: two parabolic stages (at the beginning and at the end of the process) and a linear stage between them. The aim of this paper is to clarify the laws of the development of the process of deformation localization at each stage and upon transition from one stage to another.

EXPERIMENTAL

As the material for the investigation, we used sheet aluminum of grade A85 (total content of impurities no more than 0.15 wt %). Samples with the dimensions of the gage portion of $50 \times 10 \times 2$ mm were stamped from a sheet. The average grain size D was changed by preliminary plastic deformation and subsequent recrystallization annealing; in three cases under study, it was 0.2, 0.3, and 0.4 mm. The samples were pulled in tension in an Instron-1185 testing machine at a constant velocity of the mobile grip $V_m = 3.35 \times 10^{-6}$ m/s ($\dot{\epsilon}_t = 6.7 \times 10^{-5}$ s $^{-1}$).

The accuracy of distinguishing the linear stage of strain hardening was enhanced by using a special technique, in which, in addition to the usual procedure of finding the deformation interval in which the coefficient Θ was constant, the ultrasonic velocity V_s was measured in the sample directly during the mechanical tests [7]. As is shown in [7], both Θ and V_s are constant at the stage of linear strengthening in aluminum; the use of an additional criterion simplifies the separation of this stage and increases the accuracy of determining its beginning and end in the $\sigma(\epsilon)$ flow curves for the samples of polycrystalline aluminum used.

The patterns of deformation localization were revealed directly in the process of loading using the speckle-interferometry method described in detail in [4]. This method makes it possible to obtain the distributions of the components of the plastic-distortion tensor $\beta_{i,j} = \nabla r$ (where r is the displacement vector) over the surface of observation (see, e.g., Fig. 1) and follow their evolution in time. Thus, analyzing changes in the position of zones of localized plastic flow depending on time $t = \epsilon/\dot{\epsilon}_t$, we can determine important characteristics of the process, such as the velocity V of propaga-

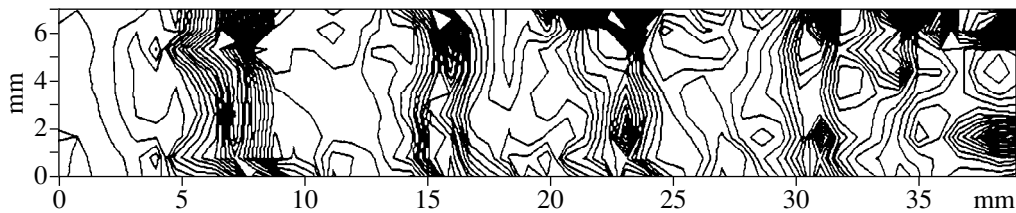


Fig. 1. Distribution of local elongations in a sample with grain size $D \approx 0.2$ mm (the contours correspond to lines of $\varepsilon = \text{const}$).

tion of deformation waves along the sample axis in the course of its extension and the wavelength λ .

RESULTS

We managed to separate the stage of linear strengthening in the curves of plastic flow of all the samples of

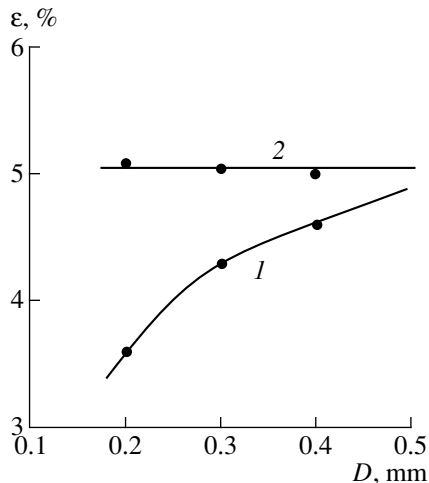


Fig. 2. The beginning (1) and the end (2) of the stage of linear strengthening of polycrystalline aluminum depending on the grain size.

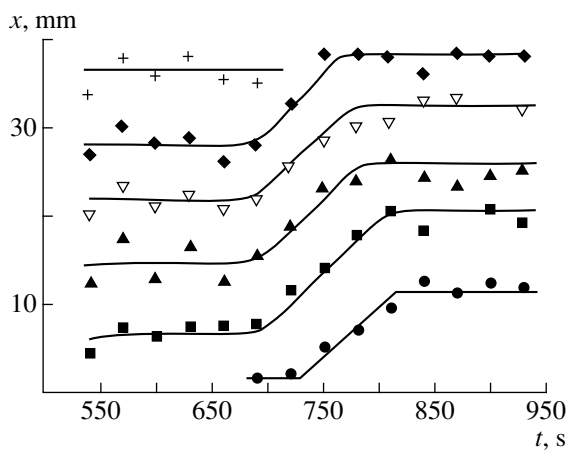


Fig. 3. Displacement of the maxima of localized deformation along the sample axis ($D \approx 0.2$ mm) upon tension at a constant rate.

polycrystalline aluminum that were studied. As is seen from Fig. 2, in all three cases the extension of this stage is small and decreases with increasing grain size D . This agrees with the data of [6], according to which the linear stage is poorly distinguishable at temperatures above 77 K, especially in coarse-grained and single-crystal aluminum. Indeed, the extrapolation of the curves obtained to their intersection shows that $|\varepsilon_c - \varepsilon_b| \rightarrow 0$ as $D \rightarrow 0.5$ mm, i.e., the segment of linear strengthening vanishes, degenerating into a point of transition from one parabolic stage to another with a discontinuous change in the exponent m in the expression

$$\sigma = \sigma_0 + A\varepsilon^m, \quad (1)$$

that describes parabolic strengthening [6].

Owing to this type of the vicissitude of stages in the curve of plastic flow of aluminum, we managed to solve the main problem of the investigation: to plot and analyze the distributions of the ε_{xx} component of the tensor of plastic distortion ($\varepsilon_{xx} \equiv \partial u / \partial x$ is the local elongation in the direction x along the axis of tension of the sample; u is the component of the vector \mathbf{r} in the x direction) in the process of deformation of the same material strengthened according to various laws. It is suitable to represent the data for the analysis of the evolution of deformation localization in the form of the dependence of the position X^* of the maxima of the quantity ε_{xx} along the axial line of the sample on the deformation time $t \sim \varepsilon$, as is shown in [3]. The dependence obtained (Fig. 3) shows that the behavior of ε_{xx} at the linear and parabolic stages of strengthening is different. It follows from Fig. 3 that the stage of linear strengthening is associated with the inclined segments of the $X^*(t)$ dependences; the velocity of propagation of deformation waves in this case is $V_w = dX^*/dt$ (see table). The stages of parabolic strain hardening in Fig. 3 correspond to the horizontal segments of the $X^*(t)$ dependences, for which $V_w = 0$. Thus, the regularities preliminarily established on a wide circle of metals and alloys [1–4] manifest themselves upon extension of polycrystalline aluminum in the total measure: at the parabolic stages, stationary systems of zones of plastic flow appear, whereas at the stage of linear strengthening, a system of zones of localized deformation that moves over the sample arises (wave process).

DEVELOPMENT OF LOCALIZED DEFORMATION

The localization of plastic deformation in polycrystalline aluminum has characteristic features that arise and develop in accordance with the change of stages of strain hardening. In the course of plastic flow, a system of zones of localized deformation spontaneously arises in samples and evolves in the form of a wave process or a stationary structure, and the type of the pattern of localization of these zones is completely determined by the law of strain hardening that is operative at a given moment. Let us discuss the main experimentally established features of the development of such patterns of localized deformation in polycrystalline aluminum.

An analysis of the shape and of the spatial position of the zones of localized deformation in a sample made it possible to establish that they represent parallel layers of plastically deformed material 3–5 nm thick. This is the case both for the parabolic and linear strengthening; the difference is that at the stage of linear strengthening such layers move synchronously. The traces of the layers at the surface of observation are almost perpendicular to the axis of tension of the sample. In this feature, they differ, e.g., from the Lüders bands, for which the slope predicted by the theory of elasticity [8] and usually observed experimentally is $\sim 55^\circ$. The difference appears to be due to the fact that the properties of real coarse-grained aluminum do not fully correspond to the idealized models of plasticity used in calculations. It follows from the analysis of the localization patterns, such as those shown in Fig. 1, that the plasticity isolines are slightly curved and their profiles (steepness of the slopes) are asymmetric.

Thus, alternating deformed and nondeformed layers of approximately equal thickness arise in aluminum polycrystals upon plastic flow. The dynamics of their evolution corresponds to a given stage and to the mechanism of strain hardening that acts at a given moment. Such a delamination of the deformed material occurs spontaneously upon tension at a constant rate and appears to represent a general property of the process of plastic deformation [1, 2].

The plastic flow develops in such a stratified material as follows. At the stages of parabolic strengthening, the entire plastic deformation proves to be almost completely localized inside immobile active zones; the material between these zones is virtually not deformed. At the stage of linear strengthening, the plastic deformation occurs so that the process gradually involves new layers of the material adjoining the leading fronts of the mobile zones of plastic flow, whereas the deformation in the layers at the back edges of these zones ceases.

At the stage of linear strengthening, the process of localization acquires a wave nature and is characterized by a wavelength λ and a velocity of wave propagation V_w (see table). In the investigated cases, the magnitude of V_w decreases with increasing coefficient Θ , which

Dependence of the parameters of deformation waves in polycrystalline aluminum on the grain size (stage of linear deformation strengthening)

Grain size D mm ($D^{-1/2}$, $\text{mm}^{-1/2}$)	Strain-hardening coefficient Θ , 10^{-8} Pa (Θ/E) $\times 10^3$	Wave-length λ , mm	Velocity of propagation of deformation waves $V_w \times 10^{-5}$ m/s
0.2 (2.24)	1.58 (2.24)	7.0 ± 0.5	8
0.3 (1.82)	1.42 (2.01)	5.8 ± 0.2	8
0.4 (1.58)	1.38 (1.96)	4.5 ± 0.2	13

qualitatively agrees with the previously established (and explained in [5]) expression

$$V_w = V_0 + J/\Theta, \quad (2)$$

which relates these two characteristics of the deformed medium. In comparison with the data of [5] on the $V_w(\Theta)$ dependence for single crystals of alloyed γ -Fe, the linearity in the case of aluminum is less -or- more poorly fulfilled. This appears to be explained by the fact that the changes in the strain-hardening coefficient Θ in this case are due to differences in the grain size D in various samples. It follows from the quantitative analysis of the relation between the strain-hardening coefficient and the grain size, the $\Theta(D)$ function has the form analogous to the well-known Hall–Petch relationship:

$$\Theta = \Theta_0 + k_\Theta D^{-1/2} \quad (3)$$

(Fig. 4) with the correlation coefficient $k = 0.97$. Other mechanical properties of aluminum, e.g., flow stress, also can change with changing D . Since J in Eq. (2) has the sense of the energy flux through the deformed sample [5], the change in the flow stress leads to changes in

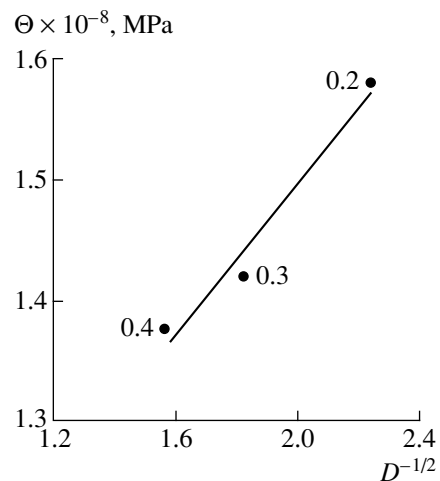


Fig. 4. Strain-hardening coefficient of polycrystalline aluminum (stage of linear strengthening) depending on the grain size. Numerals near the data points denote grain size D (mm).

J and a deviation of the $V_w(\Theta^{-1})$ dependence from linearity predicted by Eq. (2).

At the first and second stages of parabolic strengthening, the alternating actively deformed and nondeformed layers produce a stationary structure in the sample. The deformation localization at the second parabolic stage is so large that the development of such a structure leads to the formation of several (3 to 5) clearly distinguishable necks in the sample. Finally, shortly before the appearance of a macroscopic narrowing of the cross section, the entire plastic deformation is localized in one place, where later a neck is nucleated and ductile fracture of the sample occurs.

DEFORMATION LOCALIZATION AS A PROCESS OF ORDERING IN A DEFORMED MEDIUM

In the evolution of deformation structures upon plastic flow of polycrystalline aluminum, clear regularities can easily be traced which are also characteristic of other, previously investigated materials. After the yield limit is reached at the first stage of parabolic strengthening, a stationary system of localized zones of plastic flow arises in the sample. At the stage of linear strengthening, such zones move synchronously along the sample at a constant rate. The transition to the second stage of parabolic strengthening is accompanied by the formation of an immobile system of zones of localized flow.

It should be noted that upon transition from one stage to another the existing structure of zones of localized deformation becomes "broken up" and a new structure, corresponding to the law of deformation strengthening that acts at a given stage arises in its place. Due to the changes in the law of deformation, the distribution of the localized-deformation zones becomes disordered, their amplitude decreases by a factor of about two, and their number increases (some zones become divided into two). Then, in the beginning of the new stage of strengthening, a specific localization pattern (mobile or stationary) is spontaneously formed from the disordered distribution of the zones of plastic flow, gradually acquiring regular features. An important feature of the observed evolution is that it occurs in the course of the continuous extension of the sample at a constant rate.

The duration of the linear stage of strain hardening decreases in coarse-grained samples in such a way that the deformation ϵ_e corresponding to the end of this stage is virtually insensitive to D , whereas the beginning of the linear stage ϵ_b is shifted with increasing grain size into the range of larger deformations (Fig. 2), i.e., the stage of linear deformation at which the flow stress is

$$\sigma = \sigma_b + \Theta\epsilon \quad (4)$$

arises in polycrystalline aluminum at the expense of the corresponding shortening of the first parabolic stage, and this shortening is more pronounced in a fine-

grained material. This conclusion is confirmed by the following feature of the kinetics of transition from the first parabolic stage to the linear stage: while the $\sigma(\epsilon)$ diagram goes into the linear stage ($\sigma \propto \epsilon$), the zones of localized deformation still remain immobile; their motion begins after an approximately one-fourth of the duration of the stage of linear strengthening. Therefore, the duration of the transition between stages determined from the $\sigma(\epsilon)$ curve proves to be smaller than the time required for the formation of a new pattern of localized deformation. This feature agrees well with the results of [6], where it was shown that the deformation corresponding to the beginning of the second parabolic stage is virtually independent of the grain size in aluminum.

The microstructural features corresponding to the various stages of the flow curve in polycrystalline aluminum were qualitatively considered in [6]. There are also numerous works devoted to electron-microscopic investigations of the structure of deformed aluminum (see, e.g., [9]). Unfortunately, it is impossible to clarify the origin of the macroscopic localization of flow and, moreover, to understand its details at the various stages of deformation strengthening on the basis of only microscopic investigations. In this work, using the X-ray diffraction technique based on the concepts developed in [10] and a specially designed scheme [11], we established that the change of stages of the flow curve is accompanied by marked quantitative changes in the fine structure of the deformed material. It turned out that already at the first stage of parabolic strengthening, the grains of polycrystalline aluminum become divided into blocks of size χ . The onset of the stage of linear strengthening corresponds to a significant (twofold) increase in the size of such blocks, as shown in Fig. 5. After the termination of the stage of linear strengthening, the block dimensions again decrease.

It is evident that the above behavior can be considered as a result of ordering in the deformed medium. Such an approach, based on the ideas of synergetics [12], finds progressively wider application in explaining processes of structure formation upon plastic flow. Significant part of investigations is devoted to attempts of explaining the formation of dislocation structures of various complexity [13, 14], and it is only in a few works [1, 4, 15, 16] that the processes of self-organization are regarded on a macroscopic level as a spontaneous development of localization zones, and their regular behavior is considered in accordance with the multistage nature of the flow curve [1, 4, 16]. In those works, all possible patterns of localization are considered as various types of autowave processes [17] or dissipative structures that arise in open systems through which an energy flux passes.¹

¹ Autowaves, in contrast to the conventional waves that arise under the action of an external action, are eigenstates of a medium [17]. The media that admit such phenomena are called active. The main differences between deformation autowaves and plasticity waves or elastic waves in solids were considered in [5].

From such a viewpoint, a changeover of the types of localization of plastic flow is related to changes in the state of the deformed medium. Since the most significant changes upon plastic flow in a material consist in the accumulation and redistribution of lattice defects, clear concepts of the relationship between the microstructural parameters and quantitative characteristics of the patterns of deformation localization are required. In this connection, of great importance is the fact, established in this paper, that the transition from a stationary structure of localized zones to a wave-type structure is related to the above-mentioned noticeable increase in the size of blocks, which was determined by X-ray diffraction.

ON THE NATURE OF LONG-WAVELENGTH PERIODICITY

As was noted earlier [1–4], the most complex part of the problem is the explanation of the magnitude of the wavelength (distance between zones of plastic flow), which is 5–10 nm on the order of magnitude. The solution suggested in this work is based on the following model. A polycrystal loaded by an external device represents a mosaic of stressed domains of microscopic scale, which are connected with stress concentrators at which the generation of dislocation shears proves to be more probable. Each shear in this case is considered as a discontinuous (jumplike) relaxation process. Such jumps in aluminum were studied in detail at helium temperatures [18]. A relaxation jump is accompanied by the emission of elastic pulses (acoustic emission) [19]. Thus, it is characteristic of the state of a stressed material that, in the system of elastically stressed concentrators, there walk random elastic pulses, whose superimposition on the static fields of the stress concentrators increases the probability of relaxational events of plastic deformation.

Let us consider the propagation of such a pulse through a zone with a variable dislocation density (through a zone of plastic flow that had been spontaneously divided into blocks). For simplicity, we assume that the dislocation density in each block decreases from its center to the periphery in a linear law. In accordance with the known relation [20] $\varepsilon \approx b\rho l$ (where b is the Burgers vector, ρ is the dislocation density, and l is the average dislocation free path), this results in an analogous distribution of deformation in such a zone. As follows from Fig. 5, the velocity of sound in the deformed aluminum at the stage of parabolic strengthening changes with deformation ($V_s \propto \varepsilon$). This means that upon the passage of a plane wave through such a zone, the wave front rotates by a small angle α , as shown in Fig. 6, and the waves from neighboring domains, which play the role of acoustic lenses, will be focused on the symmetry axis. This results in an increase in the concentration of elastic stresses and a related increase in the probability of the occurrence of relaxational acts of plasticity in the zone of stress

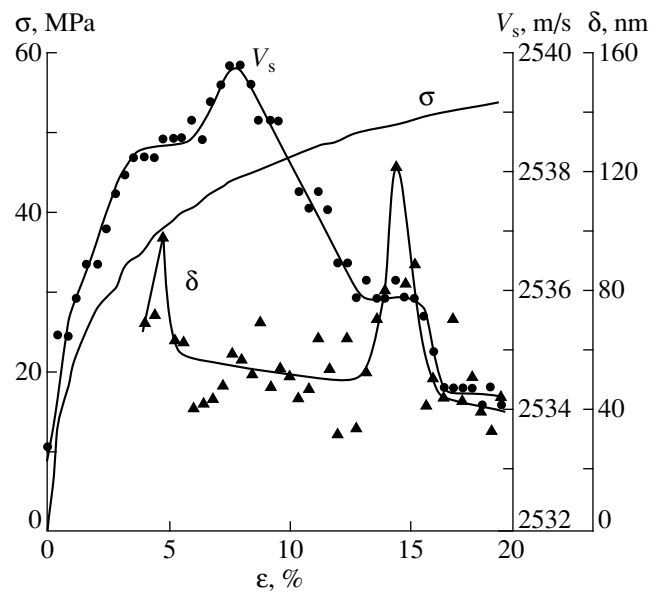


Fig. 5. Variation of the dimensions of microstructure blocks (δ) and the velocity of ultrasound propagation (V_s) upon deformation of aluminum.

focusing; this, in turn, leads to the formation of a new zone of deformation at a distance λ from the initial zone. A simple geometric calculation (see Fig. 6 for details and designations) shows that this distance is

$$\lambda \approx \chi/2 \sin \alpha \approx \chi/2 \tan \alpha \approx (\chi/2) \times 1/(\Delta V \delta / \chi V_s) \approx \chi(\chi/2\delta)(V_s/\Delta V). \quad (5)$$

In aluminum, the sound velocity is $V_s = 6 \times 10^3$ m/s and its experimentally found change in the range of plastic deformations corresponding to the stage of parabolic strengthening is $\Delta V \approx 10$ m/s [7]. For reasonable values of the cross-sectional size of blocks $\delta \approx 10^{-7}$ m (Fig. 5) and for the ratio $\chi/2\delta \approx 10$, we obtain $\lambda \approx 6 \times 10^{-3}$ m, which is close to the experimentally observed distance between the zones of localized deformation. It is important that $\lambda \gg \chi$, so that relation (5) specifies the relationship between the scales of the phenomena on the microscopic (dislocation) and macroscopic (zones of deformation localization) levels of the process of plastic flow. An analysis of the geometry of focusing shows that such an effect can manifest itself on both the ascendant and descendant branches of the $V_s(\varepsilon)$ dependence (Fig. 5).

It seems that the role of such acoustic lenses may be played by dislocation ensembles with a nonuniform distribution of dislocations (and deformations) in them, such as dislocation tangles, cells, etc. [9], as well as twins. In this case, the various schemes of distribution and behavior of macroscopic zones of localized deformation can naturally be related to changes in the geometry of the acoustic lenses (i.e., in the magnitudes of χ , δ , and their ratio) considered above or to the redistribu-

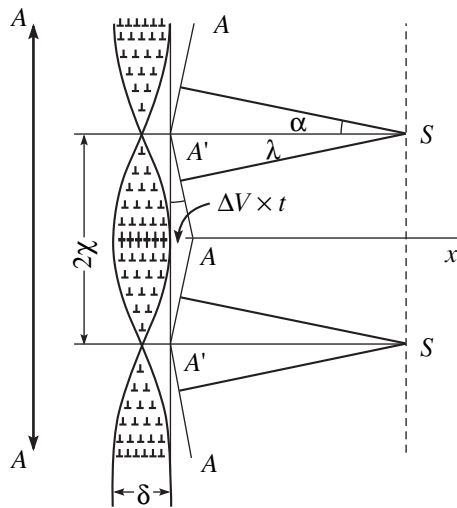


Fig. 6. To calculation of the distance between the zones of localized plastic flow: $A-A$, position of the wave front before the passage through the deformation zone; $A'-A'-A'$, position of the wave front after the passage through the deformation zone; and S , probable sites of nucleation of new deformation zones.

tion of dislocations in them in the course of plastic flow. Thus, in accordance with Eq. (5), the above-noted increase in the block dimensions δ leads to a corresponding change in the distance λ , which manifests itself in the displacement of the zone of plastic flow along the axis of tension at the stage of linear deformation strengthening.

CONCLUSION

Thus, the investigation of the evolution of the localization of plastic deformation in polycrystalline aluminum made it possible to observe the formation and decomposition of time-spatial ordered patterns of localized plastic flow. Their character and evolution completely correspond to the previously established types of relation of the patterns of localization with the law of deformation strengthening that is operative at a corresponding stage of the plastic flow curve. According to the model suggested, the formation of such ordered systems of plastic-flow zones is due to the fact that the randomly arising regions of localized deformation serve as acoustic lenses that focus the acoustic radiation generated by relaxational events at a certain distance from the operative deformation zone, which

increases the probability of the realization of the next event of plastic flow in the region of focusing.

REFERENCES

1. L. B. Zuev, V. I. Danilov, and V. V. Gorbatenko, *Zh. Tekh. Fiz.* **65** (5), 91 (1995) [*Tech. Phys.* **40**, 456 (1995)].
2. L. B. Zuev and V. I. Danilov, *Int. J. Solids Struct.* **34**, 3795 (1997).
3. L. B. Zuev, S. A. Barannikova, V. I. Danilov, *et al.*, *Zh. Tekh. Fiz.* **69** (10), 56 (1999) [*Tech. Phys.* **44**, 1179 (1999)].
4. L. B. Zuev and V. I. Danilov, *Philos. Mag. A* **79** (1), 43 (1999).
5. S. A. Barannikova, L. B. Zuev, and V. I. Danilov, *Fiz. Tverd. Tela (St. Petersburg)* **41** (7), 1222 (1999) [*Phys. Solid State* **41**, 1112 (1999)].
6. B. Jaoul, *J. Mech. Phys. Solids* **5** (1), 95 (1957).
7. L. B. Zuev, B. S. Semukhin, K. I. Bushmelyova, and N. V. Zarikovskaya, *Mater. Lett.* **42** (1), 97 (2000).
8. T. Thomas, *Plastic Flow and Fracture in Solids* (Academic, New York, 1961; Mir, Moscow, 1964).
9. P. R. Suon, *Electron Microscopy and Strength of Crystals* (Metallurgiya, Moscow, 1968), pp. 123–168.
10. B. E. Warren, *Phys. Rev.* **59**, 693 (1941).
11. L. B. Zuev, T. M. Poletika, and B. S. Semukhin, *Kristallografiya* **40**, 1071 (1995) [*Crystallogr. Rep.* **40**, 996 (1995)].
12. G. Nicolis and I. Prigogine, *Exploring Complexity: An Introduction* (Freeman, New York, 1989; Mir, Moscow, 1990).
13. G. A. Malygin, *Fiz. Tverd. Tela (St. Petersburg)* **37**, 3 (1995) [*Phys. Solid State* **37**, 1 (1995)].
14. G. A. Malygin, *Usp. Fiz. Nauk* **169**, 979 (1999) [*Phys.-Usp.* **42** (9), 887 (1999)].
15. E. S. Aifantis, *Int. J. Eng. Sci.* **30**, 1279 (1992).
16. L. B. Zuev and V. I. Danilov, *Fiz. Tverd. Tela (St. Petersburg)* **39**, 1399 (1997) [*Phys. Solid State* **39**, 1241 (1997)].
17. V. A. Vasil'ev, Yu. M. Romanovskii, and V. G. Yakhno, *Autowave Processes* (Nauka, Moscow, 1987).
18. O. V. Klyavin, *Physical Processes of Plastic Deformation at Low Temperatures* (Naukova Dumka, Kiev, 1974), pp. 5–30.
19. V. S. Boiko and V. D. Natsik, *Elementary Processes of Plastic Deformation in Crystals* (Naukova Dumka, Kiev, 1978), pp. 159–189.
20. G. Taylor, *Proc. R. Soc. London, Ser. A* **145**, 362 (1934).

Translated by S. Gorin

Decomposition of Solid Solution in an Alloy under High Plastic Strain

N. M. Vlasov, A. S. Gontar', and V. A. Zaznoba

Received August 4, 2000

Abstract—The kinetics of impurity atmosphere formation and phase precipitation near the triple joints of deformation boundaries of the cellular structure is considered. The decomposition of solid solutions of substitutional impurities when an alloy is under high plastic strain is theoretically analyzed. © 2001 MAIK “Nauka/Interperiodica”.

At large plastic strains in an alloy ($\varepsilon > 0.1$ – 0.2), the cellular structure with a characteristic cell size of 1 – $2 \mu\text{m}$ is formed [1, 2]. Microscopically, the formation of such structures is due to two processes: glide and climb of edge dislocations. Boundaries between the cells are named deformation boundaries, since they arise under plastic strain. If the structure continuity is undisturbed, the disorientation of adjacent cells generates a self-balancing system of internal stresses. The stresses are maximal near the triple joints of deformation boundaries. The dilatation field of the joints is simulated by a wedge disclination [3]. In the stress field produced by disclinations of various sign, substitutional impurities are separated: those with an atomic radius larger than that of a solvent diffuse to the positive-dilatation disclination, while those with smaller atomic radii move toward the negative-dilatation disclination. If the impurity concentration near a wedge disclination exceeds the solubility limit at a given temperature, macroscopic concentration inhomogeneities like impurity atmospheres or second-phase precipitates arise, indicating the decomposition of impurity solid solutions. In this work, we studied the kinetics of these processes near the triple joints of deformation boundaries in the cellular structure. The decomposition of solid solutions of substitutional impurities under high plastic strain of an alloy is theoretically analyzed.

The potential of interaction between an impurity atom and a wedge disclination is given by

$$V = A \ln \frac{r}{R_1}, \quad A = \frac{\mu \omega (1 + \nu) \Delta v}{3\pi(1 - \nu)}, \quad (1)$$

where μ is the shear modulus, ν is Poisson's ratio, ω is the magnitude of the disclination rotation vector, Δv is a change in the crystal volume when an impurity atom is introduced, and R_1 is the outer disclination radius.

The logarithmic divergence of the potential V at $r \rightarrow 0$ and $R_1 \rightarrow \infty$ is eliminated if these parameters

are assumed to be finite. The radius r_0 of the wedge disclination core is comparable to the characteristic thickness of deformation boundaries (several interatomic distances). The outer disclination radius R_1 equals half the cell size. Since $R_1/r_0 \approx 10^4$, diffusion problems are sometimes solved for an unbounded region. For a positive-dilatation wedge disclination (tensile stresses), $\omega < 0$; for a negative-dilatation disclination, $\omega > 0$ (compressive stresses). Near the former, impurity atmospheres with large atomic radii form; near the latter, smaller radius impurities are concentrated. Note that the sign of Δv changes with the disclination sign. Therefore, the value of the potential V remains constant. Physically, this means that diffusion of substitutional impurities with various atomic radii toward associated disclinations is governed by the same equations.

The formation kinetics of impurity atmospheres in the vicinity of a wedge disclination is described by the unsteady-state equation of diffusion in the field of the potential V subject to appropriate initial and boundary conditions:

$$\frac{1}{D} \frac{\partial C}{\partial t} = \Delta C + \frac{\nabla(C \nabla V)}{kT} \quad (r_0 \leq r \leq R_1), \quad (2)$$

$$C(r, 0) = C_0, \quad C(R_1, t) = C_0, \quad C(r_0, t) = C_{\text{eq}},$$

where D is the impurity diffusion coefficient, k is the Boltzmann constant, T is absolute temperature, C_0 is the mean impurity concentration, and C_{eq} is the equilibrium impurity concentration near the disclination core.

Let us briefly review the physical meaning of the initial and boundary conditions adopted in our model. At the initial time instant, the substitutional impurities are uniformly distributed over the volume and their concentration equals the mean concentration C_0 . The same holds at the boundary $r = R_1$. In physical terms, this means that the loss of impurity atoms that move to the disclination core (to the boundary $r = R_1$) is com-

compensated for by the atoms arriving from adjacent disclinations of the opposite sign. This follows from the self-balance of the stress system at the triple joints of cell boundaries. The condition at the boundary $r = r_0$ means that here the equilibrium impurity concentration is set instantaneously and then remains unchanged throughout the time of impurity distribution by diffusion. At $r = r_0$, the equilibrium impurity concentration is conventionally expressed as

$$C_{\text{eq}} = C_0 \exp\left(\frac{V}{kT}\right)\Bigg|_{r=r_0}. \quad (3)$$

The logarithmic coordinate dependence of the interaction potential V significantly simplified problem (2):

$$\frac{1}{D} \frac{\partial C}{\partial t} = \frac{\partial^2 C}{\partial r^2} + \frac{1 + \alpha}{r} \frac{\partial C}{\partial r}, \quad \alpha = \frac{A}{kT} < 0, \quad (r_0 \leq r \leq R_1),$$

$$C(r, 0) = C_0, \quad C(R_1, t) = C_0, \quad (4)$$

$$C(r_0, t) = C_{\text{eq}} = \left(\frac{r_0}{R}\right)^\alpha.$$

The sign of the dimensional parameter α does not depend on which (large or small) impurity atoms diffuse to associated (positive- or negative-dilatation) disclinations. In fact, when large impurity atoms diffuse to a positive-dilatation disclination, $\alpha < 0$, since $\omega < 0$ and $\Delta v > 0$. When small atoms diffuse to a negative-dilatation disclination, $\alpha < 0$, since $\omega > 0$ and $\Delta v < 0$. This immediately follows from relationship (1). The parameter α is the binding energy between an impurity atom and a wedge disclination divided by thermal energy. If $|\alpha| \ll 1$, the stress field of a wedge disclination can be considered as a small perturbation of the basic diffusion flux. If $|\alpha| \gg 1$, the disclination field governs the impurity diffusion flux. Finally, at $|\alpha| \approx 1$, the diffusion fluxes due to the potential V and concentration gradients are comparable to each other. Let us evaluate $|\alpha|$ for the Mo–Nb system. At $kT = 2 \times 10^{-20}$ J, $\mu = 1.34 \times 10^{11}$ Pa, $\nu = 0.3$, $\omega = 0.3$ rad, and $\Delta v = 2.7 \times 10^{-30}$ m³, we have $|\alpha| = 1$. The disorientation angle between adjacent cells $\omega = 0.3$ rad seems to be quite reasonable for the strain $\varepsilon \approx 0.3$.

Next, consider the kinetics of the formation of an impurity atmosphere near a positive-dilatation wedge disclination. For $\alpha = -1$, the concentration field of large impurity atoms is found from a solution to the problem

$$\frac{1}{D} \frac{\partial C}{\partial t} = \frac{\partial^2 C}{\partial r^2} \quad (r_0 \leq r \leq R_1),$$

$$C(r, 0) = C_0, \quad C(R_1, t) = C_0, \quad (5)$$

$$C(r_0, t) = C_{\text{eq}} = C_0 \left(\frac{R_1}{r_0}\right)$$

and has the form

$$\frac{C - C_0}{C_{\text{eq}} - C_0} = \frac{R_1 - r}{R_1 - r_0} - \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{\sin \frac{\pi(r-r_0)}{(R-r_0)}}{n} \exp\left[-\frac{\pi^2 n^2 D t}{(R_1 - r_0)^2}\right]. \quad (6)$$

It is seen that the stress field of an wedge disclination turns the problem with cylindrical symmetry to the plane problem; that is, a cylindrically symmetric impurity atmosphere is formed according to the plane symmetry law. This greatly enhances the formation rate of the impurity atmosphere, as directly follows from the diffusion equations in (4) and (5). Indeed, all other things being equal, at $\partial C/\partial r < 0$, the rate of change of the impurity concentration ($\partial C/\partial t$) in (5) is larger than in (4) at $\alpha = 0$ [the second term in the right of (4), $(1/r)(\partial C/\partial r) < 0$, diminishes $\partial C/\partial t$]. The impurity atmosphere here is formed because the large atoms leave adjacent negative-dilatation disclinations. In the vicinity of these disclinations, their own small-atom atmospheres form (small atoms leave positive-dilatation disclinations). The formation kinetics of the latter also obeys (5) and (6). In other words, the solid solution of substitutional impurities becomes layered. This is typical of multicomponent alloys, where the atomic radius of substitutional impurities differs from that of the matrix metal. Therefore, in multicomponent alloys, impurity atmospheres at the triple joints of deformation boundaries have different compositions. Interstitial impurities behave in a somewhat different manner. They form atmospheres only near positive-dilatation wedge disclinations, since any interstitial impurity increases the crystal volume.

If the concentration of substitutional impurities near wedge disclinations exceeds their solubility limit, new phases appear. They nucleate and grow at wedge disclinations of various signs with an equal probability. The growing precipitates capture the substitutional impurities from the solid solution and deplete it. The diffusion growth of precipitates proceeds as follows. First, a nucleus with a characteristic radius R_0 rapidly forms; then, it grows by diffusion. During this process, the concentration of substitutional impurities at the moving boundary of the precipitate changes stepwise: $C = C_1$ in the new phase and $C = C_2$ in the surrounding matrix ($C_1 > C_2$ and $C_2 < C_0$, where C_0 is the mean impurity concentration). In physical terms, this means that the precipitate boundary instantaneously captures the

impurities from the solid solution and delivers them to the new phase with a higher concentration. Mathematically, the problem of diffusion growth kinetics for the new phase near a wedge disclination is stated as

$$\frac{1}{D} \frac{\partial C}{\partial t} = \frac{\partial^2 C}{\partial r^2} + \frac{1 + \alpha}{r} \frac{\partial C}{\partial r}, \quad \alpha = \frac{A}{kT} < 0,$$

$$C(R, t) = C_2, \quad C(r, 0) = C_0 \quad (r \geq R_0), \quad C(\infty, t) = C_0, \quad (7)$$

$$(C_1 - C_2) \frac{dR}{dt} = D \left(\left| \frac{dC}{dr} \right| + \left| \frac{c\alpha}{r} \right| \right)_{r=R},$$

where R_0 is the radius of a new-phase nucleus and R is the current radius of the new phase.

It is assumed that the outer radius of a wedge disclination far exceeds the characteristic size of a precipitate. This allows us to consider the new-phase growth in the infinite matrix. If there exist large substitutional atoms ($\alpha = -1$) near a positive-dilatation wedge disclination, problem (7) takes the simpler form

$$\frac{1}{D} \frac{\partial C}{\partial t} = \frac{\partial^2 C}{\partial r^2},$$

$$C(R, t) = C_2, \quad C(r, 0) = C_0 \quad (R \geq R_0), \quad C(\infty, t) = C_0, \quad (8)$$

$$(C_1 - C_2) \frac{dR}{dt} = D \left(\frac{dC}{dr} + \frac{c}{r} \right)_{r=R}.$$

If precipitation is controlled by impurity diffusion, the precipitate size grows according to the law $R(t) = \beta \sqrt{Dt}$, where β is a dimensionless parameter of the problem. Its value is found from the equation for mass balance at the interface. In the fixed interface approximation, this equation turns to the quadratic equation for β :

$$\beta^2 - \frac{2\beta}{\sqrt{\pi}} \left| \frac{C_2 - C_0}{C_1 - C_2} \right| - \left| \frac{2C_2}{C_1 - C_2} \right| = 0. \quad (9)$$

Other approaches (for example, the introduction of the new variable $r/2\sqrt{Dt}$) will not qualitatively change the growth kinetics but will make the equation for β more complicated [4]. Moreover, one should realize that the diffusion growth of precipitates is merely an idealized model. In actual systems, second-phase precipitation is governed by two processes: the diffusion of impurities to the interface and their absorption at it. The growth kinetics law changes from $\sim \sqrt{t}$ to $\sim t$, whichever prevails. Our approximation is valid when impurity

atoms are absorbed momentarily and, hence, diffusion becomes the limiting stage of new-phase evolution. With the initial and boundary conditions remaining unchanged and $\alpha = 0$ in (7), we have

$$\frac{1}{D} \frac{\partial C}{\partial t} = \frac{\partial^2 C}{\partial r^2} + \frac{1}{r} \frac{\partial C}{\partial r},$$

$$C(R, t) = C_2, \quad C(r, 0) = C_0 \quad (r \geq R_0), \quad C(\infty, t) = C_0, \quad (10)$$

$$(C_1 - C_2) \frac{dR}{dt} = D \left(\frac{dC}{dr} \right)_{r=R}.$$

Such a statement implies that the new phase grows only owing to the impurity concentration gradient. From the equation for mass balance at the interface, one can derive the transcendental expression for the parameter β_1 , which characterizes the growth kinetics according to the law $\beta_1 \sqrt{Dt}$ [5]:

$$\beta_1 = \frac{2}{\sqrt{\pi}} \left| \frac{C_2 - C_0}{C_1 - C_2} \right| \frac{K_1(\beta_1 \sqrt{\pi}/2)}{K_0(\beta_1 \sqrt{\pi}/2)}, \quad (11)$$

where $K_0(x)$ and $K_1(x)$ are modified Bessel functions of zero- and first order, respectively.

A solution of Eqs. (9) and (11) allows us to estimate the effect of the stress field of a wedge disclination on the precipitation kinetics. Without loss of generality, we put $C_0 = 2 \times 10^{-4}$, $C_2 = 10^{-4}$, and $C_1 = 3 \times 10^{-4}$ at.%. Then,

$$\beta^2 - \frac{\beta}{\sqrt{\pi}} - 1 = 0, \quad \beta_1 = \frac{1}{\sqrt{\pi}} \frac{K_1(\beta_1 \sqrt{\pi}/2)}{K_0(\beta_1 \sqrt{\pi}/2)}. \quad (12)$$

Since $K_1(x)/K_0(x) > 1$, $\beta_1 > 1/\sqrt{\pi}$. On the other hand, from the first relationship in (12), $\beta = 1/\beta + 1/\sqrt{\pi}$, it follows that $\beta > 1/\sqrt{\pi}$. Therefore, it is impossible to arrive at the expected inequality $\beta > \beta_1$ without numerical analysis. Numerically solving Eqs. (12) yields $\beta = 1.3$ and $\beta_1 = 0.8$. For these values, the growth kinetics is described by the relationships

$$R(t) - R_0 = 1.3\sqrt{Dt}, \quad R(t) - R_0 = 0.8\sqrt{Dt}. \quad (13)$$

One can see that the stress field due to a wedge disclination (the model of triple joints at deformation boundaries) significantly enhances the growth rate of precipitates. Other values of the boundary concentration require merely a numerical refinement of β and β_1 . With small substitutional atoms present near a nega-

tive-dilatation disclination, the growth kinetics of precipitates obeys the same mathematical laws.

Thus, when an alloy is subjected to a high plastic strain, the solid solution of substitutional impurities decays. The decay is favored largely by a great number of triple joints (at deformation boundaries) that appear during the formation of the cell structure. The distinctive feature of these structural imperfections, which are simulated by wedge disclinations, is that they generate high stress fields. In fact, a wedge disclination, for which $\sigma_{ii} \sim \ln(r/R)$ and $N(t) \sim t$ (σ_{ii} is the trace of the normal stress tensor and $N(t)$ is the time law of impurity atmosphere formation), is the most powerful defect. Next are a microcrack tip ($\sigma_{ii} \sim 1/\sqrt{r}$, $N(t) \sim t^{4/5}$) and an edge dislocation ($\sigma_{ii} \sim 1/r$, $N(t) \sim t^{2/3}$) [6]. That is why impurity atmospheres and precipitates form just near the triple joints of deformation boundaries. At a high plastic strain, the concentration of such defects grows because the cells become smaller.

REFERENCES

1. V. I. Trefilov, Yu. V. Mil'man, and S. A. Firstov, *Physical Principles of Refractory Metal Strength* (Naukova Dumka, Kiev, 1975).
2. V. V. Rybin, *Large Plastic Deformation and Metal Destruction* (Metallurgiya, Moscow, 1986).
3. N. M. Vlasov and V. A. Zaznoba, *Fiz. Tverd. Tela* (St. Petersburg) **41**, 64 (1999) [*Phys. Solid State* **41**, 55 (1999)].
4. B. Ya. Lyubov, *Diffusion Processes in Inhomogeneous Solid Media* (Nauka, Moscow, 1981).
5. N. M. Vlasov and V. A. Zaznoba, *Dokl. Akad. Nauk* **363**, 472 (1998) [*Dokl. Phys.* **43**, 761 (1998)].
6. B. Ya. Lyubov and N. M. Vlasov, *Fiz. Met. Metalloved.* **47** (1), 140 (1979).

Translated by V. Isaakyan

A Statistical Model of Fullerene Formation Based on Quantum-Mechanical Calculations

I. The Most Probable Precursors of Fullerenes

N. I. Alekseev and G. A. Dyuzhev

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

e-mail: dgan@hm.csa.ru

Received August 28, 2000

Abstract—An analysis is carried out of the carbon clusters that may be immediate precursors of fullerenes. Using quantum chemistry methods it is shown that these can be two-ring clusters with a single bond between the rings. Possible initial steps of the transformation of a two-ring cluster into a fullerene have been outlined.
© 2001 MAIK “Nauka/Interperiodica”.

In [1], a scheme for calculating the kinetics of fullerene formation from multiple-ring structures was proposed. All possible transformation routes can, in principle, be followed in this scheme for a given starting object and a prediction made about the fraction of clusters closing under the conditions considered and the fraction remaining unassembled. The scheme was entirely speculative and based on the following assumptions.

(1) Intermediate clusters in the growth of fullerenes can be characterized by three parameters: the number of pentagons N_5 , the number of hexagons N_6 , and the number of vacant radicals ρ .

(2) All topologically possible routes for the rings to be bonded to the growing cluster nucleus are equivalent and feasible.

(3) To describe the evolution of the clusters quite a small number of reactions is sufficient.

(4) Incompletely assembled clusters with a large surplus of pentagons or hexagons (necks or shells) fail to develop further and the bonds in them can only break.

Before using this scheme in the calculation of fullerene formation in an arc discharge, the starting object has to be defined, the assumptions it is based upon verified, and, if necessary, corrected. The present study is devoted to solving the first of these problems.

We used both semi-empirical methods of quantum chemistry (mainly the AM1 method) and various estimates. With the help of the AM1 method, we determined the energies of given configurations and the corresponding atom positions. Specifying reactions inside the cluster, one can calculate the parameters of the corresponding transition states and determine the reaction probability based on the theory of transition state [2].

1. In order to successfully simulate the process of fullerene formation in an arc discharge it is necessary first of all to identify the most probable immediate precursors of fullerenes. As such precursors, either two-ring clusters with four triply bonded atoms [3] or three-ring clusters [4] were proposed. In these studies the dynamics of such clusters was considered for the most probable route in every instance. However, different objects were not compared using the same method and it was not determined which of them actually dominates fullerene formation under specified conditions.

Experiments on the mobility in drift tubes [5] give a definite picture of the evolution of carbon clusters up to the phase of ring formation (atoms \rightarrow molecules \rightarrow chains \rightarrow rings). It can be supposed that the next object on the route to fullerene formation will be a two-ring cluster produced as a result of the collision of rings. Its most likely configurations are shown in Fig. 1. Cluster *I* forms when two rings collide in such a way that the atom of one ring hits an atom of the other ring and the inertia of the relative motion of the rings is not enough for overcoming the elasticity of the bonds. Cluster *II* analyzed in [3] forms in a situation when the two rings come within a bond's length of one another and then part of one ring penetrates the other ring. In

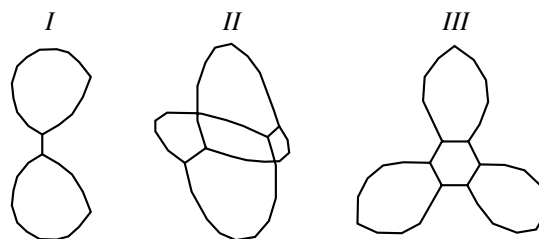


Fig. 1. Clusters with different topologies: (*I*) two-ring cluster with two triply bonded atoms, (*II*) two-ring cluster of the two-ring adduct type [3], and (*III*) three-ring cluster.

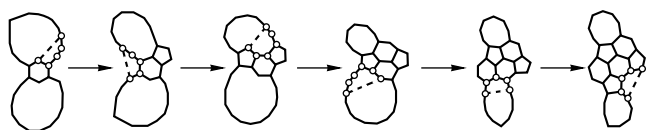


Fig. 2. First reactions in the development of cluster *I*; formation of a singly bonded frame. $N = 32$.

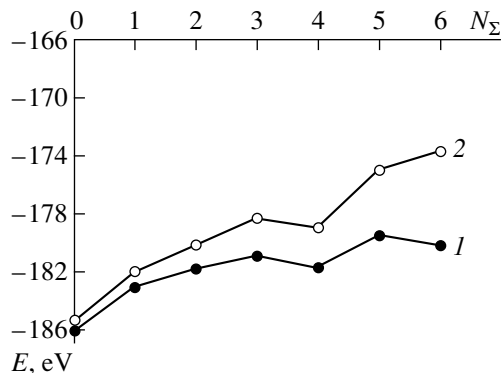


Fig. 3. Energy variations in the evolution of two-ring clusters *I* (1) and *II* (2). Identical reaction chains are taken, N_Σ is the total number of pentagons and hexagons in a cluster.

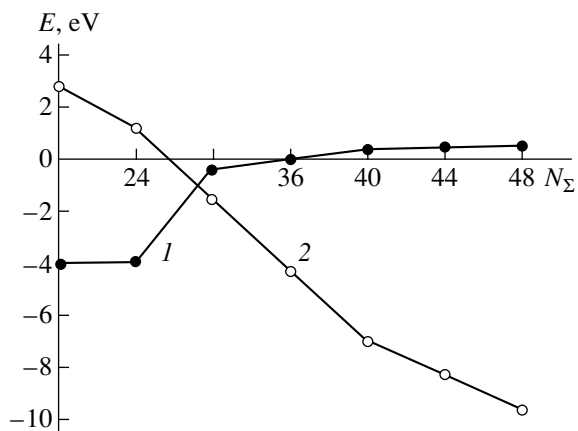


Fig. 4. Energies of clusters (*I*) and fullerenes (*2*) as functions of the number of atoms N . The zero of energy is taken as the value of the total energy of two rings of $(N/2)$ atoms.

this case the inertia of motion overcomes the elasticity of bonds and one more bond is formed. The minimum additional kinetic energy needed in this case is equal to the barrier for the rings to get close to each other and, according to our estimates, amounts to 0.5–0.6 eV (the same as in [6]). Thus, the relative occurrence of the formation reaction of cluster *II* at a temperature of 0.25 eV can be estimated at 10–20% of the total number of collisions.

2. Consider the possible ways to transform clusters *I* and *II* into fullerenes and compare their efficiencies. For cluster *I*, it will be natural to evolve through a sequence of reactions producing pentagons and hexa-

gons (possibly also tetragons). One of this type of reaction chains is shown in Fig. 2. It is seen that in every transformation phase, the cluster consists of a central singly bonded frame and the remains of the rings (ring fragments).

For clusters *II*, this route is also possible, with the growth point as one of the initial bonds. But the other original bond usually interferes with the correct closing of the structure and must eventually be broken. Yet even prior to this, other conditions being equal, cluster *II* must be more rigid and less reactive than cluster *I*. This is seen in Fig. 3, which shows the energy variation with the evolution of the clusters. The reaction chains for the two clusters are assumed to be identical. It is seen that in terms of energy, clusters *II* are in a less advantageous situation than clusters *I*. So, among the two-ring clusters, cluster *I* is a more probable precursor of fullerenes.

3. The next step in solving the problem should have been a comparison of the evolution of cluster *I* with three-ring clusters *III* (Fig. 1) formed when a two-ring cluster captures a third ring. Such clusters were considered in [4]. Under real conditions of an arc discharge, clusters *III* are formed much farther away from the discharge axis than clusters *I* and their concentration should be much lower. Therefore, they should be taken into account only if their rate of transformation into fullerenes is at least 1–2 orders of magnitude higher than that of type *I* clusters.

Thus, in the search for a fullerene precursor the fullerene synthesis process should be considered practically throughout, i.e., major routes of convolution into fullerene taken by two- and three-ring clusters of close dimensions should be established and characteristic evolution times for these routes compared. Then, concentrations of the two- and three-ring clusters in the discharge should be estimated by calculating at least some simplified transformation dynamics of the preceding clusters.

4. Consider the evolution of a two-ring cluster *I* along chains of reactions in which only pentagons and hexagons are formed (like the chain shown in Fig. 2). The reactions were analyzed beginning with a cluster of $N = 32$ atoms, this being a minimum cluster size for which a fullerene isomer exists with an energy lower than that of the two-ring cluster (Fig. 4).

Directions of possible reactions for cluster *I* are shown in Fig. 5 for a cluster with $N = 40$. In the variations shown in Fig. 6, in the course of reaction, the energy ΔE and the reaction energy barrier E_a as functions of the number of atoms are shown for some reactions. In the first step of reaction *a*, a bond forms between the atoms of the ring that are the next closest to the cross bond, resulting in a hexagon. In reaction *b*, the first and the second diagonal atoms of the rings are bonded. However, a stable pentagon fails to form (cluster *A*) and the reaction is completed with one of the atoms sticking out (cluster *C*). It can form a hexagon (“*D*”), though it is evident that the total time of such a

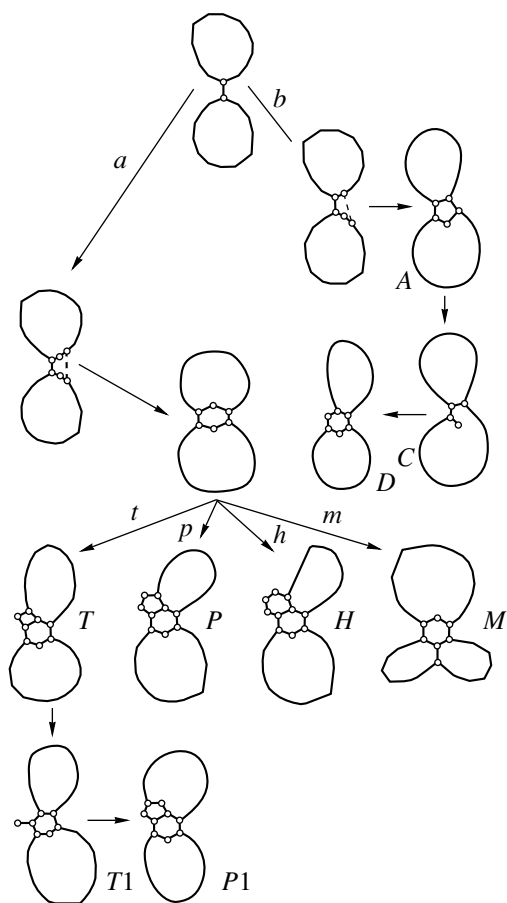


Fig. 5. Formation reactions for pentagons and hexagons in the first evolutionary step of a two-ring cluster *I*.

cascade reaction is longer than in reaction *a*, in which a hexagon is formed at the very beginning. The barrier of reaction *a* diminishes with cluster size from ~4 eV for $N = 20$ to 1.8 eV for $N = 48$.

In the second step of the route through the hexagon, the double-bonded hexagon atom binds one of the ring atoms, resulting in either tetragon *T*, pentagon *P*, or hexagon *H* (reactions *t*, *p*, and *h*). The ring fragment may also be divided into two approximately equal parts (reaction *m*). It is seen in Fig. 6 that for small clusters, the least energy is spent in the formation of the pentagon and, for larger clusters (starting with $N = 44-46$), in type *m* reactions.

Topologically, cluster *M* is nearly equivalent to a three-ring cluster considered in [4]. Though the barrier of reaction *m*, $E_a^{(m)} \sim 3.5$ eV, is considerably higher than $E_a^{(p)}$, this reaction should be taken into account, because when the number of atoms is large the three-ring clusters, as shown below, they demonstrate a higher rate of closing into a fullerene. On the other hand, estimates show that in a real arc discharge, the effective time of reaction *m* is still much larger than the time of the third ring capture. Therefore, in what fol-

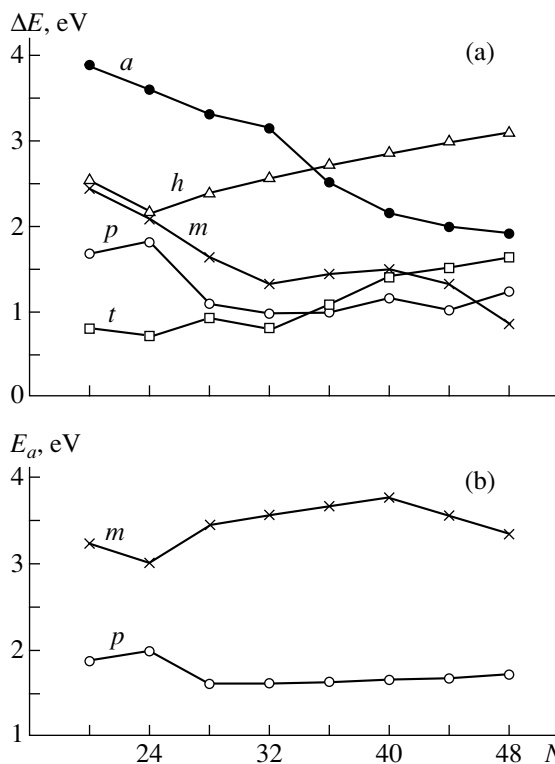


Fig. 6. Energies involved in the reactions presented in Fig. 5 for different numbers N of the atoms: (a) energy variations for different reactions; and (b) energy barriers for the reactions.

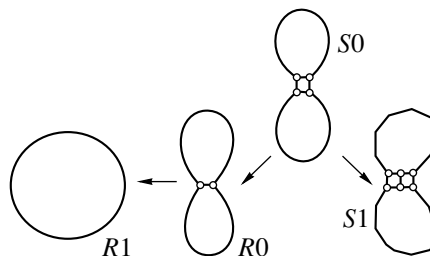


Fig. 7. Ladder route of the reactions with a central tetragon.

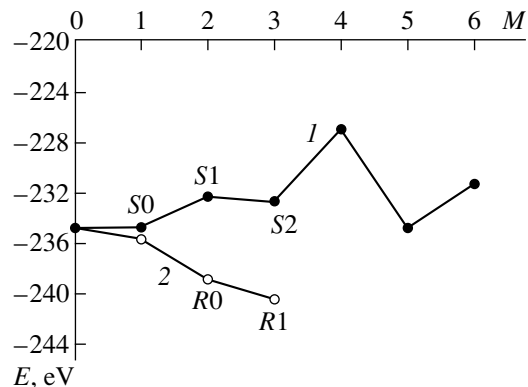


Fig. 8. Energy diagram for the evolution of the cluster with a central tetragon.

lows, reactions p and m are considered in parallel and the three-ring clusters proposed in [4] are not considered at all.

Reaction h consumes much more energy than p and, as distinct from m , nothing is gained in subsequent kinetics. Reaction t producing a tetragon at $N \geq 36$ is somewhat less advantageous compared with p . Besides, cluster T easily slips into a more preferable configuration with one of the atoms sticking out ($T1$ in Fig. 5). The same configuration is obtained most often with cluster T when the optimization is started with a bad approximation. In the next step a pentagon is formed but with a different bonding of the rings ($P1$); i.e., the end result is again a version of reaction p .

Thus, the process in which a fullerene is built only from pentagons and hexagons requires no tailoring to match the scheme but readily follows it.

5. The calculation of the development of the two-ring cluster I using AM1 shows that this cluster transforms into a cluster with a central tetragon (studied alongside other clusters in [6]) practically without meeting any barrier. It is not known if this result is true or is a consequence of the tolerance of small valence angles using this method. Therefore, it is necessary to consider possible routes of evolution of the cluster with a central tetragon (Fig. 7, $S0$), especially since it has a lower energy than cluster I .

There are two such routes. The first involves increasing the number of M tetragons (Fig. 7, transition $S0 \rightarrow S1$ and so on). However, along this route, already at $M = 4$, the cluster energy becomes very large (Fig. 8, curve 1), which makes this route a dead-end.

A different method is more likely ($S0 \rightarrow R0 \rightarrow R1$ in Fig. 7). This chain of reactions is always preferable in terms of energy, especially in the case of small

two-ring clusters (Fig. 8, curve 2), and its likelihood is indirectly confirmed in experiments with drift tubes [5]. The barrier for the second reaction of this chain amounts to 3.4 eV at $N = 32$ or 4.7 eV at $N = 40$ and, apparently, grows further with increasing N . Consequently, the larger the cluster, the lower the possibility of its developing into a simple ring. As will be shown later, the transformation into a fullerene has the opposite tendency.

ACKNOWLEDGMENTS

This study was carried out in the framework of the Russian program "Fullerenes and Atomic Clusters" and supported in part by the Russian Foundation for Basic Research, project no. 00-02-16928.

REFERENCES

1. N. I. Alekseev and G. A. Dyuzhev, *Zh. Tekh. Fiz.* **69** (9), 104 (1999) [*Tech. Phys.* **44**, 1093 (1999)].
2. H. Eyring, S. H. Lin, and S. M. Lin, *Basic Chemical Kinetics* (Wiley, New York, 1980; Mir, Moscow, 1983).
3. D. L. Strout and G. E. Scuseria, *J. Chem. Phys.* **100**, 6492 (1996).
4. V. A. Schweigert, A. L. Alexandrov, Y. N. Morokov, and V. I. Bedanov, *Chem. Phys. Lett.* **235**, 221 (1995).
5. J. M. Hunter, J. L. Fye, and M. F. Jarrold, *J. Chem. Phys.* **3**, 1785 (1993); J. M. Hunter, J. L. Fye, E. J. Roskamp, and M. F. Jarrold, *J. Phys. Chem.* **98**, 1810 (1994).
6. A. L. Aleksandrov, V. M. Bedanov, Yu. N. Morokov, and V. A. Shveĭgert, *Zh. Strukt. Khim.* **36**, 991 (1995).

Translated by B. Kalinin

A Statistical Model of Fullerene Formation Based on Quantum Mechanical Calculations

II. Verification of the Model and Kinetics of Transformation into Fullerene

N. I. Alekseev and G. A. Dyuzhev

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

e-mail: dgan@hm.csa.ru

Received August 28, 2000

Abstract—An analysis of a large number of feasible routes for the transformation of two-ring clusters into fullerenes is carried out. The applicability of a scheme for the abridged description of this transformation developed earlier has been demonstrated as well as the fact that the formation of a small quantity of fullerenes with less than 60 atoms results from the assembly kinetics alone, without resorting to symmetry considerations. It has been shown that the assembly rate increases with cluster size, at least for clusters of less than 60 atoms.
© 2001 MAIK “Nauka/Interperiodica”.

In [1], a scheme of a condensed description of the transformation kinetics of multiple-ring carbon structures into fullerenes was proposed that can be used to calculate the formation of fullerenes in an arc discharge. It has been shown in [2] that the most probable precursor of the fullerene is a two-ring cluster with a single bond between the rings. The purpose of this paper is to consider the kinetics of transformation of the two-ring cluster into the fullerene.

CLOSING KINETICS OF THE TWO-RING CLUSTER WITH THE FORMATION OF A FULLERENE: QUALITATIVE CONSIDERATION

1. Available computing facilities allow the closing kinetics of a two-ring cluster resulting in the formation of a fullerene to be calculated for clusters with the number of atoms N not exceeding 40. In particular, for $N = 32$, a calculation of the cluster closing was performed, taking into account all possible assembly routes, which gave an estimate of the effect of multiplicity of assembly routes and yielded statistical data on the probability of various reactions in the course of assembly. For $N = 40$, the bonding energy variation was considered for a fairly large number of the possible routes (about 20) and some qualitative relationships were formulated. For a number of typical routes, the reaction probabilities were calculated and used to determine the total time of evolution.

2. Figure 1 shows all reaction types adequate for evolution of the cluster towards a fullerene. Primed and unprimed reactions differ in positions of the new car-

bon bond; they are inside or outside the ring fragment, respectively.

For $N = 32$, the barriers of every reaction were calculated for many configurations of the already assembled frame. By way of example, Fig. 2 shows the barrier height E_a for reaction 5 as a function of the total

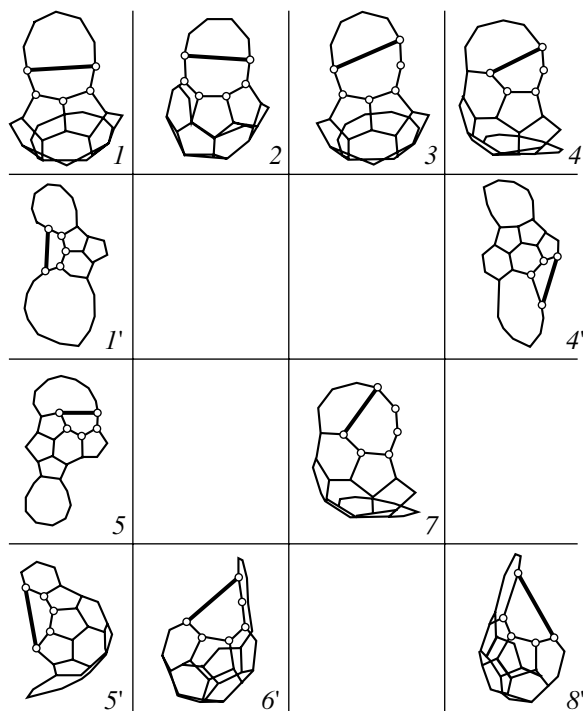


Fig. 1. Reactions taken into account in the study of the evolution of carbon clusters.

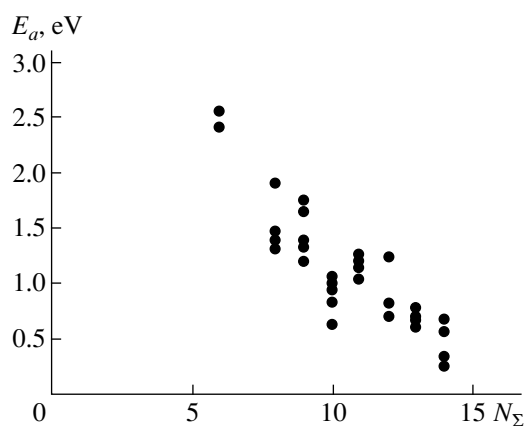


Fig. 2. Energy barriers for reaction 5.

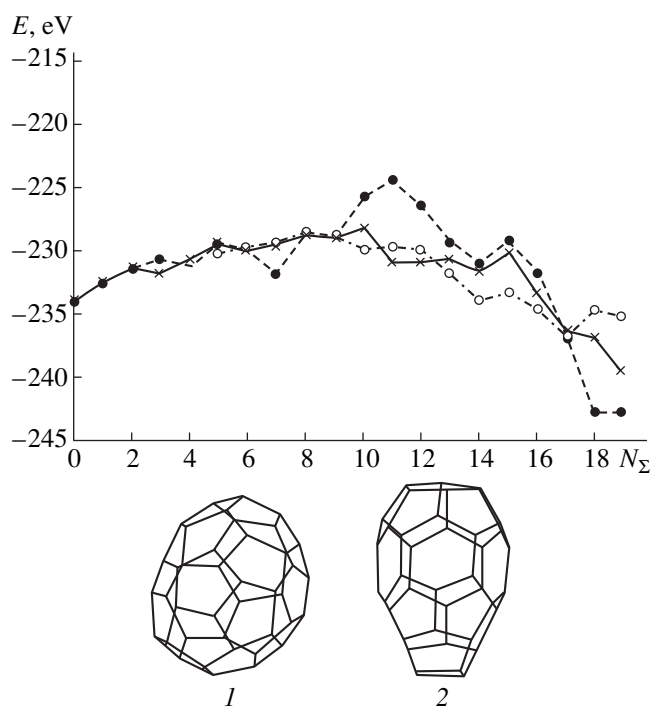


Fig. 3. Bond energy variation for some formation routes of a closed cluster of 40 atoms: (●) formation of nearly spherical fullerene 1, (○) formation of polar fullerene 2, and (×) intermediate case.

number of polygons $N_{\Sigma} = N_5 + N_6$. The scatter of the points corresponds to different evolution routes of the cluster and is a measure of the sensitivity of the reaction to a particular topology.

The probabilities of the reactions considered are essentially different. The highest barriers and the largest cup products ΔE are typical of hexagon formation reactions, in which atoms located on different sides of the ring fragment are bonded. For example, $E_a \sim 2.8$ – 3.0 eV for reaction 2, and it is 3.5 – 3.8 eV for reaction 3. This is due to the fact that the valence angles in fragments consisting of a small number of atoms are appre-

ciably less. In reaction 4, a vacant radical is additionally formed, which is a doubly connected atom with a valence angle equal to or less than that in the hexagon, $2\pi/3$ [3].

However, these reactions cannot be excluded, because the clusters would then not close. Indeed, the condition necessary for closing a cluster is as follows: at $N_5 = 6$, the number of hexagons N_6 should be at least half the number of hexagons in the assembled fullerene, $N_6^{(0)} = N/2 - 10$, and at $N_5 > 6$, the inequality $N_6/N_6^{(0)} > N_5/12$ should hold. If this condition is not met and the pentagons are too numerous, then the clusters cannot close without heptagons and octagons needed to form the constriction. This follows from the need to meet the Eulerian criterion $2N_4 + N_5 = 12 + N_7 + 2N_8$. The energy of such clusters is very large and their formation is unlikely.

Thus, if the number of pentagons in the assembling cluster is close to the critical value, the most probable of the hexagon formation reactions should take place, even though its probability is lower than that of pentagon formation.

3. In Fig. 3, the dependences are given of the bonding energy of a cluster consisting of 40 atoms ($N = 40$) on the total number of polygons in cluster N_{Σ} . Curve 1 corresponds to the case where a spherical fullerene is formed of pentagons and hexagons arranged with maximum uniformity. Curve 2 corresponds to a polar fullerene with compact groups of pentagons concentrated on the opposite poles of the cluster. Curve 3 is an intermediate configuration.

It is seen that the energy maximum for the spherical fullerene is high, because many radicals are present in the growing cluster and, in general, this entire route appears hardly probable. The route leading to the polar fullerene is also nonoptimal, because it has been found that the absolute energy minima at $N \neq 36, 48, \text{ and } 60$ are reached in an unclosed state. Between these two extremes, there are a lot of intermediate situations and it is not possible to indicate an optimum one, let alone *a priori*.

4. Thus, it appears important to calculate an ensemble of isolated clusters composed of 32 atoms and try all the possible routes to a closed (or nearly closed) structure. It was assumed that at the initial moment of time, only two-ring clusters are present. Later, some of the clusters move into energetically more advantageous states. For $N = 32$, six of such closed clusters were taken. A system of balanced equations was solved for all intermediate clusters (a total of 156). At every step, all reactions were taken into account for which the maximum and minimum probabilities differed up to two orders of magnitude at a typical temperature of 0.25 – 0.30 eV. For $N = 32$, when a fullerene should have only six hexagons, this approach is sufficient to make the control of the right ratio between N_5 and N_6 unnec-

essary. Apart from the last stage of closing, the cluster evolves only through the addition of new pentagons and hexagons.

The characteristic concentration rise time $\tau_F = 1/\omega_F$ at a temperature of 0.28 eV was found to be 0.7 s. Calculations for a few arbitrarily chosen routes leading to a closed structure yielded $\tau_F \geq 10$ s. So, taking into account the variety of routes alone reduced τ_F by more than an order of magnitude.

Of course, the practical value of this calculation is not very great, because under conditions prevailing in a real arc discharge a cluster like this is certain to trap a lot of small carbon fragments, such as C_1 , C_2 , or C_3 . However, methodologically such a calculation is important for revealing a number of the process features that support the simplified scheme [1].

(a) There is no unique well-pronounced optimal route. Variations of the concentrations of a number of clusters with $N_\Sigma = 7$ chosen fully at random are shown in Fig. 4. The curves correspond to the shortest characteristic time, which is the time needed for the establishment of a quasi-stationary flow from the initial cluster to the ultimate one in N_Σ space. It is seen that all the concentrations are of the same order of magnitude.

(b) Most ways of bonding the rings to the nucleus are realized and have approximately equal importance. Figure 5 shows variations of the concentration for three versions of the cluster with $N_\Sigma = 7$ for the same times. Curves 2 in Figs. 4 and 5 refer to the same cluster. It is again seen that all three concentrations are of the same order of magnitude.

5. Let us now estimate the characteristic time for the transformation of a two-ring cluster into a fullerene at $N \geq 40$. For the sake of simplicity, we take one chain of reactions. The reaction probabilities

$$w_{n, n+1} = (qT/h)(Q^*/Q_n)\exp[-(E^* - E_n)/T] \quad (1)$$

will be determined from the energies of the initial, final and intermediate states, E_n , E_{n+1} , and E^* , respectively. Statistical sums Q_n , Q_{n+1} , and Q^* consist of the translational, vibrational Q_{vib} , and rotational Q_{rot} terms given by the usual formulas. In an approximation of harmonic oscillators, $Q_{\text{vib}} = 1/\prod_i [1 - \exp(-h\nu_i/qT)]$, where the index i denotes the vibrational degrees of freedom $i = 1, 2, \dots, 3N - 6$. In Eq. (1) Planck's constant $h = 6.625 \times 10^{-27}$ erg s, $q = 1.6 \times 10^{-12}$ erg/eV and the temperature is in eV.

In deriving the transformation rate along the chain of reactions $1 \rightarrow 2 \rightarrow 3 \dots \rightarrow G \rightarrow F$ (where F is a fullerene and state G corresponds to the cluster preceding the energy maximum), the reverse reaction $F \rightarrow G$ is considered negligible. Then the term w_F can be estimated without solving the nonstationary problem. It will only be required to write the balance reactions for the intermediate states "k" ($k > 1$) in terms of

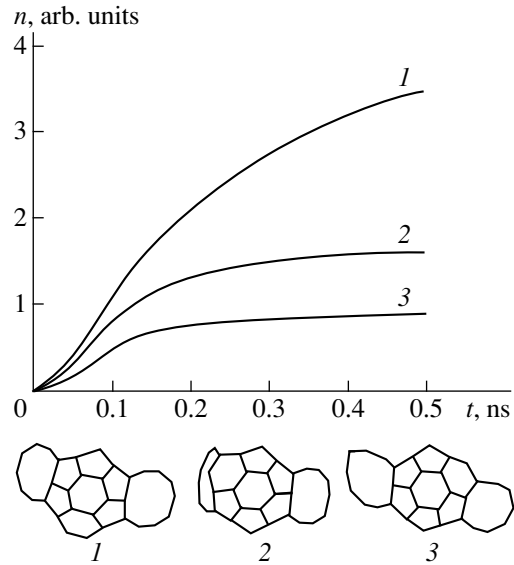


Fig. 4. Variation of the concentration with time for three arbitrarily chosen clusters (1-3).

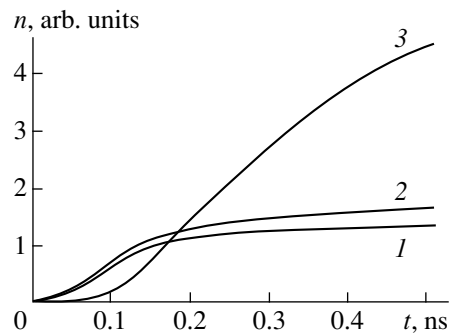


Fig. 5. Variation with time of the concentration of clusters having the same nucleus of polygons with ring fragments attached in different ways (1-3).

the population $y_k = n_k/n_k^{(0)}$ ($n_k^{(0)}$ being an equilibrium population) and set $dy_k/dt = 0$ in all equations. Together with the balance equation for n_1 ,

$$w_F = (1/n_1)dn_1/dt = w_{12}(1 - y_2). \quad (2)$$

This procedure gives the decay rate of the initial state, which is determined from a system of algebraic equations. It was found to be

$$w_F^{-1} = 1/w_{12} + 1/\Omega_2 w_{23} + 1/\Omega_3 w_{34} + \dots + 1/\Omega_G w_{GF}, \quad (3)$$

where w_{12} , w_{23} , and w_{34} are probabilities of the first and subsequent direct reactions; $\Omega_k = (Q_k/Q_1)\exp(-(E_k - E_1)/T)$.

The meaning of Eq. (3) is straightforward: the total time taken to overcome the overall potential barrier is a sum of the effective times t_k of overcoming the barriers for the component reactions, t_k being longer than the

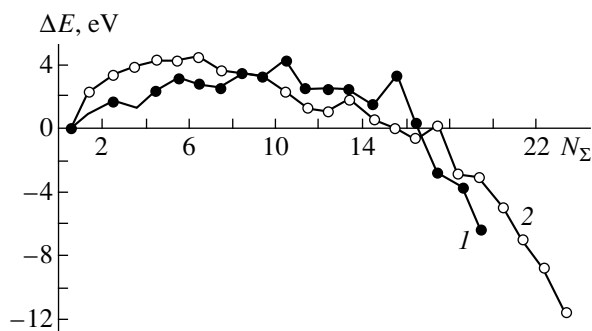


Fig. 6. Energies of two-ring clusters with different numbers of atoms: $N = 40$ (1), 48 (2).

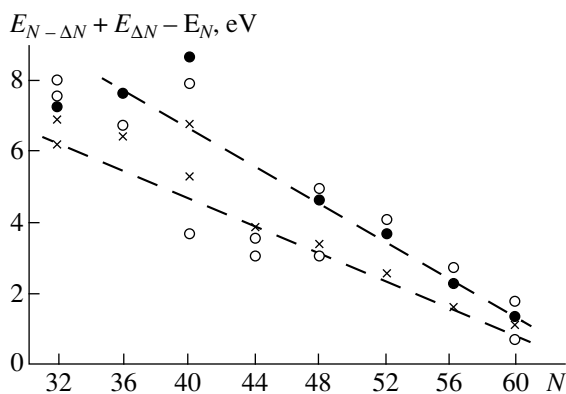


Fig. 7. Variation with N of the total energy of bonds in the case in which ΔN atoms break away in the final stage of construction.

actual reaction time by a factor of $1/\Omega_k$. Assuming that the clusters beyond the bottleneck are in equilibrium with the fullerene, formula (3) can be refined:

$$w_F^{-1} \approx 1/w_{12} + \sum_{i=2}^G 1/\Omega_i w_{i,i+1} + (n_F^\infty/n_1^0) \sum_{i=G+1}^F \Omega_{i+1}/\Omega_i w_{i,i+1}, \quad (4)$$

where $n_F^\infty = n_F(t \rightarrow \infty)$, $n_1^0 = n_1(t = 0)$.

If the energy peak is well defined (bottleneck), then

$$\omega_F \approx (qT/h)(Q^*/Q)(Q_G/Q_1) \exp(-E_Q/T), \quad (5)$$

where Q^*/Q is the typical ratio of the statistical sums of the intermediate and initial states characteristic of the new bond formation reaction and the ratio Q_G/Q_1 is of the order of $(Q_{n+1}/Q_n)^G$.

It is seen thus seen that the peak position is equally important as the barrier height.

The peak position depends on the number of atoms. Figure 6 shows energy variations in the course of construction of the fullerene from two-ring clusters with

$N = 40$ (curve 1) and $N = 48$ (curve 2). It is seen that an energy rise in the first stages of construction then gives way to an interval, in which the energy variation is weak apart from minor local peaks due to particular reactions. This is due to the fact that the gain in energy from the bonding of the atoms begins to prevail over the energy of the curved surface. However, at low N , the lowering of energy is counteracted by larger stresses in the ring fragments because of smaller angles. So, at $N = 40$, a noticeable drop in energy begins only at $N_\Sigma = 10$ – 12 (in Fig. 6 $N_\Sigma = 16$). As N is increased and the rings become larger, the energy drop starts earlier ($N_\Sigma = 6$ – 8 at $N = 48$).

Thus, with an increasing number of atoms in the cluster, its assembly rate increases. As shown above, $\tau_F > 10$ s for $N = 32$, while the characteristic time calculated taking into account the multiplicity of routes is $\tau_F \approx 0.7$ s. For $N = 40$, we have $(Q_G/Q_1) \sim 10^{-4}$ for $G \sim 8$ and τ_F is equal to 10^{-2} s. For $N = 48$, the reduction of the statistical sum in the maximum in steps 5–6 does not exceed 10^3 and the estimated value of τ_F is about 10^{-3} . Extrapolation to 60 atoms (using data for $N = 24, 32, 40, 44$, and 48) taking into account the multiplicity of routes is $\tau_F \approx 10^{-4}$ – 10^{-5} s.

SPECIFIC FEATURES OF THE FINAL STAGE OF CLOSING OF THE CARBON CLUSTER INTO A FULLERENE

Consider in more detail the final reactions, which close the cluster and give it its ultimate appearance. It is evident that with arbitrary ring sizes in the initial two-ring cluster the precise closing is a rare event (internal transformations of the Stone–Wales type in the assembled part of the clusters do not affect the surface edge topology and ring fragments).

Typical structural defects appearing in the final assembly steps are small ring fragments, which must be discarded. In Fig. 7, a dependence is shown of the variation of the total energy needed to break the bonds of ΔN atoms on the number of atoms in the cluster N . The cluster topology was assumed such that exactly this number of atoms ΔN was superfluous regarding the closing into a fullerene. In Fig. 7, solid circles correspond to $\Delta N = 2$ and crosses and open circles to $\Delta N = 4$. The scatter in the data is due to the fact that for $N = \text{const}$ different clusters can have the same size N .

It is seen that $\Delta E > 0$ as long as N is below 62–65 and the discarding reaction is never favored. On the other hand, at $36 < N < 60$, there is a wide class of nearly flawless clusters with a few extra atoms whose energy is much lower than that of the initial two-ring clusters. However, it hard for them to get rid of the extra atoms, whereas, due to reaction-active doubly bonded atoms, they should be easily incorporated into the soot particles. This the reason why the amount of fullerenes with $N < 60$ is always small, even after annealing for an

unlimited time, despite their being more favored than the initial products in some cases.

THE RATE OF TRANSFORMATION OF CARBON CLUSTERS INTO FULLERENES: THE ABRIDGED DESCRIPTION SCHEME

1. The abridged scheme proposed in [1] is efficient when the number of atoms in a cluster is large and a direct consideration of all the possible intermediate clusters is cumbersome. However, as shown above, at $N > 48-50$, several assembly routes are available with well-defined energy peaks for $N_{\Sigma} = 5-7$, and for determining the characteristic transformation time, the details of further assembly are of little consequence. It is enough to have information on clusters with $N_{\Sigma} < 5-7$ obtainable by extrapolation of the data for $N = 24, 32, 40,$ and 44 , or to generalize the calculation for $N = 32$ and not to use the abridged description scheme. However, by using this scheme it is easy to take into account an important class of reactions which have not been considered even for $N = 32$, namely, the reactions in which new pentagons and hexagons are formed in a process involving short-moving triangles and tetragons and atoms that stick out. Therefore, it will be useful to discuss this scheme again.

2. In the analysis above, it has been demonstrated that the evolution of the clusters is in fact a progression through a table (N_5, N_6) referred to in [1] as an activity table. In Fig. 8, an activity table illustrating the evolution of a cluster with $N = 44$ is presented. The area below straight line *I* is a shell region *I* containing clusters with a surplus of hexagons. This region is practically unpopulated, because the prevailing progression direction of the clusters is different, leading towards the region of excess pentagons. The area *II* above line 2, the constriction region, is also poorly populated. As already noted, a cluster coming to the boundary of this region is either reflected with the disintegration of a pentagon or a new hexagon is formed. Therefore, the progression of a cluster towards becoming a fullerene occurs within the region bounded by lines *I* and 2. For every N value, the cluster progression was followed up to the assembly stage approximately corresponding to the energy peak and influenced by the number of atoms (the fullerene diagonal *D* in Fig. 8). It was assumed that all clusters belonging to the diagonal *D* transform into fullerenes with equal probability. This is true of a fraction of clusters with $N < 60$ and of most of the clusters with $N > 62-65$ (because of the loss of 2-5 extra atoms).

3. Progression of a cluster is determined by its energy. Let the cluster energy be

$$E = E_{\text{frame}} - E_r^{(5)}(R_5 - \rho_5) - E_r^{(6)}(R_6 - \rho_6) + E_{\text{ring}} + E(\gamma), \quad (6)$$

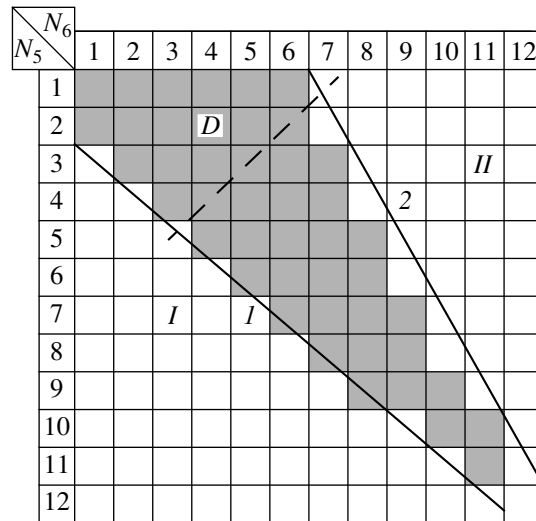


Fig. 8. Cluster activity table for $N = 44$.

where $E_{\text{frame}} - E_r^{(5)}(R_5 - \rho_5) - E_r^{(6)}(R_6 - \rho_6)$ is the frame energy minus the energy of radicals at the edge; $\rho_{5,6}$ and $R_{5,6}$ is the number of actual and potential pentagonal and hexagonal radicals; $R = R_5 + R_6 = 2m + \rho_5 + \rho_6$ is the total number of radicals; m is the number of ring fragments; E_{ring} is the total bonding energy of atoms in the ring fragments; and $E(\gamma)$ is the distortion energy at the frame-ring fragment junction.

The frame energy is localized in the bonds between the frame atoms. It depends on the shape and can be characterized by the number of potential pentagonal (R_5) and hexagonal (R_6) radicals. As mentioned above, assembled fullerenes can have two limiting configurations: nearly spherical or polar. In the first case, the energy of the curved surface is a minimum, and in the second, it is a maximum. The same properties are displayed by the frames in the process of assembly.

In a spherical frame, the numbers of radicals $R_5^{(0)}$, $R_6^{(0)}$ and the edge energy are the highest. Energy $E^{(0)}$ for this frame is expressed through the energy of the fullerene, $E_F(N_5^{(0)}, N_6^{(0)})$ as

$$E^{(0)} = \frac{B}{3N_6^{(0)} + 30} E_F(N_5^{(0)}, N_6^{(0)}),$$

where B is the number of bonds in the frame.

The quantity B is determined from topological relationships [1]:

$$\begin{aligned} N_s &= -2 + 2(N_5 + N_6) + R, & N_s &= N_i + N_A + R, \\ N_i + N_s &= 3N_5 + 4N_6 + 2, \\ B &= N_A + 2N_5 + 3N_6 + 3. \end{aligned}$$

Here, N_s is the total number of atoms in the frame; R is the total number of potential radicals; N_i is the number

of internal atoms of the frame; and N_A is the number of triply bonded atoms at the frame edge. We get $B = 3(N_5 + N_6) + R$. In a polar frame, the energy of distorted valence angles is the highest, while the number of radicals $R_5^{(1)}$, $R_6^{(1)}$ and the edge energy are the lowest. The cluster with a minimum elongation represents a structure of six pentagons and the central hexagon with a growth of a "nanotube" of hexagons. At $N_5 = 6$ there can be any number of hexagons N_6 , whereas at $N_5 > 6$ it is $N_6 = N_6^{(0)} - 1 = 6k$, where k is an integer. The number of radicals $R_6^{(1)}$ at $N_5 = 6$ is given by a periodic function of N_6 at $N_6 = 7, 13, 20, \dots$, $R_6^{(1)} = 6$ and for other N_6 values $R_6^{(1)} = 7$. The number $R_5^{(1)}$ is zero at $N_5 = 6$ and is uniquely defined for other tube closing reactions at $N_5 > 6$. In contrast to the spherical fullerene, the pair of numbers (N_5, N_6) uniquely defines the shape of the cluster for the nanotube so that it can be explicitly constructed, and its energy is $E^{(1)}$. The position of any cluster between these extreme cases is determined by the numbers of radicals R_5 , R_6 in the ranges between $R_5^{(0)}$ and $R_5^{(1)}$, $R_6^{(0)}$ and $R_6^{(1)}$, and the frame energy by linear extrapolation between the energies of the spherical cluster $E^{(0)}$ and the nanotube $E^{(1)}$:

$$E_{\text{frame}} = E^{(0)} + \frac{E^{(1)} - E^{(0)}}{R_5^{(1)} - R_5^{(0)}}(R_5 - R_5^{(0)}) + \frac{E^{(1)} - E^{(0)}}{R_6^{(1)} - R_6^{(0)}}(R_6 - R_6^{(0)}).$$

The energy of ring fragments E_{ring} given by the second term in formula (6) was determined from the following considerations. In every fragment, there are $N_R = (N - N_s)/m$ atoms and $(N_R - 1)$ bonds, each binding two atoms of the fragment. A closed ring will have a number of bonds larger by Λ ($\Lambda \sim 2$). Thus, in a closed ring the total number of bonds is $(N_R + \Lambda + 1)$, giving the total energy of the ring fragments as

$$E_{\text{ring}} = m \frac{N_R - 1}{N_R + \Lambda + 1} E_{\text{ring}}^{N_R + \Lambda + 1},$$

where $E_{\text{ring}}^{N_R + \Lambda + 1}$ is the energy of a ring of $N_R + \Lambda + 1$ atoms.

The last energy term in Eq. (6) is the energy localized in the vicinity of the frame–ring fragment bonds. This energy is minimal if the number of atoms in the rings is exactly that needed for constructing the fullerene and depends on a mismatch parameter

$$\gamma = \frac{N - N_s}{N^{(0)} - N_s},$$

where $N^{(0)} = 2N_6^{(0)} + 20$.

The energy $E(\gamma)$ was expressed as $E(\gamma) = C_0 + C_1(N_5/12)^2\gamma$, where C_0 and C_1 are constants, C_0 being close to the atomic bond energy.

4. For calculating the energy of any cluster using Eq. (6), in addition to numbers N_5 , N_6 , the numbers of three radicals are required— ρ_5 , ρ_6 , and R_5 —because the energy at the edge is determined by the number of actual radicals ρ_5 , ρ_6 , while the energy of the constructed surface depends on the number of potential radicals R_5 and R_6 . For $N = 60$, it is necessary to consider about 20 triplets of numbers (ρ_5 , ρ_6 , and R_5), but for a cluster with a constant number of atoms isolated from the contact with small carbon fragments, this problem is quite solvable.

The parameters in the expression for the energy terms were fitted using the least squares method and data for approximately 40 clusters. The energies of all clusters calculated with the obtained parameters differed from the exact values (calculated by AM1) by not more than 1 eV.

5. The performed calculations and estimates, on one hand, confirm the qualitative assumptions on which the scheme of [1] is based, and, on the other hand, suggest a need for some modifications. So, it was erroneously assumed in [1] that the configurations corresponding to bottlenecks, which have a greater number of bonds, have lesser energy than the initial cluster, and that the number of such clusters is large. Quantum mechanical calculations demonstrate that the opposite is true. Region II in Fig. 8 at $N_5 > 6$ is a natural area of attraction for the clusters and, simultaneously, a dead-end region with a very high cluster energy. As noted above, an ultimate closed cluster, if favored, skips this region at the expense of the reaction of hexagon formation at a certain assembly stage (even if the probability of these reactions is 1–2 orders lower than that of pentagon formation). In the energy diagram depicting the evolution of a real cluster, these reactions introduce characteristic peaks (Fig. 6). As a rule, the cluster energy diminishes between the peaks, because reaction 5 (Fig. 1) leads to the disappearance of a radical. The peak positions for different routes are variable.

If all the clusters with specified N_5 , N_6 , and R are treated as one object, the consequence will not be a lasting one: at the next step, an average cluster almost certainly will follow a reaction with a lower barrier and more often than not it will be the pentagon formation reaction. If the cluster already has many pentagons, it will shoot out to the dead-end part of the activity table. It can get back to the main active region only by breaking bonds. It was considered, therefore, that when an average cluster with $N_5 > 6$ is found in the last active cell adjacent to region II, a hexagon will form by the most probable reaction of this type 8 or 2' (Fig. 1). If these reactions are not possible for topological reasons, then reactions 2, 3, or 3' will take place. The most prob-

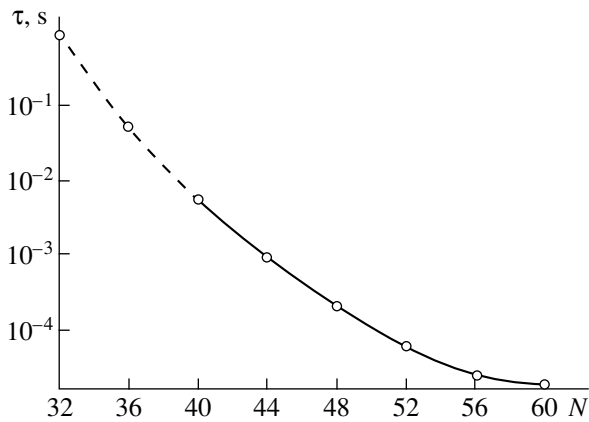


Fig. 9. Characteristic time of transformation into fullerene vs. the number of atoms in the initial cluster.

able reaction for the same topologies is reaction 5; the other reactions are frozen.

For the active part of the table (highlighted cells in Fig. 8) at $N_5 > 6$, the principle of conditional probability was also applied: only the most probable reactions of hexagon formation occur, first reaction 8, then 2, and only when the required surface configurations are lacking do other reactions become possible. Of course, this modification is applicable only to fairly small clusters ($N < 44-46$). At large N , the clusters with $N_5 > 6$ are found on the falling branch of the energy diagram (i.e., below the fullerene diagonal) and need not be considered.

6. In Fig. 9, the major result of the calculations is presented, which is the dependence of the transformation time τ_F on N . The starting objects in the calculation scheme were clusters with a central hexagon obtained as a result of a ring fragment being divided in two and the subsequent formation of a pentagon ($N_5 = 1$, $N_6 = 1$, $m = 3$). The temperature was assumed constant at 0.28 eV. A significant reduction of the assembly time with an increasing number of atoms is seen. The qualitative discussion of this result was given above.

Certainly, the long times obtained in this isolated scheme at $N < 40$ are not obtainable under gas discharge conditions where the cluster will have high chances of capturing small fragments like C_1 , C_2 , or C_3 . However, the derived relationships for the transformation of multiple-ring clusters into fullerenes open new aspects of the dynamics of their formation in an arc discharge.

CONCLUSION

In subsequent computational work on fullerene formation in gas discharge plasma, the following problems must be solved. It is necessary to develop a theory of arc discharge between graphite electrodes in inert gases at low pressures. In gas discharge physics, the arc

in which the plasma forming material is the product of anode erosion practically has not been considered. This theory should not be too complicated, because it is intended as a component of the general calculation procedure. The theory should supply not only the initial conditions of the plasma jet exiting the interelectrode gap but also the radial concentration of the charged particles, because, as shown in computer simulations, the charge residing on a cluster can change its dynamics [4].

It is necessary to clarify the role of the buffer gas in the fullerene formation process. From experiments, it is known that the best inert gas for the purpose is helium, and with increasing molecular weight of the inert gas, the fullerene formation efficiency decreases [5]. This experimental observation should be given a theoretical explanation.

The role of small clusters in the formation of polycyclic clusters and the building of fullerenes should be identified. There is thermodynamic evidence that at the temperature optimal for preparation of fullerenes (0.2–0.3 eV) and the typical total carbon concentration, the equilibrium concentration of carbon atoms and molecules is extremely low. In this case the assembly of fullerenes is possible only by internal reactions in the already formed multiple-ring clusters and only if these clusters are in an area of optimum temperature for a sufficient amount of time. It is possible, however, that these processes are all nonequilibrium and that the optimum conditions for fullerene formation are different.

ACKNOWLEDGMENTS

The authors are grateful to R.A. Suris and participants of his sector's seminar for useful discussions of the study results.

This study was carried out in the framework of the Russian program "Fullerenes and Atomic Clusters" and supported in part by the Russian Foundation for Basic Research, project no. 00-02-16928.

REFERENCES

1. N. I. Alekseev and G. A. Dyuzhev, *Zh. Tekh. Fiz.* **69** (9), 104 (1999) [*Tech. Phys.* **44**, 1093 (1999)].
2. N. I. Alekseev and G. A. Dyuzhev, *Zh. Tekh. Fiz.* **71** (5), 67 (2001) [*Tech. Phys.* **46**, 573 (2001)].
3. V. A. Schweigert, A. L. Alexandrov, Y. N. Morokov, and V. I. Bedanov, *Chem. Phys. Lett.* **235**, 221 (1995).
4. N. I. Alekseev and G. A. Dyuzhev, *Zh. Tekh. Fiz.* **69** (12), 42 (1999) [*Tech. Phys.* **44**, 1435 (1999)].
5. D. V. Afanas'ev, A. A. Bogdanov, G. A. Dyuzhev, and A. A. Kruglikov, *Zh. Tekh. Fiz.* **67** (2), 125 (1997) [*Tech. Phys.* **42**, 234 (1997)].

Translated by B. Kalinin

Thermal Stability and Radiation Hardness of SiC-Based Schottky-Barrier Diodes

A. V. Afanas'ev, V. A. Il'in, I. G. Kazarin, and A. A. Petrov

St. Petersburg State University of Electrical Engineering, St. Petersburg, 197376 Russia

e-mail: luch@cm.etu.spb.ru

Received July 21, 2000

Abstract—Pt/W/Cr/SiC Schottky-barrier diodes that retain good electrophysical parameters up to 450°C are studied. With the Auger electron spectroscopy (AES) method, it is shown that the thermal stability is provided by using a multilayer metal composition that ensures the metal/SiC interface stability. The surface-barrier structures obtained are tested for radiation hardness. They are irradiated by fast neutrons with a fluence of 4.42×10^{15} n/cm² and attendant γ radiation with a dose of 8.67×10^5 R in the concentration range of $N_d - N_a = 10^{16} - 5 \times 10^{17}$ cm⁻³. Irreversible modifications of the structures at $N_d - N_a \leq 8 \times 10^{16}$ cm⁻³ are found. The degradation of the parameters is inversely proportional to the doping level. © 2001 MAIK "Nauka/Interperiodica".

Today, silicon carbide devices are viewed as basic components of extreme electronics. Emphasis in this area is on contact systems, which must retain good electrical performance at high temperatures and when subjected to high-intensity ionizing radiation.

The effect of thermal annealing on the properties of metal/SiC (Me/SiC) surface-barrier structures was studied in [1–3], where Au, Pt, Hf, Ni, Ti, Cr, etc. were employed as barrier-forming metals. It was found that the electrical performance of the diodes irreversibly degrades when the devices are subjected to high temperatures (300–700°C). This shrinks the range of operating temperatures. For example, data on the efficiency of Au/SiC-based Schottky diodes are available only for temperatures of 300°C or below [4].

It is obvious that the thermal stability of the contacts is governed largely by the properties of the Me/SiC interface. Ways of improving the stability of the transition layer have been considered in a number of works. In particular, it was shown [5] that metals, such as Cr, W, Ta, and Mo, as well as associated multilayer compositions, are promising materials for high-temperature rectifying contacts. Such a conclusion is based on AES profiles taken from the Me/SiC interface during high-temperature treatment of the contacts. The possibility of Cr/SiC diodes being used as temperature-sensitive transducers with conversion linearity up to 400°C was demonstrated in [6].

It was the aim of this work to create SiC-based Schottky diodes whose current–voltage characteristic (CVC) remains stable up to 450°C, to discover factors responsible for their thermal stability, and to estimate the radiation hardness of the devices.

In experiments, we used *n*-SiC single crystals of polytypic modification 6H that were covered by epitaxial films ($N_d - N_a = 10^{16} - 10^{17}$ cm⁻³) grown by vacuum

sublimation on the Si(0001) face [7]. Prior to metal application, the surface was first treated by the standard technology and then (immediately before the deposition) finished by rapid argon bombardment (the energy ≈ 400 eV). Ohmic contacts were formed by applying a 0.3- μ m-thick nickel film on the *C* face of SiC by magnetron sputtering with subsequent annealing in a vacuum at 1000°C for 3 min. A Cr/W/Pt multilayer system was applied in a single vacuum cycle by magnetron sputtering of the metals onto Si wafers preheated to 200°C. The metal films were 0.1 to 0.15 μ m thick, and the surface area of the contacts was 5×10^{-4} cm². The electrophysical parameters of the diodes were stabilized by heat treatment in a vacuum at 500°C for 10 min. The devices thus prepared were intended to withstand temperatures up to 450°C.

The Schottky barrier heights derived from capacitance–voltage (*C*–*V*) curves taken from the heat-treated structures were found to vary between 0.9 and 1.05 eV, which is 0.2–0.25 eV lower than in untreated devices. This can be related to changes in the chemical composition and structure of the metal/semiconductor interface.

The rectification ratios of CVCs of the structures (Fig. 1) were $K_r = 10^7$ at 20°C and $K_r = 10^2$ at 400°C for $U_{\pm} = 1$ V. The nonideality ratio of the room-temperature CVC is $n \approx 1.12$ and drops with increasing temperature (inset in Fig. 1).

To study the thermal stability of the surface-barrier structures, we annealed them in a vacuum at 450°C for 50 h. As follows from the room-temperature CVCs (Fig. 1), the heat treatment does not change the rectifying properties of the structures. Moreover, the electrical performance of the devices does not degrade even after subsequent 10-min annealing at 700°C. AES concentration profiles from the Pt/W/Cr/SiC system (Fig. 2)

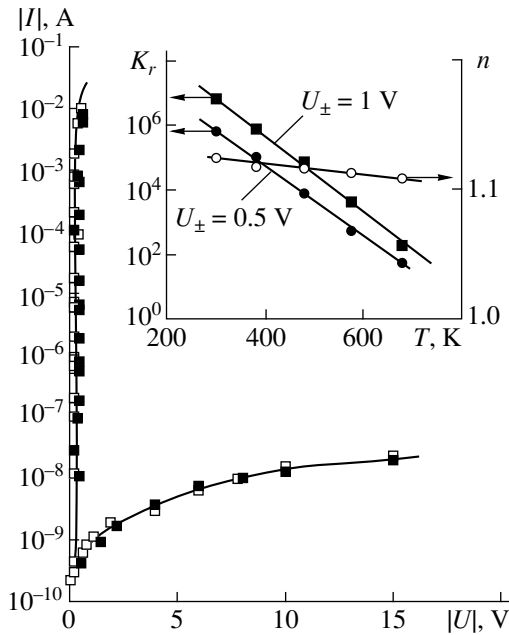


Fig. 1. CVCs of the Pt/W/Cr/SiC surface-barrier structures. □, as-prepared structure; ■, after annealing at 450°C for 50 h. The inset shows the temperature dependences of the rectification ratio K_r , and nonideality coefficient n .

show a slight modification of the Cr/SiC interface, indicating the efficiency of the tungsten barrier, which prevents interdiffusion of platinum and chromium.

The heat treatment at the stage of fabricating the diode structure and also the long-term stabilizing annealing form a transition layer of chromium carbides (hatched regions in Figs. 2a, 2c). The chromium carbides were identified from the fine structure of the KVV carbon lines (Fig. 2b). The silicon lines remain unchanged. This indicates that diffusion, causing the chromium carbides to appear, is of an activation character. The resulting carbides suppress the growth of silicide phases. Chromium carbides are known to have a high thermal stability [8]. Therefore, the transition layer with thickness of about 30 nm stabilizes the Me/SiC interface and acts as an efficient diffusion barrier at temperatures near 450°C.

As was noted, devices of extreme electronics may simultaneously be subjected to high temperatures and ionizing radiation. Therefore, of interest was to study the operation of the high-temperature Schottky diodes exposed to high-dose radiation. Obviously, radiation-induced defects, modifying the properties of the semiconductor, may severely affect the CVC of the device. A feature of defect formation in 6H SiC upon irradiation is that resulting defect centers are very deep and temperatures necessary to anneal out the defects are very high ($T \geq 1000^\circ\text{C}$). Therefore, the properties of the diodes cannot be restored at operating temperatures (400–600°C) [9]. In view of this circumstance, we tried

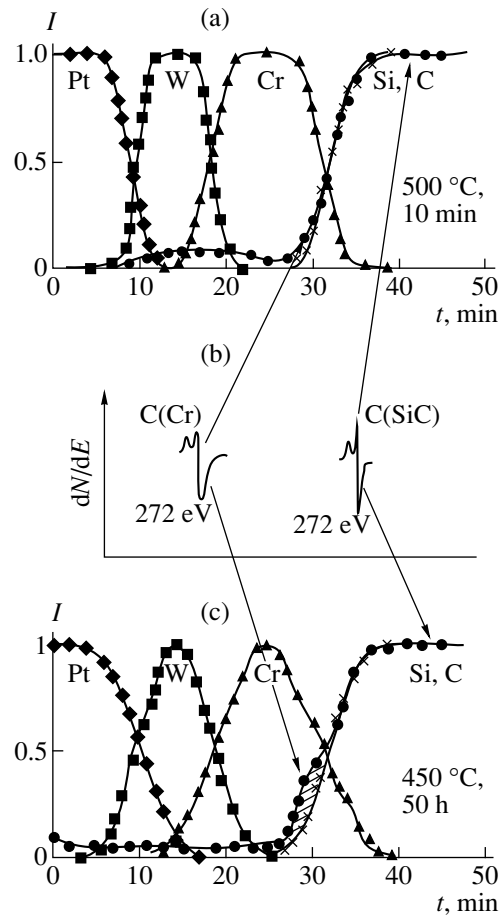


Fig. 2. (a, c) Depth profiles of the elements for the Pt/W/Cr/SiC system and (b) Auger lines of carbon (KVV) in chromium carbide and silicon carbide.

to trace a correlation between the radiation hardness of the Schottky diodes and the initial doping level in SiC.

Radiation hardness tests were performed on Schottky diodes with dopant concentrations of 10^{16} (type 1), 3×10^{16} (type 2), and 10^{17} cm^{-3} (type 3) in the epitaxial films. Irradiation was carried out with a BARS pulsed reactor (SPELS Co., Moscow). The devices to be tested were placed in the central path of the reactor at a site where the particle fluence was calibrated to maximum. The neutron energies varied between 0.1 and 3.0 MeV. Along with neutron irradiation, the devices were exposed to γ radiation. The total fluence of neutrons was $4.42 \times 10^{15} \text{ n/cm}^2$, and the dose of attendant γ radiation was $D_\gamma = 8.67 \times 10^5 \text{ R}$. After aging for a time, the devices were annealed at 600°C for 10 min, their electrophysical parameters were studied with DLTS, and C–V and I–V curves were recorded.

It was found that the CVCs of the type-3 devices remain unchanged (Fig. 3a), while those of the type-1 and type-2 devices change irreversibly (Fig. 3b). The reason is a sharp increase in the series resistance because of material compensation by radiation-induced

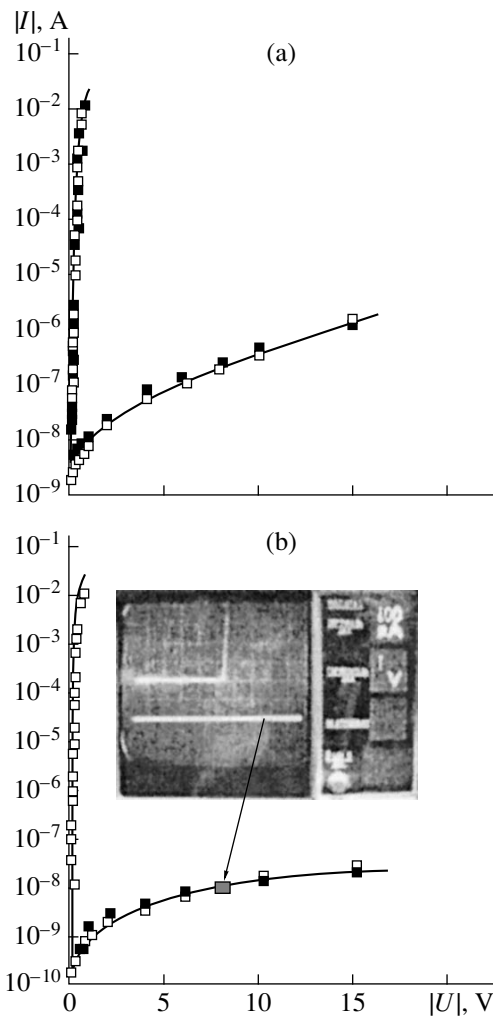


Fig. 3. CVCs of the Schottky diodes (\square) before and (\blacksquare) after fast neutron irradiation for the initial value of $N_d - N_a =$ (a) 10^{17} cm^{-3} and (b) $(1-3) \times 10^{16} \text{ cm}^{-3}$.

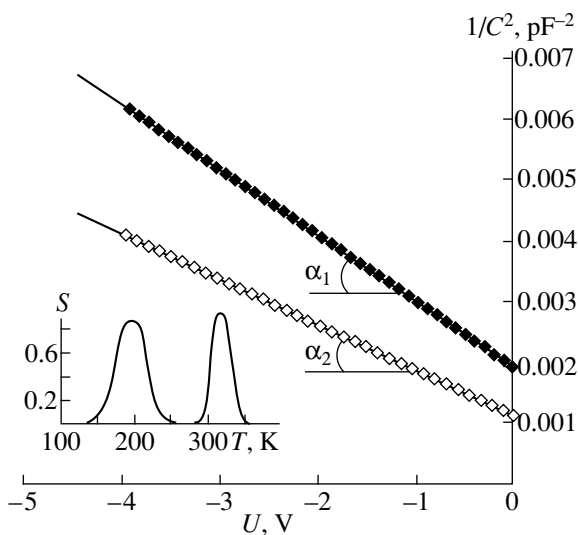


Fig. 4. C - V curves for the type-3 structures (\blacklozenge) before and (\diamond) after irradiation. The inset shows the DLTS spectra in the 6H SiC diode structure after irradiation.

defects. Annealing at 600°C does not restore the characteristics even partially.

The degree of compensation was estimated from the C - V curves (Fig. 4). In the irradiated type-3 structures, the concentration of uncompensated donors was calculated by the formula

$$(N_d - N_a)' = (N_d - N_a) \tan(\alpha_1) / \tan(\alpha_2) \quad (1)$$

and equaled $(N_d - N_a)' = 7 \times 10^{16} \text{ cm}^{-3}$. The concentration of compensating centers was estimated at $3 \times 10^{16} \text{ cm}^{-3}$.

As follows from our DLTS data (inset in Fig. 4) and data for deep radiation-induced centers in 6H SiC [9], the compensation effect after irradiation and annealing is due to deep centers with activation energies of 0.35 and 0.6–0.8 eV (reckoned from the conduction band bottom).

Thus, Pt/W/Cr/SiC Schottky diodes studied in this work retain good electrophysical parameters for 50 h at 450°C . The thermal stability is provided by using the multilayer metal composition that ensures the stability of the Me/SiC interface. The radiation hardness of the structures depends on the initial doping level. When the structures (with dopant concentrations in the range of $N_d - N_a = 10^{16} - 5 \times 10^{17} \text{ cm}^{-3}$) are simultaneously exposed to fast neutrons (the fluence $4.42 \times 10^{15} \text{ n/cm}^2$) and γ radiation (the dose $8.67 \times 10^5 \text{ R}$), the barrier properties irreversibly degrade at $N_d - N_a \leq 8 \times 10^{16} \text{ cm}^{-3}$. The degradation is inversely proportional to the doping level.

REFERENCES

1. J. R. Waldrop and R. W. Grant, *Appl. Phys. Lett.* **62**, 2685 (1993).
2. R. G. Verenchikova and V. I. Sankin, *Fiz. Tekh. Poluprovodn.* (Leningrad) **22**, 1692 (1988) [*Sov. Phys. Semicond.* **22**, 1067 (1988)].
3. L. M. Porter and R. F. Davis, in *Proceedings of the Second International High Temperature Conference, Charlotte, 1994*, Vol. 1, p. XIII-3.
4. M. M. Anikin, A. N. Andreev, A. A. Lebedev, *et al.*, *Fiz. Tekh. Poluprovodn.* (Leningrad) **25**, 328 (1991) [*Sov. Phys. Semicond.* **25**, 198 (1991)].
5. V. S. Ballandovich, V. V. Luchinin, A. A. Petrov, *et al.*, *Peterburg. Zh. Élektron.* **2**, 47 (1994).
6. V. A. Il'in, V. V. Luchinin, A. A. Petrov, *et al.*, *Mater. Sci. Eng., B* **46**, 383 (1997).
7. A. A. Lebedev, D. V. Davydov, N. S. Savkina, *et al.*, *Mater. Sci. Eng., B* **61-62**, 165 (1999).
8. V. A. Pashkin and A. M. Shiryaev, *Poverkhnost'*, No. 3, 106 (1992).
9. V. A. Il'in and V. S. Ballandovich, *Defect Diffus. Forum* **103-105**, 633 (1993).

Translated by V. Isaakyan

High-Sensitivity Moire Flaw Detection for Comparison of Periodic Composite Structures

A. M. Lyalikov

Kupala State University, Grodno, 230023 Belarus

Received March 27, 2000; in final form, July 25, 2000.

Abstract—For the moire visualization of macroscopic defects in composite periodic structures by comparing a structure under study with the reference, it is proposed that rerecorded moire patterns from each of the substructures be used as reference ones. The use of the rerecorded moire patterns obtained by filtering complex conjugate orders of diffraction improves the sensitivity of the method. © 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

For the visualization of macroscopic defects in objects with a distinct periodic structure, the moire methods are preferable [1, 2]. They are simple and have nearly the same accuracy as the interference methods [3, 4]. If a macrodefect of the periodic structure shows up as a deviation of the structure period from the mean value, the sensitivities of both methods are nearly equal but the moire methods do not require high-coherence light sources. If the object represents the superposition of several periodic substructures, one usually needs to visualize macrodefects in each of the substructures separately. In this case, preference is given to the interference methods, since the moire methods necessitate additional spatial filtering. The sensitivity of the interference methods can be improved by using higher orders of diffraction [5, 6]. In the moire methods, to do this would require filtering higher diffraction orders or using reference periodic transparencies of a higher carrier frequency [2, 7, 8].

DESCRIPTION OF THE METHOD

The scope of this work is the moire visualization of macroscopic defects in composite periodic structures by comparing the moire pattern of the overall structure with the rerecorded substructure pattern used as the reference ones. It is also proposed to rerecord the substructure patterns by separating complex conjugate diffraction orders with spatial filtering. If an entire composite structure to be studied consists of N elementary periodic structures (substructures), N separate moire patterns must be rerecorded. The use of the complex conjugate orders allows a $2n$ -fold decrease in the period of the substructure rerecorded, where n is the number of the diffraction order filtered.

The amplitude transmittance of a composite periodic structure consisting of N elementary structures can

be represented as the product of the transmittances of the elementary structures:

$$\tau = \tau_1 \tau_2 \times \dots \times \tau_N. \quad (1)$$

In (1), the amplitude transmittances of the i th elementary structure in the reference composite structure and in the composite structure under study are represented as the Fourier expansions

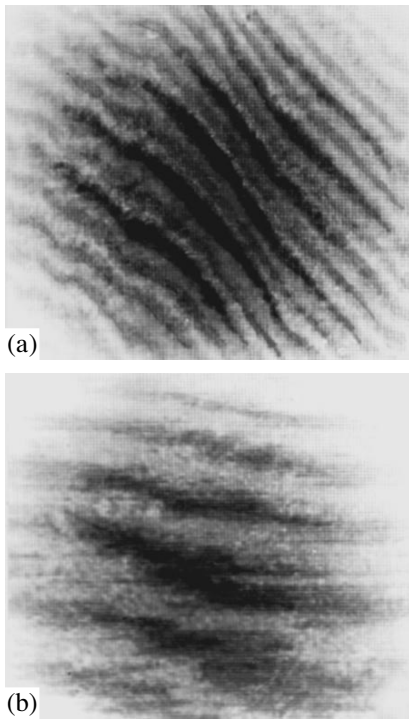
$$\tau_1 = \tau_0 + \sum_{m=1}^{+\infty} a_{im} \cos[m(\mathbf{K}_i \mathbf{r} + \varphi_i + \varepsilon_i)], \quad (2)$$

$$\tau'_1 = \tau'_0 + \sum_{m=1}^{+\infty} a'_{im} \cos[m(\mathbf{K}_i \mathbf{r} + \varphi_i)], \quad (3)$$

where τ_0 and τ'_0 are constants, a_{im} and a'_{im} are coefficients, \mathbf{K}_i is the vector of the i th structure that lies in its plane and is orthogonal to its grooves, \mathbf{r} is the radius vector of a point, φ_i is a function describing the deviation of the period of the i th elementary reference structure from the mean value $T_i = 2\pi/|\mathbf{K}_i|$, and ε_i is a function characterizing macrodefects in the structure. Rerecording of the reference composite structure to obtain separate patterns can be done by well-known methods of high-sensitivity holographic interferometry [9–11]. One-beam approach seems to be optimal for our objects, since it allows rerecording by incoherent light [9]. The magnification of the optical systems must equal unity in this process. If the n th diffraction orders are separated out and linear recording conditions are met, the amplitude transmittance of the image of the i th substructure rerecorded is represented as

$$\tau''_i \sim 1 + [2n(\mathbf{K}_i \mathbf{r} + \varphi_i)]. \quad (4)$$

A comparison of (4) and (3) shows that the period of the fringes in the pattern rerecorded is $2n$ times smaller than the period of the structure under study.



Moiré patterns visualizing macroscopic defects in the gratings with (a) vertical and (b) horizontal grooves.

The moiré visualization of macroscopic defects in an elementary structure implies the precise superposition of its rerecorded pattern on the composite structure. The resulting amplitude transmittance of the combined structure is given by the product $\tau_i''\tau$. Assume that the fringes of any of the elementary structures besides the i th one have an orientation differing from that of the fringes of the pattern superimposed. In this case, the product $\tau_i''\tau_i$ would suffice to analyze the moiré fringes. Going to the cosine of half-angle in (4) yields

$$\tau_i''\tau_i \sim \cos^2[n(\mathbf{K}_i\mathbf{r} + \varphi_i)] \times \left\{ \tau_0 + \sum_{m=1}^{+\infty} a_{im} \cos[m(\mathbf{K}_i\mathbf{r} + \varphi_i + \varepsilon_i)] \right\}. \quad (5)$$

The moiré fringes characterizing the low-frequency modulation of the elementary structure are described by the product of the first cosine by the m th term of the series, which meets the condition $m = 2n$. All other terms in (5) do not affect the moiré fringes. Representing the m th term of the Fourier series through the cosine of half-angle and retaining the necessary terms, we arrive at

$$2a_{im} \cos^2[n(\mathbf{K}_i\mathbf{r} + \varphi_i)] \cos^2\left[\frac{m}{2}(\mathbf{K}_i\mathbf{r} + \varphi_i + \varepsilon_i)\right]$$

$$= \frac{a_{im}}{2} \left\{ \cos^2\left[2n\left(\mathbf{K}_i\mathbf{r} + \varphi_i + \frac{\varepsilon_i}{2}\right)\right] + \cos^2(n\varepsilon_i) + 2 \cos\left[2n\left(\mathbf{K}_i\mathbf{r} + \varphi_i + \frac{\varepsilon_i}{2}\right)\right] \cos(n\varepsilon_i) \right\}. \quad (6)$$

The moiré fringes corresponding to the low-frequency modulation of a high-frequency elementary structure are described by the east term in (6):

$$a_{im} \cos\left[2n\left(\mathbf{K}_i\mathbf{r} + \varphi_i + \frac{\varepsilon_i}{2}\right)\right] \cos(n\varepsilon_i). \quad (7)$$

If pattern (4) is recorded under nonlinear conditions, its amplitude transmittance will be other than the cosine function and contain higher harmonics. Eventually, the superposition of such a pattern on the structure under study would complicate the moiré pattern.

Setting the last cosine in (7) equal to zero, we find the centers of the moiré fringes:

$$\varepsilon_i = (2l + 1)\frac{\pi}{2n}; \quad l = 0, 1, 2, \dots \quad (8)$$

The moiré fringes are observed in both reflected and transmitted light.

If the pattern rerecorded is precisely aligned with the composite structure under study, moiré pattern (7) is tuned to an infinitely wide fringe. The in-plane rotation of the image rerecorded by a small angle α provides tuning to fringes of a finite width. In this case, the moiré pattern is described by the product

$$a_{im} \cos\left[2n\left(\mathbf{K}_i\mathbf{r} + \varphi_i + \frac{\varepsilon_i}{2}\right)\right] \cos[\mathbf{K}_0\mathbf{r} + n\varepsilon_i], \quad (9)$$

where $\mathbf{K}_0 = n(\mathbf{K}_i - \mathbf{K}_i')$ and \mathbf{K}_i' is the structure vector of the rotated rerecorded pattern.

The vector \mathbf{K}_0 is orthogonal to the moiré fringes, and its absolute value determines their width: $T_0 = 2\pi/|\mathbf{K}_0|$. It is evident that $T_0 = T/(2n\alpha)$ at small α .

The fact that the function $\varepsilon_i/2$, describing a macroscopic defect of the elementary structure, is multiplied by the factor $2n$ means a $2n$ -fold increase in the sensitivity of the flaw detection method. Tuning of the moiré pattern to finite-width fringes makes it possible to determine the sign in the visualized distribution of the function ε_i [12].

EXPERIMENTAL TEST

The method proposed was applied to visualize macroscopic defects in a composite structure that represents two mutually orthogonal amplitude gratings with equal periods: $T_1 = T_2 = 0.2$ mm. The substructure patterns of the reference composite structure were recorded with the separation of the first diffraction

orders. An FG-690 holographic film was used as a record medium. The figure shows reflection moire patterns observed on the diffusing screen upon superimposing the patterns rerecorded on the composite structure under study. The moire patterns visualize the functions ϵ_1 and ϵ_2 for the gratings with (a) vertical and (b) horizontal grooves with a twofold sensitivity. The rotation of the gratings rerecorded to tune to the finite-width moire fringes was the same.

REFERENCES

1. I. Havanessian and Y. Y. Hung, *Appl. Opt.* **10**, 2734 (1971).
2. I. P. Sukharev and B. N. Ushakov, *Investigation of Strains and Stresses by the Moire Fringe Method* (Mashinostroenie, Moscow, 1969).
3. N. Wadworth, M. Marchant, and B. Billing, *Opt. Laser Technol.* **5** (3), 119 (1973).
4. N. G. Vlasov and A. E. Shtan'ko, in *Proceedings of the IX All-Union School on Holography* (Leningr. Inst. Yad. Fiz., Leningrad, 1977), pp. 256–266.
5. A. M. Lyalikov, *Opt. Spektrosk.* **86**, 489 (1999) [*Opt. Spectrosc.* **86**, 429 (1999)].
6. A. M. Lyalikov and M. Yu. Serenko, *Zh. Tekh. Fiz.* **69** (10), 126 (1999) [*Tech. Phys.* **44**, 1242 (1999)].
7. A. I. But' and A. M. Lyalikov, *Kvantovaya Élektron. (Moscow)* **23**, 381 (1996).
8. A. M. Lyalikov, *Opt. Zh.* **65** (3), 54 (1998) [*J. Opt. Technol.* **65**, 220 (1998)].
9. J. Schwider, in *Proceedings of the III All-Union School on Holography* (Leningr. Inst. Yad. Fiz., Leningrad, 1972), pp. 247–254.
10. I. S. Zeřlikovich and N. M. Spornik, *Holographic Diagnostics of Transparent Media* (Universitetskoe Izd., Minsk, 1988).
11. I. S. Zeřlikovich and A. M. Lyalikov, *Usp. Fiz. Nauk* **161** (1), 143 (1991) [*Sov. Phys. Úsp.* **34**, 74 (1991)].
12. Ch. M. Vest, *Holographic Interferometry* (Wiley, New York, 1979; Mir, Moscow, 1982).

Translated by A. Chikishev

Optical Properties of Laser-Produced Plasmas from Cd and Layered Crystalline CdAl₂Se₄ Targets

A. K. Shuaibov, L. L. Shimon, and M. P. Chuchman

Uzhgorod State University, Uzhgorod, 88000 Ukraine

Received June 29, 2000

Abstract—An experimental investigation into emission by plasmas from Cd and crystalline CdAl₂Se₄ targets irradiated with a pulse-periodic YAG : Nd³⁺ laser is presented. The laser operates at a pulse width of 20 ns and provides $1\text{--}2 \times 10^9$ W/cm² at the focus. The spectral and time characteristics of the emission are examined. The main findings are as follows: (1) The strongest lines correspond to the transitions from the 6³S₁ and 5³D₂ levels of Cd I. (2) With the Cd target, the recombination bottlenecks are 5²D_{5/2} of Cd II ($E_{\text{up}} = 20.11$ eV) and 8³D_{1,3} of Cd I ($E_{\text{up}} = 8.60\text{--}8.65$ eV). (3) The average electron temperature outside the core of the plume is 0.64 eV. (4) For the 508.6-nm line of Cd I, the high filling rate of the 6³S₁ upper energy level may be related to the rapid recombination of cadmium ions with electrons via the intermediate levels of Cd⁺ at 6.82 and 7.24 eV. © 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

Cadmium-containing crystals are widely used as targets for the laser deposition of thin compound films in microelectronics [1]. A mass-spectrometric diagnostics of the plasmas from CdGa₂S₄ and CdGa₂Se₄ single crystals irradiated with an YAG : Nd³⁺ laser (delivering over 10^9 W/cm²) was reported in [2]. Optical diagnostics [3, 4] can also provide information necessary for the optimal fabrication of high-quality films. Such experiments allow one to determine n_e , T_e , and the concentrations of excited (metastable) atoms and ions in multispecies laser-produced plasmas. Since the plasmas from the crystals are characterized by very complex emission spectra [5], it may be wise to take targets made of individual elements of the crystal under study (a metal or a semiconductor). This approach may enable one to precisely interpret the emission spectra and may deepen the understanding of the laser–crystal interaction.

To the best of our knowledge, no data have been reported on the optical properties of plasmas from Cd–Al–S(Se) crystals under the action of a neodymium laser generating a relatively low radiant-flux density ($\leq 10^9$ W/cm²) at the focus. Nor have we found any quantitative data on the emission intensity distribution and other properties of plasmas from Cd targets (in the case of a 10^9 W/cm² YAG : Nd³⁺ laser).

This paper presents an experimental investigation into the optical properties of plasmas from Cd and layered crystalline CdAl₂Se₄ targets irradiated with a moderate-power YAG : Nd³⁺ laser.

EXPERIMENTAL SETUP

The plasmas were produced in a vacuum by means of a pulse-periodic YAG : Nd³⁺ laser ($\lambda = 1.06$ μm and $\tau = 20$ ns) providing $1\text{--}2 \times 10^9$ W/cm² at the focus. The vacuum chamber with the targets was pumped down to 3–12 Pa. We used highly pure cadmium, whose metal-impurity percentage was $10^{-4}\text{--}10^{-2}\%$ at most. The CdAl₂Se₂ target was a layered crystal grown and purified by a technique similar to that reported in [2]. Its impurity percentage was below $10^{-6}\text{--}10^{-4}\%$. The emission by the plasmas was examined with an MDR-2 monochromator, an FÉU-106 photomultiplier (PMT), and a KSP-4 recorder. The monochromator–PMT system was calibrated in a range of 200–600 nm with an SI8-200 band lamp and a DVS-25 hydrogen lamp. Emission intensity was measured to an accuracy of 3–5%. For the core of the plume, time characteristics were obtained with an ÉLU 14-FS linear electron multiplier and a 6 LOR-04 fast oscilloscope. (In this case, the distance r from the surface to the point of interest is within 1 mm.) For wider pulses ($\tau \geq 1$ μs) and the region outside the core ($r \geq 3\text{--}5$ mm), we used a Foton pulsed PMT and a CI-99 oscilloscope, which insured a time resolution of 20–30 ns. Emission spectra were interpreted on the basis of data borrowed from [6–8].

SPECTRAL AND TIME CHARACTERISTICS OF THE EMISSION

Figure 1 and Table 1 refer to the Cd target. Figure 1 shows a survey emission spectrum. Table 1 presents interpreted spectral data, the fractional intensities of Cd I and Cd II lines, and respective upper energy levels. The rightmost column lists the values of $\Delta J/k_\lambda$, i.e., the

Table 1. Emission spectrum of the laser-produced plasma from the Cd target

No.	λ , nm	Degree of ionization	E_{up} , eV	Transition	J/k_{λ} , arb. units	$\Delta J/k_{\lambda}$, %
1	508.6	Cd I	6.39	$5^3P_2-6^3S_1$	1.00	21
2	480.0	Cd I	6.39	$5^3P_1-6^3S_1$	0.60	13
3	467.8	Cd I	6.39	$5^3P_0-6^3S_1$	0.20	5
4	441.3	Cd I	6.61	$5^3P_1-6^3S_0$	–	–
5	361.1	Cd I	7.37	$5^3P_2-5^3D_2$	0.70	15
6	346.7	Cd I	7.37	$5^3P_1-5^3D_{2,1}$	0.25	5
7	340.4	Cd I	7.37	$5^3P_0-5^3D_1$	–	–
8	+ 326.1	Cd I	3.80	$5^1S_0-5^3P_1$	0.20	5
	+ 325.3	Cd I	7.76	$5^3P_1-7^3S_1$		
9	313.3	Cd I	7.76	$5^3P_1-7^3S_1$	0.03	<2
10	308.1	Cd I	7.76	$5^3P_0-7^3S_1$	0.02	5
11	298.1	Cd I	8.10	$5^3P_2-6^3D_{3,2}$	0.25	5
12	288.1	Cd I	8.10	$5^3P_0-6^3D_{2,1}$	0.15	<5
13	283.7	Cd I	8.10	$5^3P_0-6^3D_1$	0.05	<2
14	+ 276.7	Cd I	8.44	$5^3P_2-7^3D_{3,2}$	0.20	5
	+ 274.9	Cd II	10.28	$5^2P_{3/2}-6^2S_{1/2}$		
15	266.8	Cd I	8.44	$5^3P_1-7^3D_2$	0.05	<2
16	266.0	Cd I	8.60	$5^3P_2-8^3D_3$	0.05	<2
17	+ 255.3	Cd I	8.65		0.10	2
	+ 254.5	Cd I	8.60	$5^3P_0-8^3D_1$		
18	231.3	Cd II	11.12	$5^2P_{3/2}-5^2D_{5/2}$	0.10	<5
19	228.8	Cd I	5.41	$5^1S_0-5^1P_1$	0.30	7
20	+ 226.5	Cd II	9.26	$5^2S_{1/2}-5^2P_{1/2}$	0.25	5
	+ 226.7	Cd I	5.47			

ratio of the intensity of an individual spectral line to the total intensity of the line emission in the region 200–600 nm.

The emission lines of Cd I and Cd II were observed against a faint continuous background. The spectra indicated no presence of cadmium ions with higher degrees of ionization. The highest intensities were measured for the Cd I lines located at 508.6, 479.9, and 361.1 nm. The upper level of the first and second line is 6^3S_1 ($E_{\text{up}} = 6.39$ eV), and that of the third is 5^3D_3 ($E_{\text{up}} = 7.37$ eV). The intensity of a Cd I resonance line (at $\lambda = 326.1$ nm) was fairly low, which may stem from considerable self-absorption in the plasma. For erosion laser-produced plasmas expanding into a vacuum, it is commonly held that excited levels of metal ions and atoms are filled mainly due to electron–ion recombination, e.g., $\text{Cd}^+ + e + e - \text{Cd}^* + e$ [4, 9]. Accordingly, Table 1 enables us to reveal the bottlenecks of the recombination stream flowing downwards through the energy levels of Cd II and Cd I. (From here on, *recombination bottleneck* means the highest ionic or atomic energy

level from which radiative transitions are detected [4].) The recombination bottlenecks were found to be $5^2D_{5/2}$ of Cd II ($E_{\text{up}} = 20.11$ eV) and $8^3D_{1,3}$ ($E_{\text{up}} = 8.60$ –8.65 eV).

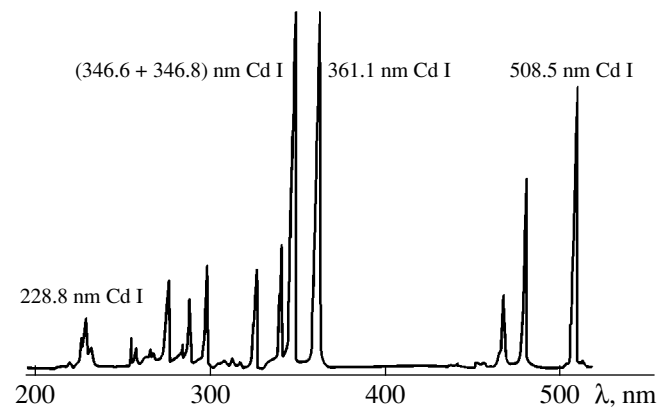


Fig. 1. Plasma from the Cd target: a survey emission spectrum ($r = 7$ mm). The plume originates from the surface of the target.

Table 2. Emission spectrum of the laser-produced plasma from the CdAl₂Se₄ target

No.	λ , nm	Element	E_{up} , eV	Transition	J/k_{λ} , arb. units	$\Delta J/k_{\lambda}$, %
1	+ 432.0	Se II	16.88		0.15	<2
	432.9	Se I	9.18			
2	426.5	–	–	–	0.15	<2
3	421.3	Se II	14.92		0.85	8
4	404.7	Se III	19.38		0.65	6
5	396.2	Al I	3.14	$3^2P_{3/2}-4^2S_{1/2}$	0.70	6
6	394.4	Al I	3.14	$3^2P_{1/2}-4^2S_{1/2}$	0.55	5
7	365.5	Al II			0.40	<5
8	361.1	Cd I	7.37	$5^2P_2-5^3D_3$	0.15	<2
9	+ 337.9	Se III			1.00	10
	338.7					
10	+ 325.2	Cd I	7.76	$5^3P_2-7^3S_1$	1.00	10
	326.1		3.8	$5^1S_0-5^3P_1$		
11	309.3	Al I	4.02	$3^2P_{3/2}-3^2D_{5/2}$	0.70	7
12	308.2	Al I	4.02	$3^2P_{1/2}-3^2D_{3/2}$	0.45	<5
13	288.1	Cd I	8.1	$3^3P_1-6^3D_{2,1}$	0.25	2
14	+ 275.7	Cd I			0.20	2
	274.9					
15	266	Cd I	8.6	$5^3P_2-8^3D_3$	0.10	<2
16	258.1	Cd I	8.6	$5^3P_1-8^3D_2$	0.25	2
17	243.4	Cd II			0.15	<2
18	232.1	Cd I	–	–	1.00	10
19	228.8	Cd I	5.41	$5^1S_0-5^1P_1$	0.55	5
20	+ 226.5	Cd II			0.50	5
	226.8					
21	381.2	Se III	5.47	$5^2S_{1/2}-5^2P_{1/2}$	0.55	5

The average electron temperature was determined from the distribution of J/k_{λ} for Cd I. The values of T_e were found from the intensity ratios of the lines [10]. The transition probabilities of the Cd I lines were borrowed from [11, 12]. Since the spread in E_{up} for the Cd I lines was modest ($\Delta E \leq 2$ eV), we evaluated T_e from the slope of $\ln(I_1/I_2)$ plotted against $E_1 - E_2$, where I_1 and I_2 are line intensities and E_1 and E_2 are the

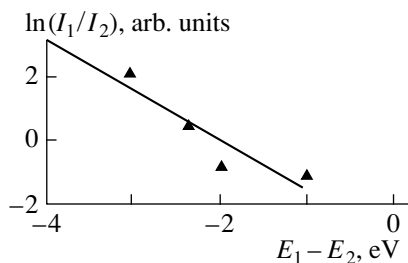


Fig. 2. Plasma from the Cd target: emission-intensity ratio vs. the difference between the corresponding upper energy levels for Cd I spectral lines.

respective upper energy levels (Fig. 2). Due to self-absorption in the plasma, the resonance Cd I line was not covered by the procedure, which allowed us to obtain an acceptable scatter of points in Fig. 2.

Figure 3 shows emission-intensity oscillograms of the strongest Cd I lines for the core of the plume from the Cd target ($r = 1$ mm). It is seen that the leading edges are 5–15 ns long and the trailing edges decay almost linearly by 5–15% of J_{max} per 100 ns. The 508.6-nm line is the exception: its intensity falls at a rate as high as 3% of J_{max} per 1 ns in the linear segment. As is known, the transition probabilities for the 508.6- and the 479.9-nm lines are close to each other: they are 1.56×10^{-8} and $1.41 \times 10^{-8} \text{ s}^{-1}$, respectively [11]. Also, the lower level for the 508.6-nm Cd I line is metastable, with a lifetime of 130 s, and that for the 479.9-nm line has $\tau_{ik} = 2.4 \times 10^{-6} \text{ s}$ [12]. Consequently, the short duration of the pulsed emission and the high decay rate (and, hence, the short recombination-filling time for the 508.6-nm Cd I line [4]) may be due to the fact that the width of the emission pulse is self-limited owing to the

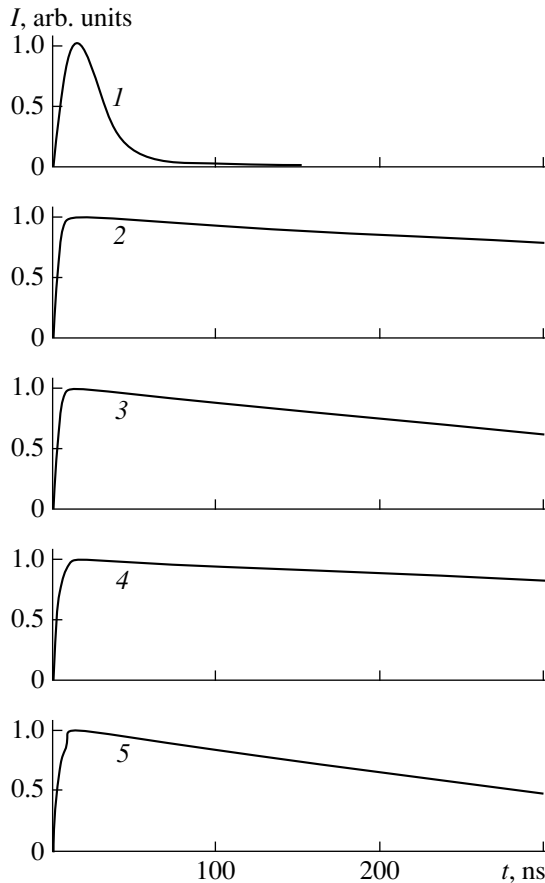


Fig. 3. Plasma from the Cd target: emission-intensity oscillograms for Cd I spectral lines observed in the core of the plume ($r = 1$ mm): (1) 508.6, (2) 479.9, (3) 467.8, (4) 361.1, and (5) 346.7 nm.

metastability of the lower level. Furthermore, 5^3P_2 is the lower level for the 361.1-nm Cd I line as well. This line gives a slowly decaying pulse, and its τ_{ik} approximates to those of the lines at 508.6 and 479.9 nm. Therefore, the difference in duration between the 508.6-nm line and the other strong lines of Cd I mainly stems from the ways in which the corresponding upper levels are filled. The distinctive manner in which Cd I (6^3S_1) is excited by slow electrons was first noted in [13]. It has been demonstrated that the excitation at near-threshold electron energies (≤ 8.00 eV) goes only via an intermediate Cd⁺ level ($E_{up} = 6.82$ eV) of half-width 0.68 eV and that a higher intermediate level (7.24 eV) also exists [13]. Thus, the most probable reason for the high intensity and upper-level filling rate of the 508.6-nm Cd I line is the rapid recombination of Cd⁺ ions with electrons via the intermediate levels of Cd⁺.

Figure 4 and Table 2 refer to the CdAl₂Se₄ target. Figure 4 shows an emission spectrum over the range 220–435 nm, with the plume originating from the surface of the target. Table 2 presents interpreted spectral

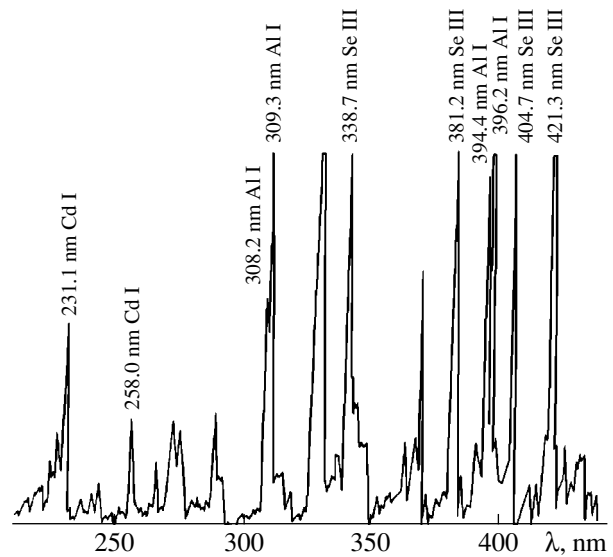


Fig. 4. Plasma from the CdAl₂Se₄ target: an emission spectrum ($r = 7$ mm). The plume originates from the surface of the target.

data and the values of J/k_λ and $\Delta J/k_\lambda$. The salient features of the spectrum are Cd I and Al I lines, which are typical of this wavelength region. For comparison, we performed an experiment with Al targets (under similar irradiation and detection conditions). It was found that the plume spectrum comprises only four strong Al I lines located at 396.2, 394.4, 309.3, and 308.2 nm. They account for 12, 14, 37, and 37% of the total intensity, respectively. With the CdAl₂Se₄ target, selenium showed up as strong lines of Se II and Se III ions. The surface irradiation of a slab made of highly pure selenium caused no plume. Instead, a faintly glowing core was observed on the surface; its emission spectrum was continuous under the specified irradiation conditions. This implies that the formation of excited Se ions is related to the features of crystal destruction in the laser plasma. A mass-spectrometric analysis of CdGa₂Se₄, a compound of similar properties, has revealed that the majority of charged particles in the laser-produced plasma are Se⁺, Ga⁺, and Cd⁺ ions with corresponding percentages of 46, 28, and 14% [2]. We believe that the plasma from a CdAl₂Se₄ target must have basically the same ionic composition, although we detected a large number of excited Se III ions. With respect to the laser deposition of thin compound films, the most intense lines of Cd I, Al I, and selenium ions can be used to evaluate the plasma properties (n_e , T_e , the concentrations of atoms and ions in metastable states, etc.) that determine the film quality [14, 15].

CONCLUSIONS

We have conducted an experimental investigation into emission by laser-produced plasmas from Cd and

layered crystalline CdAl₂Se₄ targets irradiated with a pulse-periodic YAG : Nd³⁺ laser of moderate output pulse power. The findings are as follows:

(1) The strongest lines correspond to the transitions from the 6S₁ and 5³D₃ levels of Cd I.

(2) With the Cd target, the recombination bottlenecks are 5²D_{5/2} of Cd II ($E_{\text{up}} = 20.11$ eV) and 8³D_{1,3} of Cd I ($E_{\text{up}} = 8.60$ – 8.65 eV).

(3) The average electron temperature outside the core of the laser plume (at $r = 7$ mm) is 0.64 eV.

(4) For the 508.6-nm line of Cd atoms, the filling of the upper energy level may be related to the rapid recombination of Cd⁺ with electrons via intermediate levels of Cd⁻ ($E_{\text{up}} = 6.82$ and 7.24 eV), as suggested by the emission dynamics of the core plasma of the laser plume.

(5) For the CdAl₂Se₄ target, the strongest lines correspond to Cd I, Al I, Se II, and Se III.

The results of this study may be useful for the diagnostics of plasmas from cadmium- and aluminum-containing crystals by the transient emission or absorption spectroscopy.

REFERENCES

1. A. A. Shchuka, V. G. Dneprovskii, and V. G. Dudoladov, *Élektron. Tekh.* **72** (24), 38 (1973).
2. I. É. Kacher, I. I. Opachko, and M. Yu. Rigan, *Ukr. Fiz. Zh.* **34**, 1728 (1989).
3. R. V. Arutyunyan, V. Yu. Baranov, L. A. Bol'shov, *et al.*, *Laser Emission Action on Materials* (Nauka, Moscow, 1989), pp. 88–100.
4. L. T. Sukhov, *Laser Spectrum Analysis* (Nauka, Novosibirsk, 1990).
5. B. K. Kotlyarchuk, D. I. Popovich, and V. Ya. Pentko, *Zh. Tekh. Fiz.* **57**, 1824 (1987) [*Sov. Phys. Tech. Phys.* **32**, 1091 (1987)].
6. M. A. Elyashevich, *Atomic and Molecular Spectroscopy* (GIFML, Moscow, 1962).
7. A. N. Zaidel', V. K. Prokof'ev, S. M. Raïskii, *et al.*, *Tables of Spectral Lines* (Nauka, Moscow, 1969; Plenum, New York, 1970).
8. A. R. Striganov and G. A. Odintsova, *Tables of Spectral Lines of Atoms and Ions: A Handbook* (Énergoatomizdat, Moscow, 1982).
9. V. I. Derzhiev, A. G. Zhidkov, and S. I. Yakovlenko, *Ion Radiation in Nonequilibrium Dense Plasma* (Énergoatomizdat, Moscow, 1986).
10. *Plasma Diagnostics*, Ed. by W. Lochte-Holtgreven (American Elsevier, New York, 1968; Mir, Moscow, 1971).
11. W. L. Wise and G. A. Martin, *Wavelength and Transition Probabilities for Atoms and Atomic Ions, Part II: Transition Probabilities* (US Department of Commerce, National Bureau of Standards, Washington, 1980); *Natl. Stand. Ref. Data Ser.* **68** (IV), 359 (1980).
12. A. A. Radtsig and B. M. Smirnov, *Reference Data on Atoms, Molecules, and Ions* (Énergoatomizdat, Moscow, 1986; Springer, Berlin, 1985).
13. O. B. Shpenik, I. P. Zapesochnyi, V. V. Souter, *et al.*, *Zh. Éksp. Teor. Fiz.* **65**, 1797 (1973) [*Sov. Phys. JETP* **38**, 898 (1974)].
14. A. D. Akhsakhalyan, Yu. A. Bityurin, and S. V. Gaponov, *Zh. Tekh. Fiz.* **52**, 1584 (1982) [*Sov. Phys. Tech. Phys.* **27**, 969 (1982)].
15. S. V. Gaponov, A. A. Gudkov, A. A. Fraerman, *et al.*, *Zh. Tekh. Fiz.* **52**, 1843 (1982) [*Sov. Phys. Tech. Phys.* **27**, 1130 (1982)].

Translated by A. Sharshakov

Response of a Gyrotron to Small-Amplitude Low-Frequency-Modulated Microwaves Reflected from a Plasma

G. M. Batanov*, L. V. Kolik*, Yu. V. Novozhilova**, M. I. Petelin**, A. E. Petrov*,
K. A. Sarksyan*, N. N. Skvortsova*, and N. K. Kharchev*

**Institute of General Physics, Russian Academy of Sciences, Moscow, 119991 Russia*

***Institute of Applied Physics, Russian Academy of Sciences, Nizhni Novgorod, 603600 Russia*

Received June 1, 2000

Abstract—Experiments in the L-2M stellarator revealed the intense noise modulation of the gyrotron power and the change in its mean value under the action of the noise modulation of radiation reflected from the plasma column. The effect observed is explained in terms of the resonant locking of the gyrotron self-oscillations due to wave reflection from the fluctuating plasma load. © 2001 MAIK “Nauka/Interperiodica”.

1. It is commonly accepted that the use of quasi-optical systems for transmitting and focusing gyrotron radiation in plasma magnetic confinement systems makes it unnecessary to specially protect the gyrotron from the action of the reflected wave. Indeed, even at a relatively high reflection coefficient ($\sim 10\%$), it is hardly possible that the front of the reflected wave can be matched to the front of the incident wave, which guarantees a significant weakening of the reflected waves that propagate backward through the quasi-optical transmission line and arrive at the gyrotron output window. Thus far, this approach has dominated in experimental investigations and seemed to be quite reasonable. However, in our experiments on plasma electron cyclotron heating at the second harmonic of the electron gyrofrequency in the L-2M stellarator, we encountered an unexpected phenomenon—the response of a gyrotron loaded with a strongly absorbing gyrotropic plasma to reflected radiation. The aim of this paper is to describe this effect.

2. A four-mirror quasi-optical line is used in the L-2M stellarator to transport gyrotron radiation and focus it onto the outer horizontal port. The distance between the gyrotron and the input window is about 5 m. A resonant quartz input window 85 mm in diameter is positioned in the waist of the last focusing mirror. The cylindrical surface of the rear wall of the stellarator vacuum chamber, together with the flange of the port situated on it, defocuses the beam; as a result, the intensity of the microwave beam reflected backward through the input port is damped substantially (the reflection coefficient is $\sim 0.7\%$). At the beam axis, between the last mirror and the quartz window, a bidirectional coupler with two detectors for monitoring the envelop of the gyrotron pulse and the reflected radiation is installed. Such monitoring is very useful when adjusting the gyrotron operating conditions because the envelop of the gyrotron pulse is very sensitive to small

variations in the beam current and cathode voltage of the gyrotron. In the L-2M stellarator, a GYKOM gyrotron operating at a frequency of 75 GHz, with a power of up to 400 kW is used. The standard operating conditions are as follows: the power is 150–250 kW, the pulse duration is 10–12 ms, the voltage is 60–63 kV, and the beam current is 17–20 A. In addition to the incident and reflected power signals measured with the help of the bidirectional coupler, we also measured the radiation transmitted through the plasma column. The transmitted signal passed through the port situated at the inner side of the stellarator chamber wall and was recorded by detectors equipped with rectangular waveguide sections, which made it possible to measure radiation with the given orientation of the electric field. The detectors were mounted in absorbent rubber cylinders, which served to blind the detectors and narrow their directional patterns.

In experiments, the electric field vector of a linearly polarized wave was perpendicular to the stellarator magnetic axis, lying in the equatorial plane of the toroidal chamber. The polarization of the reflected wave was usually the same as that of the incident wave. In the absence of a plasma, the passed wave also had the same polarization as the incident wave. In the presence of the plasma column, the linearly polarized incident wave splits into an extraordinary and ordinary wave with elliptical polarization. In a hot plasma, the extraordinary wave is absorbed almost completely in the gyroresonance region during a single pass through the plasma column (the absorption coefficient is not less than 0.9), whereas the ordinary wave passes without attenuation. The signal passed through the plasma is the sum of the partially absorbed extraordinary wave and the unabsorbed ordinary wave. As the electron density smoothly varies during the discharge, the passed wave undergoes Faraday rotation, which manifests itself as oscillations in the detector signal amplitude [1].

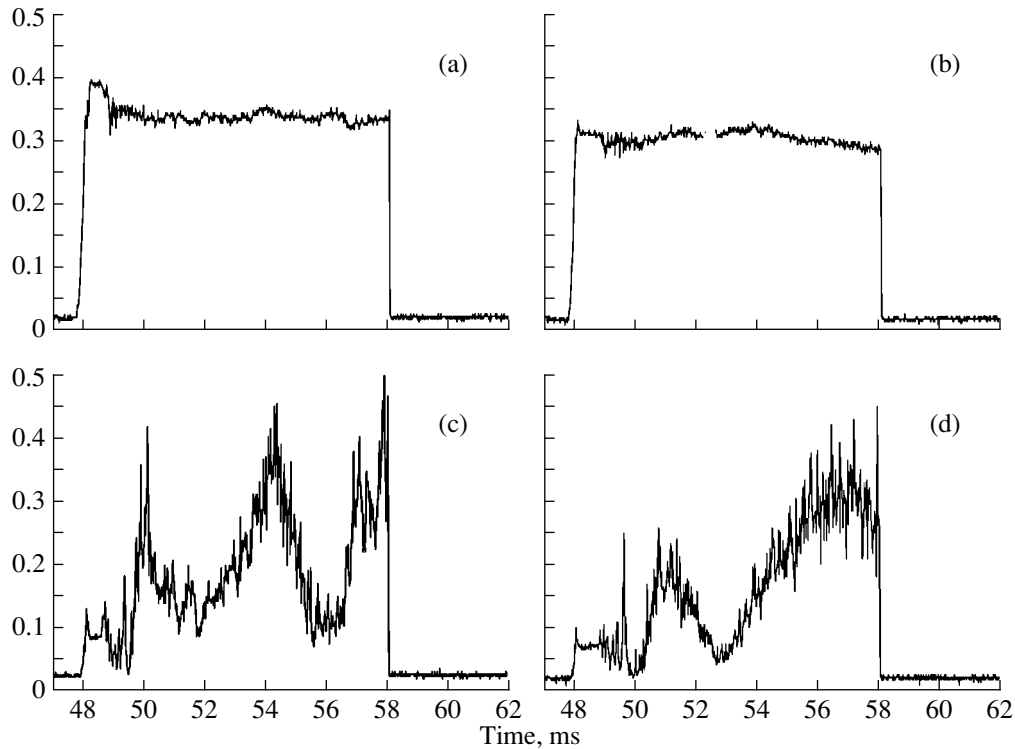


Fig. 1. Signals of (a, b) the gyrotron power (direct signal) and (c, d) radiation passed through the plasma for two stellarator discharges. The sampling time is 16 μ s.

The detector signals were processed with the help of an analog-to-digital converter (ADC). The ADC sampling frequency was 0.1–16 μ s. The time constant of the detectors was about 1 μ s.

3. In the absence of a plasma, the gyrotron generates a flat-top microwave pulse with a rise time of 0.5 ms and a fall time of 20–30 μ s. The gas breakdown and the formation of the plasma column last for 1–1.5 ms. During this time, the gyrotron signal remains unchanged (Fig. 1). After the plasma column is formed, noise oscillations arise at the plateau of the incident-power signal. Depending on the gyrotron operating conditions, the averaged pulse power can decrease (Fig. 1a), remain constant, or increase (Fig. 1b).

Higher amplitude oscillations are seen in the signal passed through the plasma (Figs. 1c, 1d). In this case, the detector was oriented so as to measure the field component parallel to the stellarator magnetic axis. A low-level flat-top burst in the signal arises because of the insufficient adjustment of the detector. Broad peaks are produced due to the Faraday effect.

The effect of the plasma on the gyrotron operation, which manifests itself in the gyrotron power oscillations, may be attributed to the modulation of either the reflection coefficient or the spectrum of the reflected radiation. Indeed, intense noise fluctuations are seen in the reflected signal measured with a bidirectional coupler (Fig. 2). In this case, the modulation amplitude of the reflected signal is substantially higher than that of

the gyrotron signal and is as high as the noise modulation of the microwave radiation passed through the plasma (Figs. 1c, 1d). That the modulation amplitude of the radiation reflected from the plasma and passed through the plasma is much higher than the modulation amplitude of the gyrotron signal shows that the modulation of the reflected and passed radiation is due to the scattering by plasma fluctuations. The modulation of the radiation passed through the plasma is related to the small-angle scattering by plasma fluctuations and the interference of the scattered and passed waves (the phase-contrast version of the small-angle scattering technique [2]). The modulation of the reflected signal can be attributed to either backward scattering of the incident wave by plasma fluctuations or small-angle scattering of the radiation reflected backward from the rear wall of the stellarator chamber. We note that the slow (on a time scale of milliseconds) oscillations of the reflected signal are similar to the oscillations of the transmitted signal. The reason for these slow oscillations is also similar to that for the slow oscillations of the transmitted signal: we observe the interference between the extraordinary and ordinary waves reflected from the plasma and those reflected from the rear wall of the chamber and passed through the gyrotropic plasma column.

Note also that, after the plasma column is produced, the amplitude of the reflected signal drops by one order

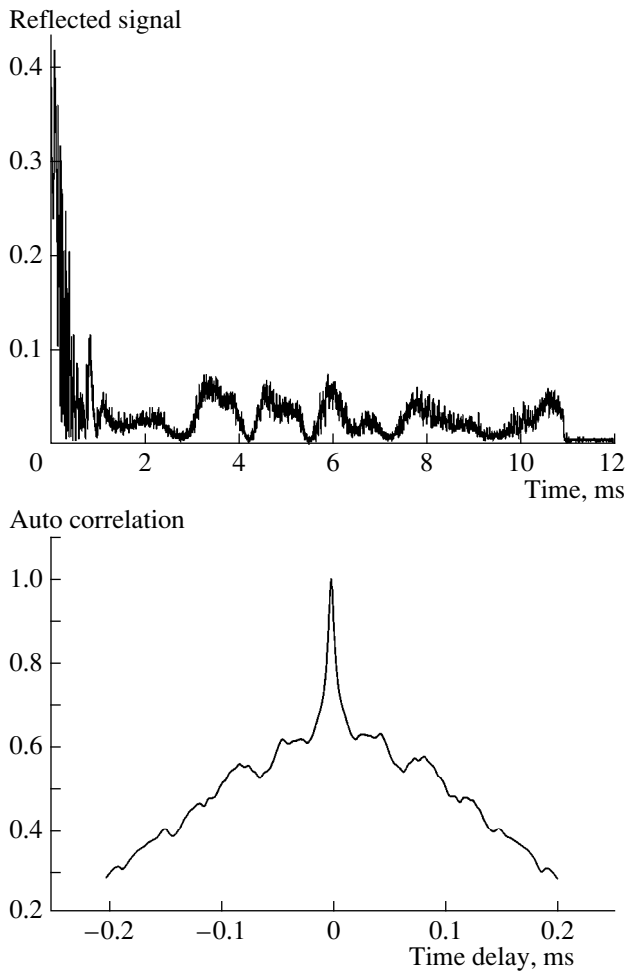


Fig. 2. Reflected signal and its autocorrelation function. The sampling time is $0.1 \mu\text{s}$. The averaging is performed over four time intervals: 3–4, 4.5–5.5, 7.5–8.5, and 9.8–10.8 ms.

of magnitude (the reflection coefficient falls to 0.1%), as seen in Fig. 2.

4. To study the influence of the amplitude and phase of the reflected microwave on the gyrotron power, we carried out test measurements in the absence of a discharge in the stellarator. For this purpose, a thin mica film with a thickness of $110 \mu\text{m}$ was placed in the waist of the microwave beam in front of the stellarator window; the film plane was perpendicular to the beam axis. From shot to shot, the film was gradually displaced along the beam axis, which allowed us to vary the amplitude and phase of the reflected wave produced as a result of the interference between two waves reflected from the stellarator and the film. The phase variation attained 0.688π , and the reflection coefficient varied from its minimum value of 0.014 up to 0.065. Within this range of the amplitudes and phases of the reflected wave, we did not observe any variations in the gyrotron power. These measurements permitted us to determine the reflection coefficients from the stellarator chamber and the mica film: the former was 0.67%; the latter, 4%.

Consequently, neither variations in the time-independent reflection coefficient from 0.014 to 0.065 nor variations in the time-independent phase of up to 0.688π influence the gyrotron power, whereas noise fluctuations in the reflection coefficient, which is on the order of 0.1%, produce the modulation of the gyrotron power and change its mean value by 10% and higher.

The relation between the noise fluctuations of the plasma column and the gyrotron power was studied numerically by calculating the autocorrelation functions of gyrotron radiation (the direct signal) and the reflected and scattered radiation. The autocorrelation functions were computed using the method developed by van Milligen [3] (with a low-frequency digital filter with a 1-kHz band and a 6-dB attenuation per octave). This allowed us to cut the dc component of the signal and its slow variations due to the interference of the waves with different polarizations. Figures 3 and 4 show the gyrotron signals and their autocorrelation functions calculated by averaging over 4 ms with a sampling time of $2 \mu\text{s}$. The delay in the appearance of the noise fluctuations in Fig. 4 is due to the delay in the gas breakdown and the formation of a plasma in the stellarator under conditions when the resonance region is displaced from the magnetic axis toward the outer side of the stellarator chamber. In both cases, the correlation time for three-fourths of the noise power is about $16 \mu\text{s}$; i.e., the noise frequencies lie above 30 kHz. For the reflected signal, the situation is quite different. As seen from Fig. 2, only one-third of the noise power has a correlation time of $16 \mu\text{s}$ and no less than one-half of the noise power lies within the range of lower frequencies. An even stronger difference is observed for the scattered and gyrotron radiation spectra. Figure 5 illustrates the scattered and gyrotron signals measured simultaneously in a discharge, and Fig. 6 shows the autocorrelation functions for these signals. It follows from Fig. 6 that no more than one-fourth of the noise power in the scattered signal lies in the range of short correlation times (shorter than $16 \mu\text{s}$), whereas three-fourths of the gyrotron noise power is concentrated here.

Hence, in addition to the fact that the noise modulation amplitude of the scattered and reflected radiation substantially exceeds the modulation amplitude of gyrotron radiation, we also see the difference between their spectra; namely, the low frequency modulation is dominant in the scattered and reflected radiation spectra.

5. The influence of microwave radiation reflected from a plasma on the gyrotron spectrum is easily described in terms of a self-oscillator that has undergone an external quasi-resonance signal.¹ Recall [4, 5] that the self-oscillator phase can be captured by an arbitrarily small signal if its frequency falls within a narrow

¹ The equivalent approach is based on a model in which the effect of reflection from a fluctuating plasma is treated as the modulation of the eigenfrequency and quality factor of the resonator [1].

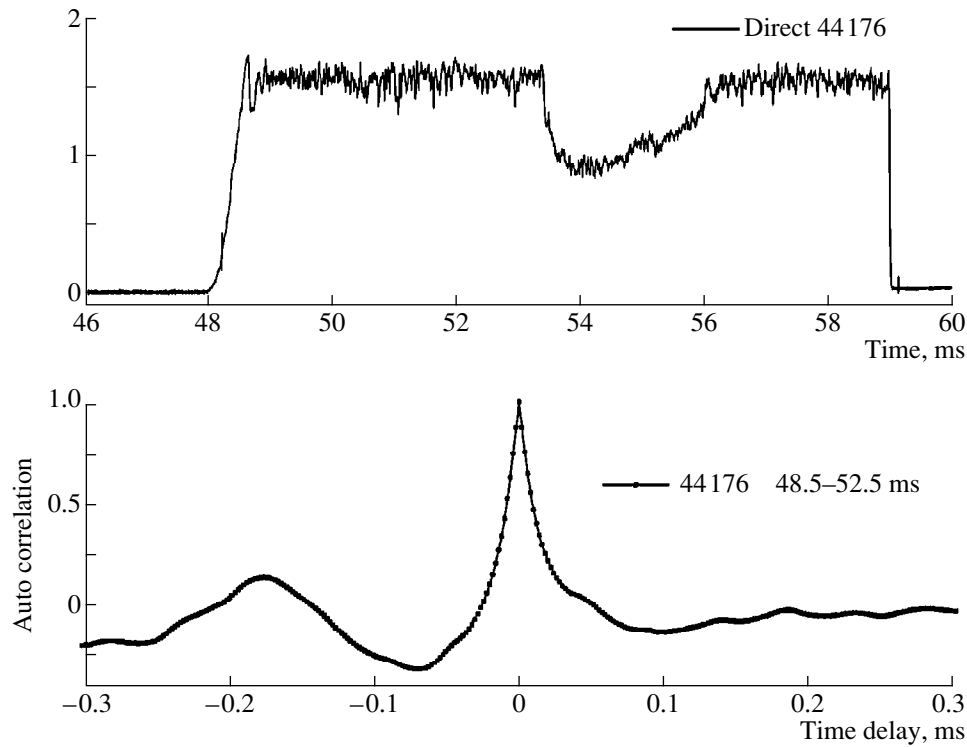


Fig. 3. Gyrotron signal and its autocorrelation function. The resonance point lies at the magnetic axis of the stellarator.

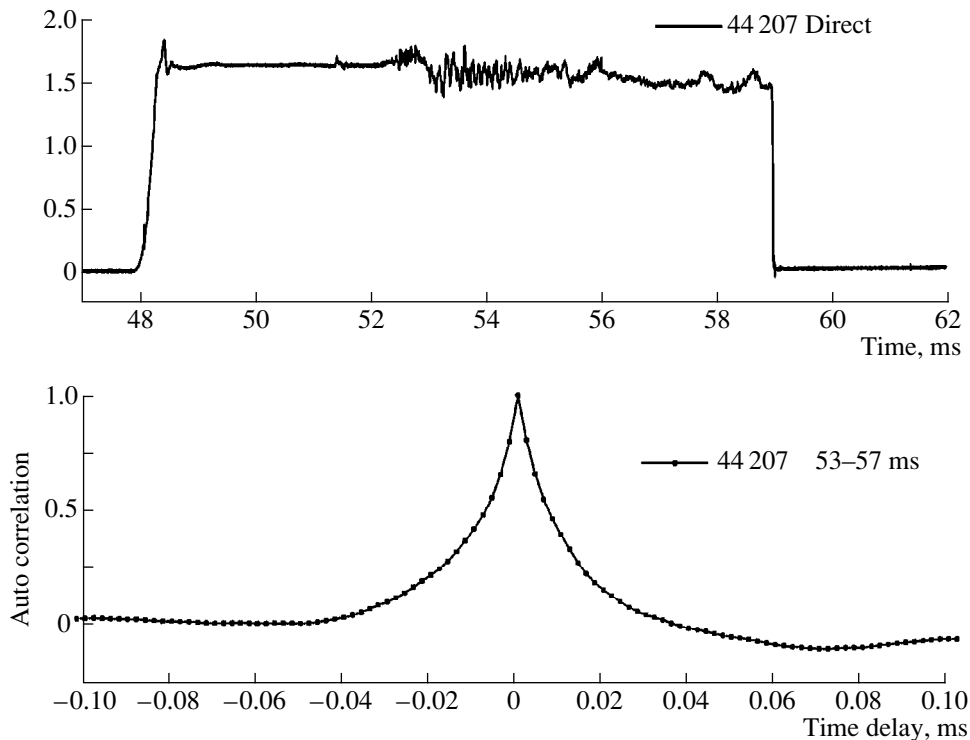


Fig. 4. Gyrotron signal and its autocorrelation function. The resonance point is displaced from the magnetic axis toward the outer side of the chamber.

band proportional to the amplitude of this signal. Otherwise, if the controlling-signal frequency falls outside the locking band ($\Delta f = R\omega/Q$, where R is the amplitude

reflection coefficient from the load and Q is the quality factor of the resonator), then the additional oscillations provoked by this signal are merely imposed on the free

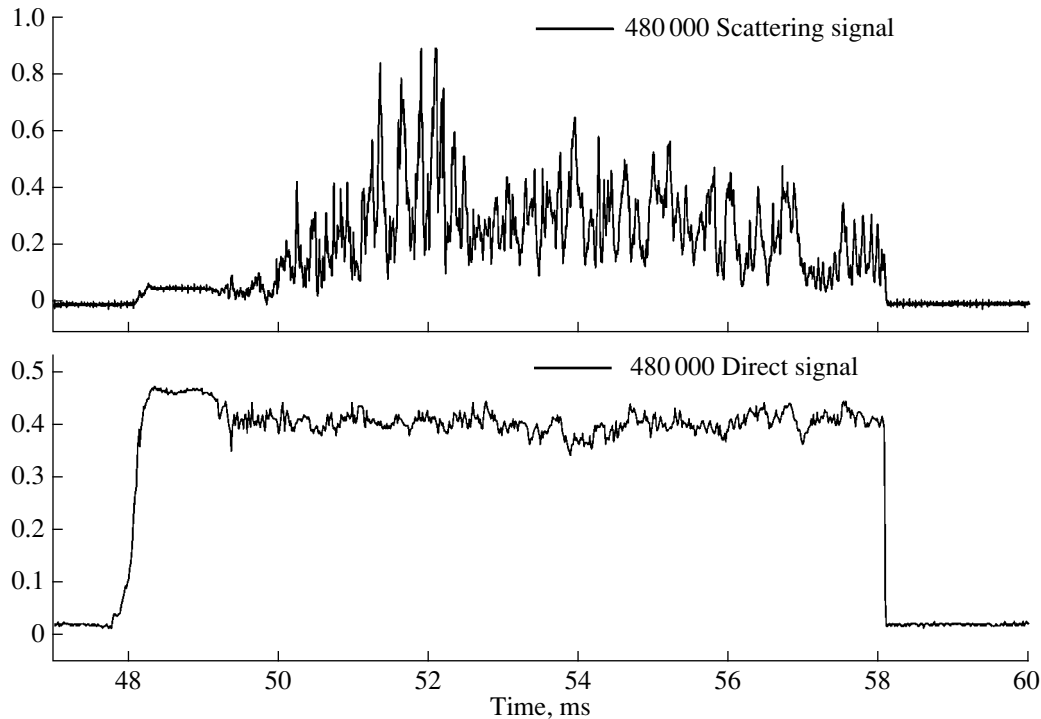


Fig. 5. Gyrotron signal (the sampling time is $16 \mu\text{s}$) and scattered radiation signal (the sampling time is $1 \mu\text{s}$) measured in the same discharge.

(not synchronized) self-oscillations. A relatively weak response of the self-oscillator to the signal outside the locking band is resonant within the frequency band on the order of $p \sim \omega/Q$.

The role of the quasi-resonant controlling signal may be played by secondary fields that arise due to scattering of the primary self-oscillator field by the fluctuations of self-oscillator parameters, including its load. Being scattered by broadband fluctuations, each spectral component of the primary self-oscillator field gives rise to a secondary broadband controlling signal. It follows from above that only the low-frequency component of the controlling signal spectrum (within the band on the order of p) affects the self-oscillator.

Therefore, the spectrum of a self-oscillator with weakly fluctuating parameters consists of a relatively narrow primary peak standing on a relatively broad secondary pedestal (Fig. 7). The primary-peak width is determined by the resonant locking of the self-oscillator phase by the lowest frequency fluctuations of the self-oscillator parameters and is proportional to their amplitude, whereas the width of the pedestal that arises due to broadband low-frequency fluctuations is on the order of p .

In our experiments, the pedestal was very low and the primary self-oscillation peak was only detected. The broadening of this peak due to the locking effect apparently takes place even when the gyrotron is perfectly matched to the output transmission line (Fig. 7); in this case, the role of the controlling signal is played

by perturbations caused by the scattering of the self-oscillator field by low-frequency fluctuations of the parameters of the self-oscillator itself. Under our experimental conditions, the reflection from the mica film with constant parameters cannot produce the additional components in the controlling fluctuation spectrum and only results in a slight displacement of the central frequency of the gyrotron self-oscillations. However, the equally small reflection from the plasma gives rise to controlling fluctuations with a much higher amplitude

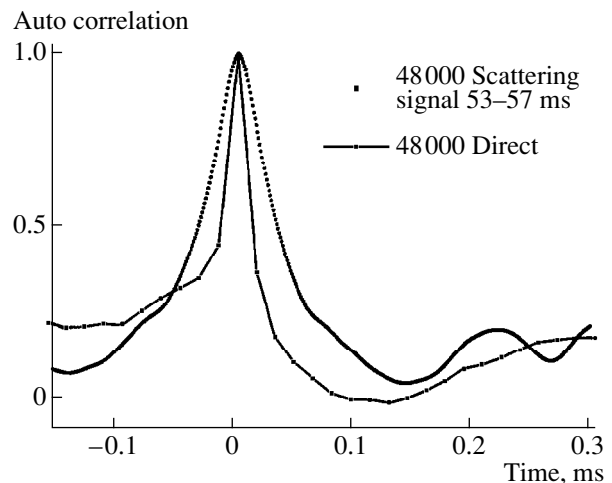


Fig. 6. Autocorrelation functions for the signals shown in Fig. 5.

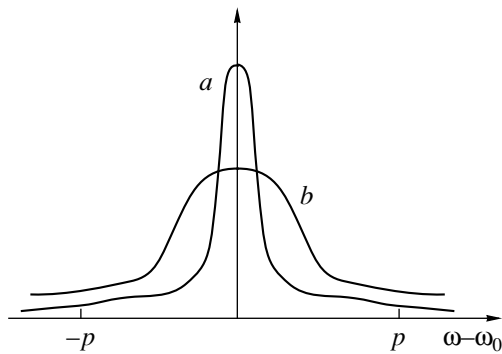


Fig. 7. Spectra of a self-oscillator with (a) the matched output and (b) the partially reflecting fluctuating load; ω_0 is the frequency of an oscillator with perfectly stable parameters.

in comparison with the fluctuations arising immediately in the gyrotron working volume, thus resulting in a significantly broader self-oscillation peak in the output radiation spectrum (Fig. 7).

6. The results obtained can be summarized as follows:

The modulation of the gyrotron power and the change in its mean value under the action of plasma fluctuations has been found experimentally at low values of the reflection coefficient (~ 0.001) of the gyrotron radiation from the plasma column. The observed effect can be attributed to the resonant frequency locking of gyrotron oscillations by the waves that arise from the scattering of gyrotron radiation by turbulent plasma fluctuations, propagate backward through the transmission line, and arrive at the gyrotron output window.

It is shown that both variations in the reflection coefficient from the steady-state load in the range 0.014 – 0.065 and the variations in the reflected-wave phase by 0.68π have no effect on the gyrotron power.

It is demonstrated that the noise modulation amplitude of the nearly forward scattered radiation or the backward reflected radiation is several times larger than the amplitude of the gyrotron power modulation and that the low frequency modulation is dominant in the spectra of the scattered and reflected radiation, in contrast to the gyrotron noise spectrum.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 00-02-17507.

REFERENCES

1. E. V. Suvorov and A. A. Fraïman, *Fiz. Plazmy* **6**, 1161 (1980) [*Sov. J. Plasma Phys.* **6**, 639 (1980)].
2. A. G. Zhukovskii, *Fluctuations of Plasma Density* (Énergoatomizdat, Moscow, 1993).
3. B. van Milligen, E. Sánchez, T. Estrada, *et al.*, *Plasma Phys.* **2**, 3017 (1996).
4. A. N. Malakhov, *Fluctuations in Self-Oscillation Systems* (Nauka, Moscow, 1968).
5. M. I. Rabinovich and D. I. Trubetskov, *Introduction to the Theory of Oscillations and Waves* (Nauka, Moscow, 1984), Sec. 16.1.

Translated by N. Larionova

A Simple Quadrupole–Octupole Electromagnetic Lens

L. P. Ovsyannikova and T. Ya. Fishkova

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

e-mail: L.Ovsyannikova@pop.ioffe.rssi.ru

Received July 17, 2000

Abstract—Charged particle beam trajectories in a simple electromagnetic quadrupole–octupole lens are numerically calculated based on analytical expressions for the +potential distribution. Locations of the lens foci in the image space are determined over a wide range of initial conditions. Relationships between the electrostatic and magnetic lens components that provide correction for chromatic aberrations of off-axis beams, which are different from classical beams, are obtained. Phase contours at the lens exit are calculated for beams occupying the major part of the aperture, and conditions for transforming the phase contour with minimal distortions are found. © 2001 MAIK “Nauka/Interperiodica”.

In our earlier paper [1], a modified combined quadrupole lens capable of simultaneously correcting for chromatic and geometric aberrations was proposed. This lens has the form of a cylinder split along its generatrices into eight equal parts. Electrostatic potentials applied to the lens sections form the quadrupole and octupole components of the lens. The four electrodes located between the quadrupole electrodes are simultaneously the pole pieces of the magnetic lens. This type of lens system can be used in charged particle transport channels, microanalyzers, electron microscopes, and other devices. In this paper, we obtain an analytical expression for the scalar potential of the combined lens with infinitesimally narrow gaps between electrodes and poles.

The purpose of this work is to theoretically study the electromagnetic quadrupole–octupole lens described above in order to provide a distortion-free transmission of wide charged-particle beams and of narrow beams passing far from the axis. Numerous previous works use the aberration theory, which is valid at best for beams that occupy half of the aperture at most. In our approach, the beam can occupy almost the entire aperture, because we calculate the trajectories based on analytical formulas that we obtained for the electrostatic and magnetic fields.

In Cartesian coordinates, the potential of the above finite-length lens can be written as

$$\begin{aligned} \Phi_{FM}(x, y, z) = & 1/\pi k(z) \\ & \times \{ (W - V) \arctan[\sqrt{2}(x^2 - y^2 + 2xy)/[1 - (x^2 + y^2)^2]] \\ & - (W + V) \arctan[\sqrt{2}(x^2 - y^2 - 2xy)/[1 - (x^2 + y^2)^2]] \\ & + 2U \arctan[2(x^4 + y^4 - 6x^2y^2)/[1 - (x^2 + y^2)^4]] \}. \end{aligned}$$

Here and below, the coordinates are normalized by the radius R of the lens aperture; $\pm V$ and $\pm U$ are the potentials of the electrostatic quadrupole and octupole,

respectively; and $\pm W$ is the scalar potential of the magnetic quadrupole. The function $k(z)$ describes the field distribution along the optical axis, which is the same for the electric and magnetic components. This circumstance is not significant, because it does not restrict the generality of further conclusions, although it simplifies the calculations. The function $k(z)$ is specified numerically or is approximated by an appropriate analytical function. In particular, the field in short lenses is most frequently represented by a Gaussian approximation; in sufficiently long lenses, by a rectangular function.

Parameters of the charged particle beam trajectories are calculated from second-order differential equations of motion. To solve these equations, we developed a computer code EMKOL using MathCAD. This code reduces each second-order differential equation to a system of two first-order differential equations and solves them with an error below 10^{-6} . The code is capable of predicting trajectories in a single lens and in a system of lenses as well.

As an example, we studied an electromagnetic quadrupole–octupole lens with a rectangular field distribution with an effective length of $L = 4R$. Focusing modes for the parallel beam were calculated, which are necessary for using the lens in imaging systems. Phase contours of wide beams at the exit of the lens were also calculated, which is of interest for concentrating systems.

For electrostatic quadrupole lenses of different strengths with parallel entrance beams, Fig. 1 shows the distances from the field edge to the points of intersection between the lens axis and trajectories as well as the angles with which the trajectories leave the lens. For small entrance point offsets from the axis $x_0 < 0.1R$, this distance determines the focal length of a lens of a particular strength, and it coincides with the distance calculated from the formula given in [2]. Figure 1 shows that, as the entrance point offset increases, the trajectory–axis intersection point approaches the lens and the

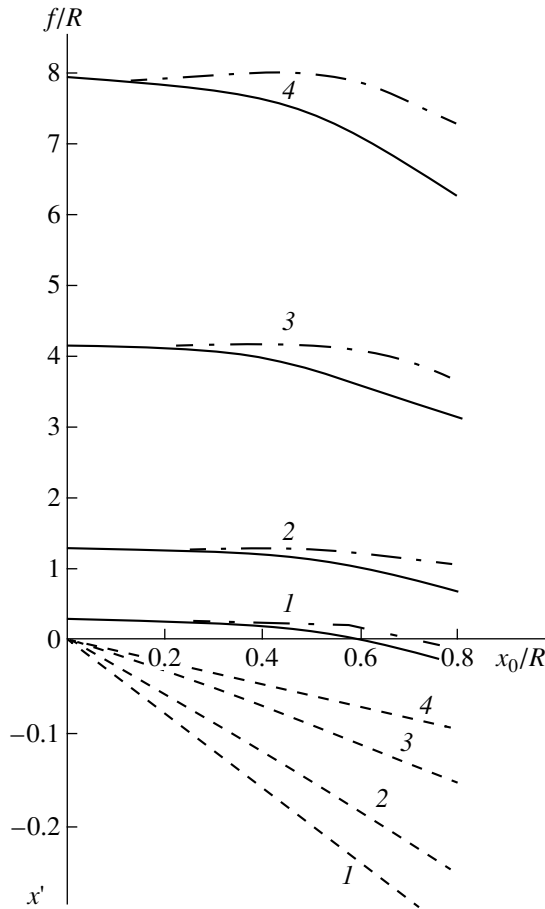


Fig. 1. Coordinate of the trajectory–axis intersection point for the quadrupole lens (solid lines) and the angle with which the trajectories leave the lens (dashed lines) vs. the entrance point coordinate of the parallel beam at $V/\Phi_0 =$ (1) 0.15, (2) 0.10, (3) 0.05, and (4) 0.03. Dash-and-dot lines refer to the quadrupole–octupole lens with $U = 0.1V$.

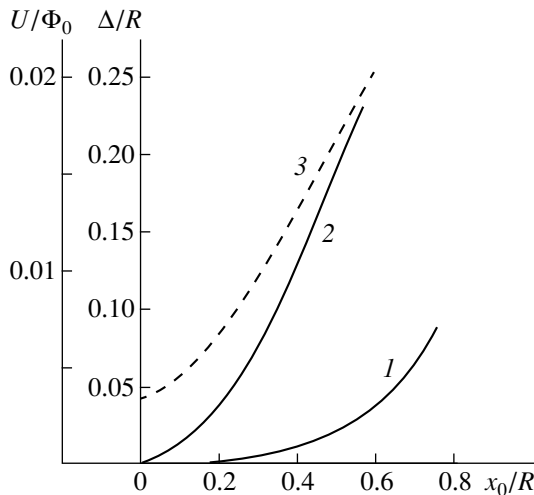


Fig. 2. (1) Transverse and (2) longitudinal spot expansion and (3) strength of the correcting octupole vs. the entrance point coordinate of the central beam trajectory.

trajectory inclination angle grows. The lower the lens strength, the greater the longitudinal spot extent. The transverse focus blur is almost independent of the lens strength and remains below $0.1R$.

A narrow parallel entrant beam is generally focused along a curvilinear trajectory. Figure 2 illustrates the transverse and longitudinal beam extension at the point the most remote trajectory intersects the axis versus the initial coordinate of the central beam trajectory. The entrance beam width is $\Delta x_0 = 0.2R$. It is seen that the transverse focus blur is an order of magnitude smaller than the longitudinal blur. It should be noted that, for annular entrance beams, the longitudinal spot extension remains the same, while the transverse one increases twofold.

To improve the beam transmission conditions, we used the octupole. We studied wide beams that occupy up to 0.8 of the lens aperture and narrow beams that are far offset from the axis. In the first example, the octupole potential is $U = 0.1V$. In this case, the change in the position of the beam–axis intersection point with the entrance coordinate is less rapid (dash-and-dot line in Fig. 1) and the longitudinal displacement is more than two times smaller than that for the quadrupole lens. The transverse dimension of the spot in the focal plane is two times smaller. In the second example, the spot blur is corrected by the octupole component so as to make the focal distance independent of the beam entrance point coordinate. Curve 3 in Fig. 2 is a plot of the octupole strength U/Φ_0 at the quadrupole strength of $U/\Phi_0 = 0.1$, where Φ_0 is the accelerating potential. The transverse beam dimension is two orders of magnitude smaller than that for the quadrupole lens.

It is known that the chromatic aberration in the paraxial region of the combined quadrupole electromagnetic lens can be corrected. If the electric and mag-

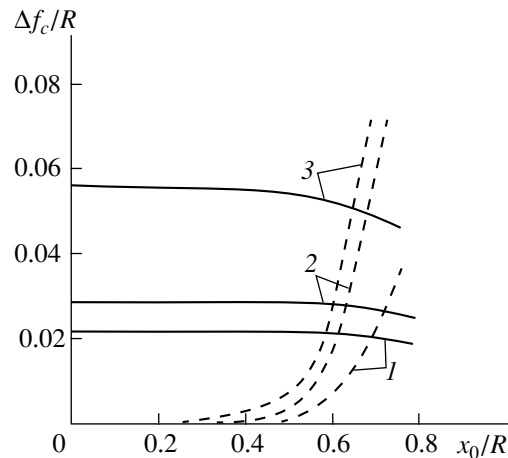


Fig. 3. Longitudinal chromatic aberration in the electrostatic (solid lines) and combined lenses (dashed lines) vs. the coordinate of the beam entrance point with the lens strength as a parameter: (1–3) same as in Fig. 1.

netic fields are colocated, their strengths are interrelated as $G_E = -0.5G_M$ [2], where $G_E = V/\Phi_0$ and $G_M = W/c\sqrt{-2e/m\Phi_0}$. Then, the strength of the achromatic combined quadrupole lens is $G_{ME} = G_M + G_E = -G_E$. However, as the beam moves away from the axis, a noticeable chromatic aberration reveals itself. Figure 3 illustrates the longitudinal chromatic aberration in the electrostatic lens and in the combined lens meeting the paraxial beam correction condition; the initial beam energy varies within 1%. The chromatic aberration in the electrostatic lens is seen to vary slightly with the entrance point position. In the achromatic lens, the region of correction depends on the lens strength. Outside this region, the chromatic aberration rapidly grows and, when $x_0 > 0.6R$, reaches values typical of the electrostatic lens.

We determined the electrostatic-to-magnetic field strength ratios at which the region of chromatic aberration correction becomes wider. Figure 4 shows this ratio for the parallel beam and the residual chromatic aberration for the initial beam energy spread within 1% as a function of the beam entrance point position. These quantities were found to be almost independent of the lens strength over a sufficiently wide range. The comparison of Figs. 3 and 4 shows that the residual chromatic aberration in lenses with a refined G_E/G_{ME} ratio is below 10^{-3} when the distance from the axis is less than 0.7 of the lens radius, while for the classical achromatic lens, it is one to two orders of magnitude higher within 0.6–0.7 of the aperture radius, depending on the lens strength.

We calculated the coordinate of the trajectory–axis intersection point for the achromatic lens, as well as the angle with which the beam leaves the lens (Fig. 5). It should be noted that, in lenses with a refined electrostatic-to-magnetic component ratio, this coordinate is almost the same as in the electrostatic lens (compare Figs. 1, 5). For a lens which is achromatic in the paraxial approximation, it rapidly grows as the trajectory moves away from the axis, while the behavior of the trajectory inclination angle versus the entrance coordinate is significantly different from a linear function. These results reveal that geometrical aberrations in the classical achromatic lens are much greater than those in the improved lens.

The quadrupole–octupole electromagnetic lens is capable of concentrating wide beams with large angles of inclination. To study these modes, one should investigate how the lens transforms different phase contours of the entrance beams. As an example, consider an elliptic entrance phase contour with semiaxes $x_0 = 0.7R$ and $x'_0 = 0.35$. Figure 6 shows halves of the phase contours at the exit of the following lenses. On the left are the electrostatic and magnetic quadrupoles of equal strengths ($G_E = G_M = 0.1$). On the right are the classical achromatic lens ($G_E = 0.1, G_E/G_{ME} = -1$), the improved achromatic quadrupole ($G_E = 0.1, G_E/G_{ME} = -0.42$),

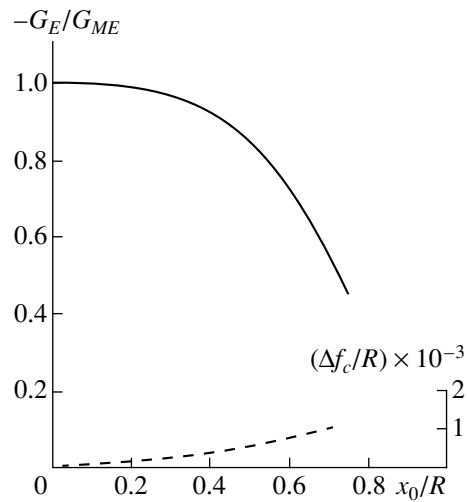


Fig. 4. Ratio of the electrostatic component to the total strength of the electromagnetic lens that provides the condition for the lens to be achromatic over a wide range of initial conditions (solid line); the dashed line shows the residual chromatic aberration of the electromagnetic lens.

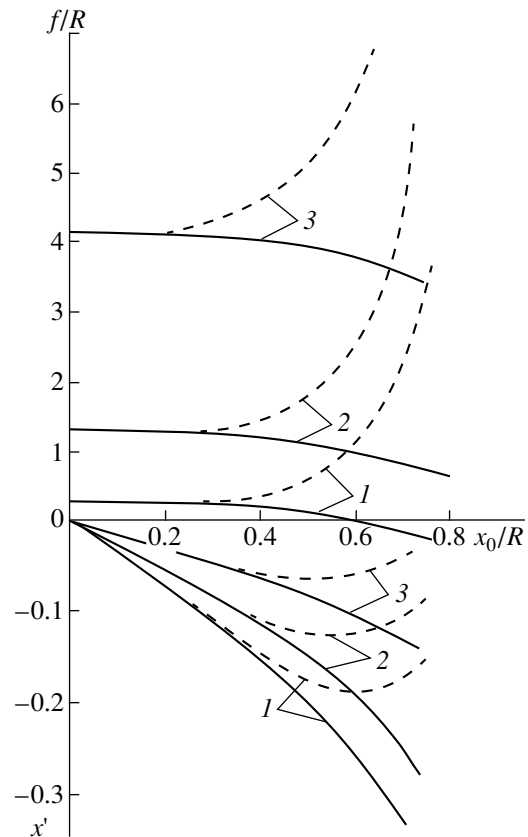


Fig. 5. Coordinate of the trajectory–axis intersection point for achromatic lenses and the angle with which the trajectories leave the lens: solid lines represent the electromagnetic lens with the strength shown in Fig. 4 and dashed lines represent the classical achromatic lens; (1–3) same as in Fig. 1.

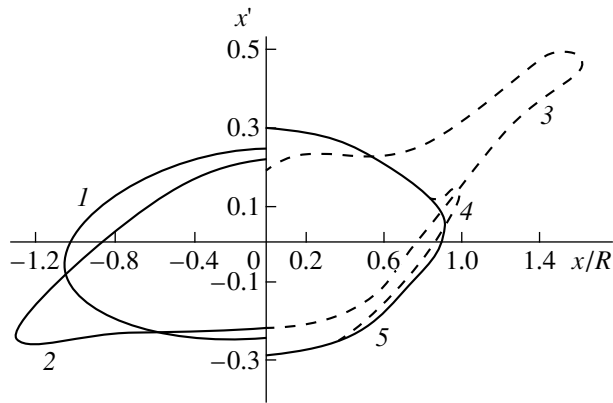


Fig. 6. Phase contours at the exit of different quadrupole lenses: (1) electrostatic lens; (2) magnetic lens; (3) classical achromatic lens; and (4, 5) improved achromatic electromagnetic quadrupole and quadrupole–octupole lenses, respectively.

and the quadrupole–octupole ($G_E = 0.1$, $G_E/G_{ME} = -0.55$, $U/\Phi_0 = -0.046$) lenses. This figure shows that the classical achromatic lens introduces the greatest distortions into the phase ellipse. As for the intensity of

the exit beam, it peaks in the improved achromatic quadrupole and quadrupole–octupole lenses. The comparison of the improved achromatic lenses have shown that the octupole is free from the phase ellipse distortions typical of the quadrupole lens. However, the octupole features somewhat higher chromatic aberrations, though they remain two times lower than those in the electrostatic lens.

We have thus shown that the above approach improves the performance of the quadrupole lens for charged particle beams occupying the major part of the aperture. The lens has a simple design and its parameters are the same as or better than those of quadrupole lenses combined with paraxial beams.

REFERENCES

1. L. P. Ovsyannikova and T. Ya. Fishkova, *Pis'ma Zh. Tekh. Fiz.* **25** (21), 5 (1999) [*Tech. Phys. Lett.* **25**, 847 (1999)].
2. S. Ya. Yavor, *Focusing of Charged Particles by Quadrupole Lenses* (Atomizdat, Moscow, 1968).

Translated by A. Khzmalyan

Thermodynamics and Kinetics of Shear-Induced Melting of a Thin Layer of Lubricant Confined between Solids¹

V. L. Popov

*Institute of Strength Physics and Materials Science, Siberian Branch, Russian Academy of Sciences,
Tomsk, 634021 Russia*

e-mail: popov@phys.uni-paderborn.de

Received August 28, 2000

Abstract—Using Landau’s phenomenological theory of phase transitions, the shear-induced melting of a thin layer of substance confined between two crystalline surfaces is considered. The kinetics of melting and solidification is considered for static and alternating loads. The possibility of two consecutive “melting” (solid-to-liquid) transitions is discussed. As a result of the first transition, modulation of the microscopic density becomes zero only in the direction of shear (partial melting) and as a result of the second, it also disappears in the normal direction (complete melting). © 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

The development of experimental tools for studies of frictional processes at the atomic level (atomic force microscope [1], quartz microbalance [2], surface force apparatus [3]), as well as the development of computer techniques, led to a sharp increase in the last decade in the number of studies in the field of the friction between solids either in direct contact or separated by a thin lubricating layer (see, for example, [4–6]). Experimental physicists can now investigate well-defined systems under closely controlled conditions (for instance, the possibility of measuring the thickness of the lubricant layer and the relative displacement of the solid surfaces to within fractions of the interatomic distances). As a result, frictional process, formerly a purely technical problem, for the first time became an object of extensive basic research.

One of the findings of these studies is that a layer of lubricant several atomic layers thick can behave like a solid [4]. Such thin layers are by no means exotic but exist practically in any friction unit operating in the so-called boundary lubrication regime. When two solids are pressed against each other (or two microirregularities are brought into close contact), the thickness of the lubricant layer steadily decreases until about 10 molecular layers are left. This layer demonstrates considerable elasticity and remains trapped between the two surfaces if the pressure is not too high [4]. At this point, it tends to form either a layered structure in the direction perpendicular to the bounding surfaces, or a long-range crystalline or quasi-crystalline order in the plane of the layer (experimental investigations of these aspects can be found in [7, 8] and a study by the molecular dynamics method in [6, 9]).

Computer simulation of the slip under shear stress of solids separated by a thin layer of the solidified “liquid” [6, 9] has shown that the slip occurs often because of the melting of the intervening layer. For instance, in thin crystallized layers confined between atomically flat surfaces of mica and loaded mechanically, the shear modulus steadily decreases as the shear deformation is increased and ultimately becomes zero; when the load is removed, it increases continuously from zero up [7]. Obviously, in this case, melting occurs by the second order phase transition. It should be emphasized that this is the case of local melting of a layer confined between two crystalline solids retaining their crystalline structure. It is evident that in this case, the energy of the crystalline state of the lubricating layer will be different for different orientations of the crystal lattice. It can be shown that the standard Landau argument that the melting at a second order phase transition is impossible fails in this case (this problem is discussed in more detail in the last section of this paper).²

It is essential that both the process of shear-induced melting and the process of crystal structure recovery are not instantaneous but have a character of slow relaxation. High-molecular compounds studied in [7], due to confinement of the system by solids and steric limitations, displayed shear modulus relaxation times between 1 and 10 s. At times much less than the above characteristic time, the medium behaves as an abso-

¹ This study was performed during the author’s stay at the Forschungszentrum Jülich (Germany).

² The possibility of a thin layer melting at the second order phase transition does not contradict Landau’s conclusion that melting is always the first-order phase transition [14], since the latter is valid only for unbounded media and transitions from the state with a discrete symmetry group into a state which is isotropic and invariant with respect to rotations. In the case considered here, the isotropy, even though the medium is in a liquid state, is distorted both by the confining surfaces and by the presence of a selected direction in the plane of the layer (the direction of deformation).

lutely ideal linearly elastic body (in particular, the response contains only the basic frequency under a periodic external action at a fixed frequency).

It is obvious that in the study of these layers over the time range given below, the characteristic time of the shear modulus relaxation using only temperature and stress as thermodynamic parameters is inapplicable and that a more detailed study of nonequilibrium kinetics of the system should be undertaken. The slowness of the shear modulus variation allows for studying the kinetics within the framework of the macroscopic phenomenological theory. Basically, this theory considers the states of partial thermodynamic equilibrium at a specified (slowly varying) order parameter. For considering the states of partial equilibrium as having a real physical meaning, it is necessary that the kinetics of the order parameter be slow. The shear modulus displays just this property as seen in the experiment. This is obvious from the theoretical point of view as well, because the shear modulus is a result of long-range correlations in the system and can change only by the staged mutual adjustment of atoms, which in any case requires much more time than, for example, the displacements occurring in the propagation of an elastic wave.

In this paper we propose a theoretical description of the melting of the mentioned thin intermediate layer in the framework of Landau's theory of phase transitions [10].

A MELTING PHASE TRANSITION INDUCED BY TEMPERATURE AND SHEAR

The starting point of Landau's thermodynamics theory of phase transitions is the choice of the order parameter φ , which is a quantity characterizing the qualitative changes of the body state at the point of phase transition. In the case of melting, a natural candidate for the role of the order parameter is the limiting value of the pair correlation function at infinity or of the amplitude of the periodic part of the microscopic function of medium density, which is equivalent. In a liquid state, the microscopic function of medium density is homogeneous and the order parameter is equal to zero, while in the crystalline state the density function has a nonzero periodic component accounting for the presence of the crystal lattice. The stability of the medium with respect to shear is closely related to the order parameter defined in this way, the statement that the medium is in a crystalline state simply meaning that any two neighboring molecular layers in this state create periodic potentials for each other. The melting means the vanishing of the periodic component of the interaction potential. In the case of purely harmonic modulation of the density function, it can be shown that the shear modulus μ is proportional to the square of the order parameter

$$\mu = \varphi^2. \quad (1)$$

At temperatures in the vicinity of the phase transition temperature, the expansion of the free energy density in powers of the order parameter has the form

$$f_{\text{therm}} = a\varphi^2 + \frac{b}{2}\varphi^4. \quad (2)$$

Here, a and b are functions of the thermodynamic state parameters, namely, of temperature and pressure. In the vicinity of a critical temperature T_c ,

$$a(T) = \alpha(T - T_c). \quad (3)$$

Function (2) describes the system in a state of minimum energy, where there are no elastic deformations. In the crystalline phase due to nonvanishing transverse rigidity of the system, the medium can be transformed into a state of metastable strain. For description of this state, it is necessary to add to the free energy Eq. (2) the free energy of the elastic deformation $\mu\varepsilon_{\text{el}}^2/2$. In a spatially nonuniform system the expansion in terms of gradients of the order parameter should be considered as well. Thus, taking Eq. (1) into account, the total free energy density can be written in the form

$$f = \alpha(T - T_c)\varphi^2 + \frac{b}{2}\varphi^4 + \frac{1}{2}\varphi^2\varepsilon_{\text{el}}^2 + \frac{g}{2}(\nabla\varphi)^2, \quad (4)$$

where ε_{el} is the elastic deformation.³

The elastic stress in the layer is defined as

$$\sigma_{\text{el}} = \frac{\partial f}{\partial \varepsilon_{\text{el}}} = \varphi^2\varepsilon_{\text{el}} = \mu\varepsilon_{\text{el}}. \quad (5)$$

At $T < T_c$ and $\varepsilon_{\text{el}} = 0$, the coefficient by the second power of the order parameter is negative, meaning that the body is in the solid state with a shear modulus:

$$\mu = \varphi^2 = \frac{\alpha(T_c - T)}{b} \neq 0. \quad (6)$$

As the elastic distortions increase, the absolute value of the coefficient by the second-order term becomes less. Accordingly, the order parameter and shear modulus decrease:

$$\mu = \varphi^2 = \frac{\alpha(T_c - T) - \varepsilon_{\text{el}}^2/2}{b}. \quad (7)$$

At a certain deformation value

$$\varepsilon_{\text{el},c} = \sqrt{2\alpha(T_c - T)} \quad (8)$$

the transverse rigidity of the medium becomes zero. The elastic stress as a function of the elastic deformation

³ Note that the free energy (4) has absolutely the same form as the Ginzburg–Landau free energy in the theory of superconductivity. The elastic deformation in (4) corresponds to the phase gradient in the Ginzburg–Landau theory.

tion is given by

$$\begin{aligned} \sigma_{el} &= \mu \varepsilon_{el} = \Sigma(\varepsilon_{el}) \\ &= \begin{cases} \frac{\alpha(T_c - T) - \varepsilon_{el}^2/2}{b} \varepsilon_{el}, & \varepsilon_{el} < \varepsilon_{el,c}, \\ 0, & \varepsilon_{el} > \varepsilon_{el,c}. \end{cases} \end{aligned} \quad (9)$$

At $\varepsilon_{el} = \varepsilon_{el}^*$, where

$$\varepsilon_{el}^{*2} = \frac{2}{3} \alpha(T_c - T) = \frac{1}{3} \varepsilon_{el,c}^2, \quad (10)$$

it reaches its maximal value,

$$\sigma_0 = \frac{1}{b} \left(\frac{2}{3} \alpha(T_c - T) \right)^{3/2}. \quad (11)$$

Expression (9) is plotted in Fig. 1 with the experimental results taken from [7]. Note that in the experiment this dependence can only be followed up to the stress maximum since the falling branch of the $\sigma(\varepsilon_{el})$ dependence corresponds to the unstable states of the system. The layer subjected to gradually increasing shear stress will loose its stability with respect to shear when a critical shear value given by Eq. (1) has been reached. For analysis of its further dynamics, it is necessary to take into consideration plastic deformation.

PLASTIC AND VISCOUS LAYER PROPERTIES AT FINITE DEFORMATIONS

If a plastic deformation is present in a medium, the elastic deformation can be represented as the difference between the total ε and the plastic ε_{pl} deformations

$$\varepsilon_{el} = \varepsilon - \varepsilon_{pl}. \quad (12)$$

If the plastic deformation varies with time,⁴ then additional viscous stress is generally associated with it:

$$\sigma_{visc} = \eta \dot{\varepsilon}_{pl}, \quad (13)$$

where η is the dynamic viscosity.

The total stress in the layer is

$$\sigma = \sigma_{el} + \sigma_{visc}. \quad (14)$$

It can be shown that the plastic flow in the elastic state proceeds by thermoactivated nucleation of the critical dislocation rings, whose energy U_0 is inversely proportional to the applied stress:

$$U_0 = A(T)/\sigma. \quad (15)$$

Having overcome this barrier, the dislocations move with high velocities determined by the viscous friction arising from interaction with the phonons and electrons [11], which are approximately in proportion to the

⁴ Plastic deformation can develop in an elastic layer at finite elastic deformations because the state of elastical deformation is metastable and at finite temperatures thermoactivated relaxation of the elastic stress is possible.

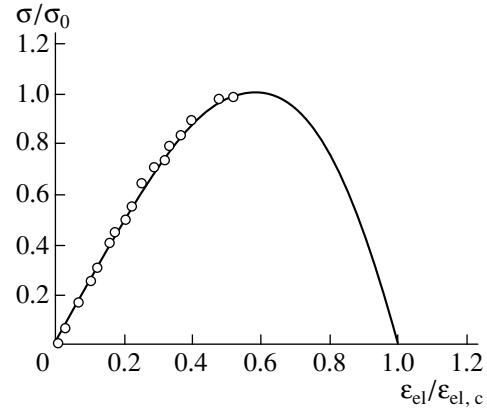


Fig. 1. Dependence of the elastic shear stress on the elastic deformation in the layer. The circles show experimental data [7].

applied stress. Accordingly, the macroscopic rate of the plastic deformation is given by an expression of the form

$$\dot{\varepsilon}_{pl} \propto \sigma e^{\frac{A(T)}{\sigma k T}}. \quad (16)$$

This expression shows that we are indeed dealing with an ideally elastic body in the sense that the rate of plastic deformations identically becomes zero in the linear as well as in all finite orders of the stress (the point $\sigma = 0$ is an essentially singular point of expression (16)). Still, at finite stress there is always some finite rate of plastic deformations.

It can be shown that the coefficient $A(T)$ vanishes at the critical temperature. This means that at the point of the transition from the solid to the liquid state (at $\varepsilon_{el} = \varepsilon_{el,c}$), the rate of the plastic deformation rate varies continuously. Therefore, in the qualitative analysis, we assume that between the elastic and plastic deformations, a linear relaxation relationship exists in the vicinity of the transition point

$$\dot{\varepsilon}_{pl} = \varepsilon_{el}/\tau, \quad (17)$$

where τ is a phenomenological coefficient having the meaning of the characteristic time of the internal stress relaxation (Maxwell relaxation time).

TRANSITION OF THE LAYER FROM AN ELASTIC TO A VISCOUS STATE

Using Eqs. (5), (7), (9), and (17) we write

$$\sigma = \Sigma(\varepsilon_{el}) + \frac{\eta}{\tau} \varepsilon_{el}. \quad (18)$$

The plot of this dependence is presented schematically in Fig. 2 for two different values of the parameter η/τ corresponding to two possible scenarios of the shear-induced melting. In the first case, as the stress increases, a steplike melting of the layer occurs at

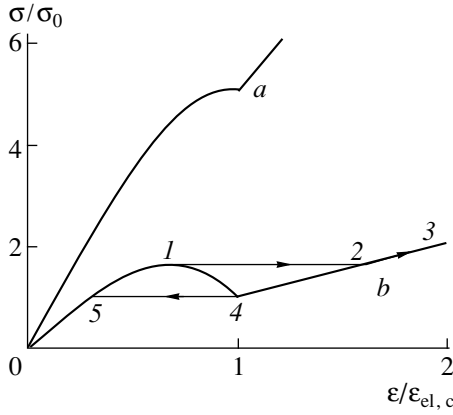


Fig. 2. Dependences of the shear stress on the elastic deformation in a layer undergoing a melting phase transition. (a) Large viscosity $\left(\frac{\eta \varepsilon_{el,c}}{2\tau \sigma_0} = 20\right)$, (b) low viscosity $\left(\frac{\eta \varepsilon_{el,c}}{2\tau \sigma_0} = 4\right)$.

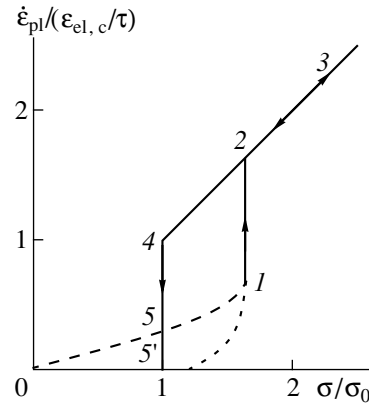


Fig. 3. The plastic flow rate as a function of the shear stress in a loaded/unloaded layer as given by Eqs. (18) and (17) (solid curve). With the use of a more realistic law of plastic flow in form (16), the dependence shown schematically by the dotted line results.

point 1 and the system goes to point 2, corresponding to viscous flow. A further increase in stress results in a monotonic increase of the deformation rate. If the stress is then decreased, the layer remains in the liquid state down to point 4, where it solidifies abruptly (transition to point 5). The corresponding diagram with the coordinates $(\dot{\varepsilon}_{pl}, \sigma)$ is shown in Fig. 3. A similar hysteretic behavior was observed experimentally [7, 8] and in computer simulations [6].

At very high viscosity (or short relaxation time τ), there can be no extrema in the function $\sigma(\varepsilon_{el})$ as shown in Fig. 2 (curve *a*). In this case, the melting proceeds continuously and results in a continuous and monotonic increase of the plastic deformation rate with stress.

FLUCTUATIONAL CONTRIBUTION TO THE ELASTIC STRESS IN THE LIQUID PHASE

It is interesting to note that strong fluctuations of the order parameter in the vicinity of the transition point produce an elastic contribution to the stress above the transition point as well, i.e., in the liquid phase. Indeed, by averaging Eq. (5) we obtain

$$\sigma_{el, fluct} = \langle \varphi^2 \rangle \varepsilon_{el}. \tag{19}$$

The average value of $\langle \varphi^2 \rangle$ in the liquid phase outside the fluctuation region is given by the following expression [10]:

$$\langle \varphi^2 \rangle = \int_{-\infty}^{\infty} \langle |\varphi_k|^2 \rangle \frac{Ldk}{2\pi}, \tag{20}$$

where

$$\langle |\varphi_k|^2 \rangle = \frac{k_B T}{L2[\alpha(T - T_c) + gk^2]}, \tag{21}$$

k_B is the Boltzmann constant, and L is the system extent in the shear direction.

Equations (19)–(21) determine the contribution of fluctuations to the total stress above the critical temperature. A similar contribution is added below the phase transition point. Note that under externally applied stress, stress fluctuations in the directions perpendicular to shear are suppressed because of a symmetry violation. That is why the integration in Eq. (20) is carried out along only one component of the wave vector (in the shear direction). Substitution of Eq. (21) into Eq. (20) results in the following expression for the fluctuation component of the order parameter:

$$\langle \varphi^2 \rangle = \frac{k_B T}{4\sqrt{g\alpha(T - T_c)}}. \tag{22}$$

Note that the fluctuations of the force of friction have principally different character at the direct and reverse transitions. With increasing stress, the static elastic deformation changes stepwise from the value corresponding to the maximum stress in the elastic state to that corresponding to viscous flow of the layer (transition from point 1 to point 2 in Fig. 2). Therefore, the layer can be found in the vicinity of the phase transition point by any stress. When the stress decreases, solidification of the layer occurs at the point corresponding to the phase transition to the elastic state. In the vicinity of this point, the fluctuations should abnormally increase, and at the transition point, the fluctuation should be scale invariant [10]. These conclusions are completely supported by the measurements of the friction force fluctuations in the vicinity of the layer solidification transition [12].

Let us find the temperature dependence of the plastic deformation rate for small stress and temperatures above the phase transition temperature. In this case, the stress is given by the expression

$$\sigma = \langle \varphi^2 \rangle \varepsilon_{el} + \frac{\eta}{\tau} \varepsilon_{el} = \frac{k_B T}{4\sqrt{g\alpha}(T - T_c)} \varepsilon_{el} + \frac{\eta}{\tau} \varepsilon_{el}. \quad (23)$$

For the plastic deformation rate, by virtue of Eq. (17), we obtain

$$\dot{\varepsilon}_{pl} = \frac{\sigma}{\eta} \frac{1}{1 + \sqrt{t_0/(T - T_c)}}, \quad (24)$$

where

$$t_0 = \frac{(\tau k_B T_c)^2}{16\eta^2 g \alpha}. \quad (25)$$

The dependence (24) is presented in Fig. 4. Similar dependences have been observed in computer experiments [6].

KINETICS OF THE ORDER PARAMETER EQUILIBRATION

If the temperature or deformation of the layer changes stepwise, a certain time is then necessary for the order parameter to attain a new equilibrium value. The kinetics of the order parameter can be determined from the following considerations. The derivative $-\partial f/\partial \varphi$ determines a generalized thermodynamic force, which makes the order parameter approach its equilibrium value. In the vicinity of the phase transition point, this derivative is small and the kinetic equation will be linear in thermodynamic force:

$$\dot{\varphi} = -\gamma \frac{\partial f}{\partial \varphi}, \quad (26)$$

where γ is the kinetic coefficient, which has a dimension of frequency in our case.

Its explicit form is

$$\dot{\varphi} = -\gamma(2\alpha(T - T_c)\varphi + 2b\varphi^3 + \varphi \varepsilon_{el}^2(t)). \quad (27)$$

Using this kinetic equation, let us analyze the experimental results obtained in [7, 13]. In [7], thin layers of liquid lubricants were investigated with molecules in the form of linear chains consisting of 3-methyl-undecane ($=\text{CH}_3\text{-C}_{11}$), alkane with eleven CH_2 groups with one methyl group near the third carbon atom, and perfluoro-heptalyne [$\text{CF}_3\text{O}(\text{CF}_2\text{CF}_2\text{O})_7\text{CF}_3$]. A thin layer of lubricant was placed between two atomically flat mica surfaces, one of which was bent to form a cylinder and glued to a glass lens. The mica surfaces were set into periodic motion relative to each other at frequencies from 1 to 250 Hz, and the force and relative displacement of the surfaces were measured. In the elastic region, the response of the layer during one period was almost ideally linear. At the same time, the response

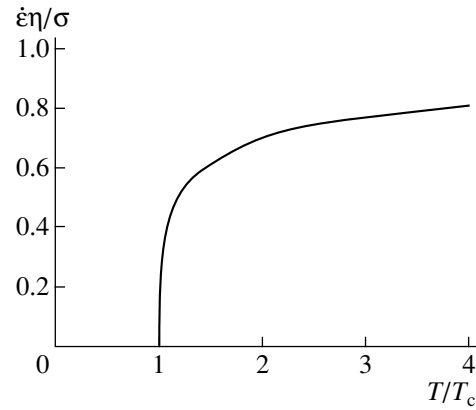


Fig. 4. The plastic flow rate as a function of temperature in the vicinity of the melting temperature.

varied smoothly from period to period with a characteristic relaxation time from 10 to 30 s for $\text{CH}_3\text{-C}_{11}$. The equilibrium value of the elastic modulus decreased monotonically as the oscillation amplitude was increased. It is evident that in the case considered the order parameter relaxation time considerably exceeded the period of oscillations. The relaxation time was measured in experiments with other polymeric lubricants, for instance, squalane $\text{C}_{30}\text{H}_{62}$, which has globular molecules in contrast lower than the period of oscillations, so the load could be considered quasistatic [13].

Let us show that the root-mean-square value of the steady-state order parameter as a function of the root-mean-square of the elastic deformation can be obtained in general form irrespective of the relation between the period of external force and the order parameter relaxation time.

Dividing Eq. (27) by φ and taking into account that $\dot{\varphi}/\varphi \equiv \partial \ln \varphi / \partial t$, Eq. (27) can be represented in the form

$$\frac{\partial \ln \varphi}{\partial t} = -\gamma(2\alpha(T - T_c) + 2b\varphi^2 + \varepsilon_{el}^2(t)). \quad (28)$$

If $\varepsilon_{el}(t)$ is a periodic function of time, then the solution of Eq. (28) will be an asymptotically periodic function of time. Averaging Eq. (28) over one period and taking into account that $\langle \partial \ln \varphi / \partial t \rangle = 0$ due to the periodicity of φ , we obtain

$$\langle \varphi^2 \rangle = \frac{\alpha(T_c - T) - \langle \varepsilon_{el}^2 \rangle / 2}{b}. \quad (29)$$

At $\langle \varepsilon_{el}^2 \rangle$ exceeding the critical value $2\alpha|T_c - T|$, the average value is $\langle \varphi^2 \rangle \equiv 0$. Equation (29) is similar to Eq. (9); however, in the case of periodic action, it relates the time-averaged values $\langle \varphi^2 \rangle$ and $\langle \varepsilon_{el}^2 \rangle$. We see that under an alternating applied force, the layer undergoes the same melting phase transition as in the case of the static external force. At the transition point in this

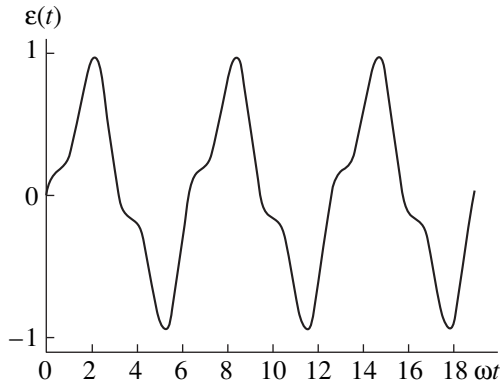


Fig. 5. Typical response of the layer to an oscillating stress. Asymmetry is caused by a third harmonic in the response of the layer (see Eq. (34)).

case the amplitude ε_{el} exceeds the critical value for the transition $\varepsilon_{el,c}$ under static load. Exactly this behavior is observed experimentally [7].

If the frequency is much higher than the reciprocal of the relaxation time, then the solution of Eq. (27) or of the equivalent Eq. (28) can be easily obtained in explicit form. So, for the first-order approximation in the parameter γ/ω , we obtain

$$\varphi(t) = \varphi_0 \left\{ 1 - \gamma \int_0^t (\varepsilon_{el}^2(t) - \langle \varepsilon_{el}^2 \rangle) dt \right\}. \quad (30)$$

Substituting

$$\varepsilon_{el}(t) = \sqrt{2\langle \varepsilon_{el}^2 \rangle} \sin \omega t, \quad (31)$$

we find

$$\varphi(t) = \varphi_0 \left\{ 1 - \frac{\gamma}{2\omega} \langle \varepsilon_{el}^2 \rangle \sin 2\omega t \right\}. \quad (32)$$

Obviously, this expansion is true if

$$\frac{\gamma}{2\omega} \langle \varepsilon_{el}^2 \rangle \ll 1. \quad (33)$$

The stress $\sigma(t) = \varphi^2(t)\varepsilon(t)$ in the linear approximation in the parameter $\gamma/2\omega$ is given by the expression

$$\sigma(t) = \varphi_0^2 \sqrt{2\langle \varepsilon_{el}^2 \rangle} \left[\sin \omega t + \frac{\gamma}{2\omega} \langle \varepsilon_{el}^2 \rangle \cos \omega t - \frac{\gamma}{2\omega} \langle \varepsilon_{el}^2 \rangle \cos 3\omega t \right]. \quad (34)$$

It has a component in phase with $\varepsilon(t)$ and a component out of phase by $\pi/2$, which determine the elastic (μ) and dissipative (μ') parts of the shear modulus,

respectively:

$$\begin{aligned} \mu &= \varphi_0^2, \\ \mu' &= \varphi_0^2 \frac{\gamma}{2\omega} \langle \varepsilon_{el}^2 \rangle = \mu \frac{\gamma}{2\omega} \langle \varepsilon_{el}^2 \rangle. \end{aligned} \quad (35)$$

If the stress is taken in the form of a harmonic function

$$\sigma = \sqrt{2} \langle \sigma^2 \rangle \sin \omega t, \quad (36)$$

then for the deformation, we obtain

$$\begin{aligned} \varepsilon(t) &= \frac{\sqrt{2\langle \sigma_{el}^2 \rangle}}{\varphi_0^2} \left[\sin \omega t - \frac{\gamma}{2\omega} \frac{\langle \sigma_{el}^2 \rangle}{\varphi_0^4} \cos \omega t \right. \\ &\quad \left. + \frac{\gamma}{2\omega} \frac{\langle \sigma_{el}^2 \rangle}{\varphi_0^4} \cos 3\omega t \right]. \end{aligned} \quad (37)$$

Note the presence in the response of the third harmonic, which has the same amplitude as the dissipative component of deformation. The occurrence in the response of the squalane layer under periodic loading of a well pronounced third harmonic (with the phase coinciding with that given by Eq. (37)) was established in [13]. A typical response at $\gamma \langle \sigma_{el}^2 \rangle / 2\omega \varphi_0^4 = 1/3$ is presented in Fig. 5.

ORDER PARAMETER RELAXATION AT FIXED STRESS

In experiments, it is often not the displacement but the externally applied stress that is fixed. Expressing from the condition of constant stress,

$$\sigma = \varphi^2 \varepsilon_{el} = \text{const}, \quad (38)$$

the elastic deformation as a function of the order parameter and substituting it into Eq. (27), we obtain the kinetic equation

$$\dot{\varphi} = -\gamma \left[2\alpha(T - T_c)\varphi + 2b\varphi^3 + \frac{\sigma^2}{\varphi^3} \right], \quad (39)$$

which determines the order parameter dynamics at a given stress. Introducing the dimensionless variables

$$\tilde{\varphi} = \varphi/\varphi_0, \quad \tilde{\sigma} = \sigma/\sigma_0, \quad \tilde{t} = t/t_0, \quad (40)$$

where $\varphi_0 = (\alpha(T_c - T))^{1/2}$ is the equilibrium value of the order parameter in the absence of deformation (see Eq. (6)), $\sigma_0 = (1/b)(2/3\alpha(T_c - T))^{3/2}$ is the maximal attainable steady-state elastic stress in the layer given by Eq. (11), and $t_0 = (2\alpha(T_c - T)\gamma)^{-1}$ is the characteristic time of the order parameter relaxation to the equilibrium value, Eq. (39) can be rewritten in the form

$$\frac{\partial \tilde{\varphi}}{\partial \tilde{t}} = \tilde{\varphi} - \tilde{\varphi}^3 - \frac{4}{27} \frac{\tilde{\sigma}^2}{\tilde{\varphi}^3}. \quad (41)$$

The function on the right side of this equation is shown in Fig. 6 at several values of the dimensionless stress. In the absence of external stress ($\tilde{\sigma} = 0$), the equilibrium value, which is given by the intersection point of the corresponding curve with the abscissa axis, is $\tilde{\phi} = 1$. If the stress is increased rapidly to $\tilde{\sigma} = 0.5$, the order parameter will decrease and the kinetics of the approach to the equilibrium value will be exponential. If the stress is increased up to the critical value, $\tilde{\sigma} = 1$, the kinetics of the order parameter in the vicinity of the equilibrium point, which is at $\tilde{\phi}^* = \sqrt{2/3}$, will be determined by the equation

$$\frac{\partial(\tilde{\phi} - \tilde{\phi}^*)}{\partial \tilde{t}} = -6\sqrt{2/3}(\tilde{\phi} - \tilde{\phi}^*)^2. \quad (42)$$

Its solution

$$\tilde{\phi}(t) = \frac{\tilde{\phi}^*}{1 + 6\sqrt{2/3}\tilde{\phi}^*t} \quad (43)$$

shows that at critical stress, the approach to the equilibrium value obeys the power law.

At still higher stresses, the derivative $\partial\tilde{\phi}/\partial\tilde{t}$ remains negative at all values of the order parameter, which decreases until reaching zero. Obviously, this decrease proceeds with acceleration and the system attains the zero of the order parameter in a finite time. Indeed, at low values of the order parameter the greatest contribution comes from the last term in Eq. (41), so this equation can be rewritten in the form

$$\frac{\partial\tilde{\phi}}{\partial\tilde{t}} = -\frac{4}{27}\frac{\tilde{\sigma}^2}{\tilde{\phi}^3}. \quad (44)$$

Its solution has the form

$$\tilde{\phi}^2 = \sqrt{\tilde{\phi}(0)^4 - \frac{16}{27}\tilde{\sigma}^2t}, \quad (45)$$

supporting the above statement.

The numerical solutions of Eq. (41) for the critical stress ($\tilde{\sigma} = 1$), as well as for lower and higher stresses than the critical one, are shown in Fig. 7. The presented kinetics is amazingly similar to the kinetics observed in experiment [7]. The fact that the measurements in [7] were carried out under alternating stress does not compromise the results presented above. So, it can be shown that if the stress frequency is much higher than the reciprocal of the order parameter relaxation time (as was the case in the experiments described in [7]), Eq. (41) is valid provided that $\tilde{\phi}$ is understood as a mean of the order parameter over the stress period and $\tilde{\sigma}^2$ as a mean-square stress.

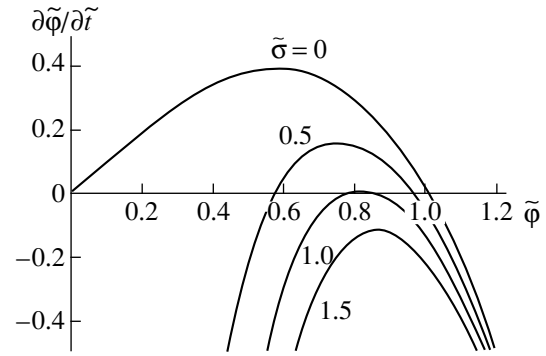


Fig. 6. The rate of variation of the normalized order parameter as a function of the order parameter.

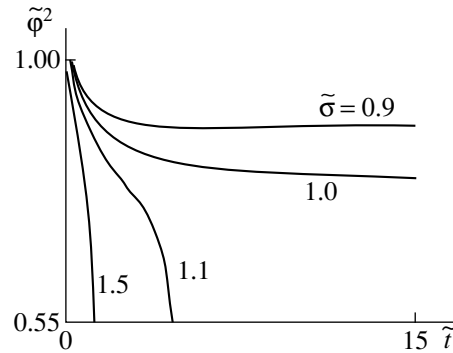


Fig. 7. Kinetics of the shear modulus (order parameter squared) at stresses below, at, and above the critical stress.

Note that in the vicinity of the point $\phi = 0$, Eq. (38), and the kinetic Eq. (39) as well, become, strictly speaking, inapplicable. Taking into account viscosity, instead of Eq. (38), we will have

$$\varepsilon_{el} = \frac{\sigma}{\phi^2 + \eta/\tau}, \quad (46)$$

and the kinetic equation in the vicinity of the zero of the order parameter takes the form (we return to the initial dimensional variables)

$$\dot{\phi} = -\gamma \frac{\sigma^2}{(\phi^2 + \eta/\tau)^2} \phi. \quad (47)$$

It is obvious that if

$$\phi^2 < \eta/\tau, \quad (48)$$

the kinetics changes from the square-root law of Eq. (45) to an exponential law with the characteristic relaxation time τ' ,

$$\frac{1}{\tau'} = \frac{\gamma\sigma^2}{(\eta/\tau)^2}. \quad (49)$$

KINETICS OF THE TRANSITION TO SOLID STATE

If in the experiment the resulting strain is maintained constant, then the kinetic equation taking into account the finite rate of plastic deformation is obtained from Eq. (27) by substituting Eq. (46)

$$\dot{\phi} = -\gamma \left[2\alpha(T - T_c)\phi + 2b\phi^3 + \phi \frac{\sigma^2}{(\phi^2 + \eta/\tau)^2} \right]. \quad (50)$$

Going back to the dimensionless variables (40), this equation can be written in the form

$$\frac{\partial \tilde{\phi}}{\partial \tilde{t}} = \tilde{\phi} - \tilde{\phi}^3 - \frac{4}{27} \frac{\tilde{\sigma}^2 \tilde{\phi}}{(\tilde{\phi}^2 + k)^2}, \quad (51)$$

where

$$k = \frac{\eta}{\tau} \frac{1}{\phi_0^2}. \quad (52)$$

The linearization of Eq. (50) in the neighborhood of the point $\tilde{\phi} = 0$ shows that the liquid state becomes unstable at

$$\tilde{\sigma} = \frac{3}{2} \sqrt{3k}. \quad (53)$$

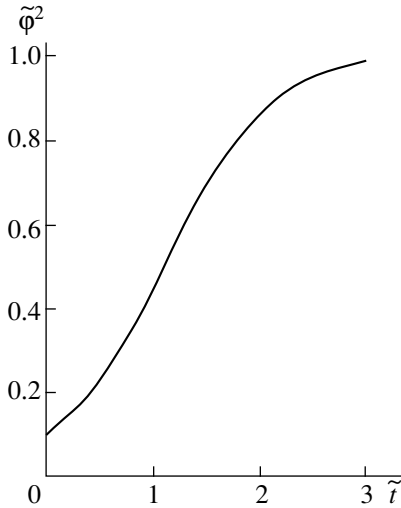


Fig. 8. Kinetics of the shear modulus increase after removal of load.

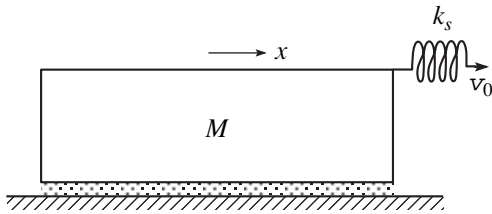


Fig. 9. Schematic of the tribological system.

Thus, if $k \ll 2/(3\sqrt{3})$, the strain of solidification will be substantially lower than the strain of melting.

Let us analyze the solidification kinetics in the limiting case where the strain drops instantaneously from the critical value (53) down to zero. The solution of kinetic Eq. (50) at $\tilde{\sigma} = 0$ is given in an implicit form by

$$\ln \frac{\tilde{\phi}}{\tilde{\phi}(0)} - \frac{1}{2} \ln \frac{1 - \tilde{\phi}}{1 - \tilde{\phi}(0)} - \frac{1}{2} \ln \frac{1 + \tilde{\phi}}{1 + \tilde{\phi}(0)} = \tilde{t}. \quad (54)$$

The corresponding time dependence of the normalized shear modulus $\tilde{\phi}^2$ from an initial value of $\tilde{\phi}^2(0) = 0.1$ is presented in Fig. 8.

STABILITY CONDITIONS FOR STEADY SLIDING

The dynamics of any tribological system is determined not only by the forces of friction acting directly in the sliding plane, but also by the elastic properties of the system as a whole. The latter in many cases can be described in the framework of the following model presented schematically in Fig. 9. A spring with elastic constant k_s is tied to a block of mass M placed on a flat substrate. The contact area of the block with the floor is denoted by A . The free end of the spring moves with a constant velocity v_0 . The equation of motion of the block has the form

$$M\dot{x} = k_s(v_0 t - x) - F_0, \quad (55)$$

where F_0 is the friction force

$$F_0 = A(\sigma_{el} + \sigma_{visc}). \quad (56)$$

For the rate of plastic flow determining the viscous contribution to the strain, we will use the approximation

$$\dot{\epsilon}_{pl}(\epsilon_{el}) = \begin{cases} \frac{1}{\tau} e^{-\frac{U_0(1-\tilde{\epsilon}^2)}{k_B T \tilde{\epsilon}}} & \tilde{\epsilon} < 1 \\ \frac{\tilde{\epsilon}}{\tau}, & \tilde{\epsilon} > 1, \end{cases} \quad (57)$$

which is based on the thermoactivated nucleation of dislocations. The form of this dependence for different values of the parameter $U_0/k_B T$ is shown in Fig. 10. For a layer of thickness d , using Eqs. (54), (57), and the relation $x = d\epsilon$ and neglecting the inertia term, the equation of motion is

$$k_s(v_0 t - x) = A \left(\sigma_{el} + \frac{\eta}{\tau} R(\epsilon_{el}/\epsilon_{el,c}) \right). \quad (58)$$

Differentiating Eq. (12) with respect to time and taking into account Eq. (57), we obtain

$$\dot{x} = \dot{\epsilon}_{el} d + \frac{1}{\tau} R(\epsilon_{el}/\epsilon_{el,c}) d. \quad (59)$$

Kinetic Eq. (27) can be transformed to

$$\dot{\mu} = -2\gamma[(\varepsilon_{el}^2 - \varepsilon_{el,c}^2)\mu + 2b\mu^2]. \quad (60)$$

Equations (58)–(60) completely determine the dynamics of the block's movement over the substrate. The steady-state solutions of this system of equations satisfy the following three equations:

$$\frac{1}{\tau}R(\varepsilon_{el}/\varepsilon_{el,c})d = v_0, \quad (61)$$

$$\mu = \frac{\varepsilon_{el}^2 - \varepsilon_{el,c}^2}{2b}H\left(1 - \frac{\varepsilon_{el}}{\varepsilon_{el,c}}\right), \quad (62)$$

$$-k_s = A\left(\varepsilon_{el}\mu + \frac{\eta}{\tau}R\left(\frac{\varepsilon_{el}}{\varepsilon_{el,c}}\right)\right). \quad (63)$$

Here, $H(x)$ is the Heaviside function, $\xi = x - v_0t$.

In order to study the linear stability of this solution, we will linearize the system of Eqs. (58)–(60) in the vicinity of the steady-state solution (61)–(63):

$$d\delta\dot{\varepsilon}_{el} + \frac{d}{\tau\varepsilon_{el,c}}R'(\varepsilon_{el}/\varepsilon_{el,c})\delta\varepsilon_{el} = \delta\dot{\xi}, \quad (64)$$

$$\delta\dot{\mu} = -2\gamma((\varepsilon_{el}^2 - \varepsilon_{el,c}^2)\delta\mu + 2\varepsilon_{el}\mu\delta\varepsilon_{el} + 4b\mu\delta\mu), \quad (65)$$

$$-k_s\delta\xi = A\left(\delta\mu\varepsilon_{el} + \mu\delta\varepsilon_{el} + \frac{\eta}{\tau\varepsilon_{el,c}}R'\delta\varepsilon_{el}\right). \quad (66)$$

Substituting into Eqs. (64)–(66):

$$\delta\varepsilon_{el}, \delta\mu, \delta\xi \propto e^{pt}, \quad (67)$$

we come to the following condition of the solvability of the system of equations (64)–(66)

$$Sp^2 + Np + Q = 0, \quad (68)$$

where

$$S = k_s d + A\left(\mu + \frac{\eta}{\tau\varepsilon_{el,c}}R'\right), \quad (69)$$

$$N = \left(k_s d + A\left(\mu + \frac{\eta}{\tau\varepsilon_{el,c}}R'\right)\right) \times (2\gamma|\varepsilon_{el}^2 - \varepsilon_{el,c}^2|) + k_s d \frac{R'}{\tau\varepsilon_{el,c}} - 4A\gamma\mu\varepsilon_{el}^2, \quad (70)$$

$$Q = k_s d \frac{R'}{\tau\varepsilon_{el,c}} 2\gamma|\varepsilon_{el}^2 - \varepsilon_{el,c}^2|. \quad (71)$$

The stability condition for the solutions is that all coefficients in Eq. (68) should have the same sign. Since S and Q are always positive, this condition reduces to the requirement that the coefficient N be positive. The equation $N = 0$, therefore, describes the boundary on the (ε_{el}, k_s) plane separating regions of the steady-state and non-steady-state movements. Solving it with respect to k_s and taking into account Eq. (62), the

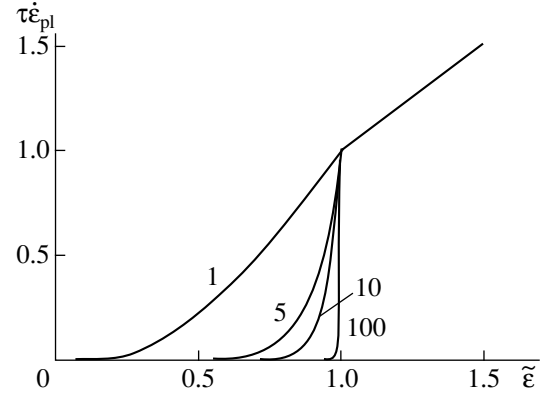


Fig. 10. The plastic deformation rate as a function of elastic deformation at different values of the activation energy. Numbers by the curves are values of the parameter $U_0/k_B T$.

equation of the curve bounding the region of steady states can be obtained:

$$\frac{k_s d}{A\mu_0} = \frac{2\tilde{\varepsilon}(1 - \tilde{\varepsilon}^2)H(1 - \tilde{\varepsilon}) - \left\{(1 - \varepsilon^2)H(1 - \tilde{\varepsilon}) + \frac{\eta}{\tau\mu_0\varepsilon_{el,c}}R'\right\}|1 - \tilde{\varepsilon}^2|}{|1 - \tilde{\varepsilon}^2| + \frac{1}{2\gamma\tau\varepsilon_{el,c}^3}R'}. \quad (72)$$

The form of the boundary of the steady states in coordinates $(k_s, \dot{\varepsilon}_{pl})$ can be obtained by solving Eq. (57) with respect to $\tilde{\varepsilon}$ and substituting the function $\tilde{\varepsilon} = \tilde{\varepsilon}(\dot{\varepsilon}_{pl})$ into Eq. (72). In the limit of large values of the parameter $U_0/k_B T$, Eq. (72) can be presented in a simple analytical form. First at all, note that

$$R' = \frac{U_0(1 + \tilde{\varepsilon}^2)}{kT\tilde{\varepsilon}^2}\tau^{-1}R = \frac{U_0(1 + \tilde{\varepsilon}^2)}{kT\tilde{\varepsilon}^2}\dot{\varepsilon}_{pl}. \quad (73)$$

Then, in the case $U_0/k_B T \gg 1$, the rate of plastic deformation is essentially nonzero only in the immediate vicinity of the critical value $\tilde{\varepsilon} = 1$ and we can replace $\tilde{\varepsilon}$ by its critical value throughout Eq. (72) with the exception of terms containing

$$(1 - \tilde{\varepsilon}^2) \approx -\frac{k_B T}{U_0} \ln(\dot{\varepsilon}_{pl}\tau).$$

Ultimately, we arrive at the equation

$$\frac{k_s d}{2A\mu_0} = \frac{-\ln(\dot{\varepsilon}_{pl}\tau)(1 - k_1(\dot{\varepsilon}_{pl}\tau))}{-\ln(\dot{\varepsilon}_{pl}\tau) + k_2(\dot{\varepsilon}_{pl}\tau)}, \quad (74)$$

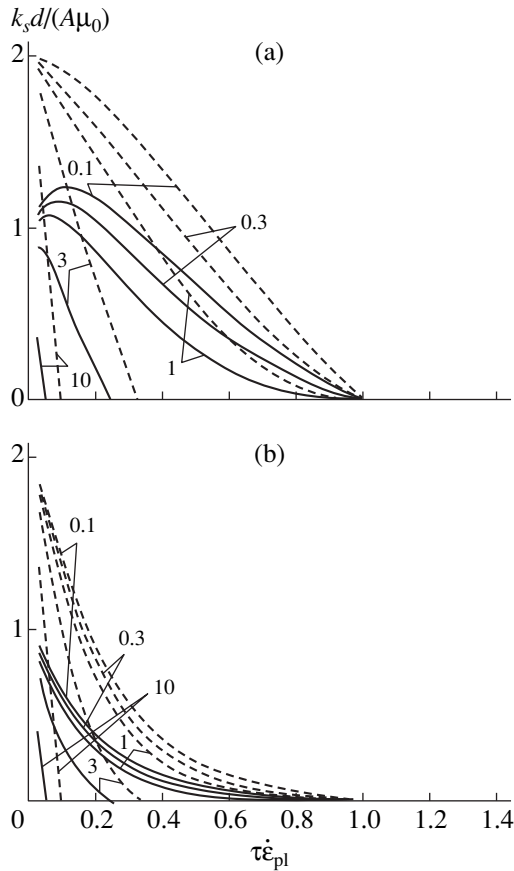


Fig. 11. Curves bounding the region of steady-state motion. The motion is steady-state in the region above the corresponding curve and non-steady-state below. Numbers by the curves are values of the parameter $U_0/k_B T$. Solid curves are exact solutions of Eq. (72), dashed curves are the approximate solutions of Eq. (74). k_2 : (a) 1.0 and (b) 10.

where

$$k_1 = \frac{\eta}{\tau \mu_0 \epsilon_{el,c}} \frac{U_0}{k_B T}, \tag{75}$$

$$k_2 = \frac{1}{\gamma \tau \epsilon_{el,c}^2} \left(\frac{U_0}{k_B T} \right)^2. \tag{76}$$

The stability diagram for the steady-state movement is presented in Fig. 11.

PHYSICAL ORIGIN OF THE SHEAR-INDUCED MELTING PHASE TRANSITION

Returning to the problem of the possibility of a second-order phase transition between crystalline and liquid states, we emphasize that the peculiar feature of this transition in our case is that it occurs under conditions that considerably change the symmetry of both the crystalline and liquid states, in the first place because of the confining solids and, in second place, by elastic deformation. Orientation of the surfaces of the solids

determines the preferential plane, and the direction of shear determines the preferential direction in that plane. The vanishing of the order parameter and the related vanishing of the shear modulus imply that the microscopic density, which is periodic in the crystalline state, becomes uniform in the liquid state [14]. The preferential character of the direction of shear creates a situation in which, as a result of melting, the microscopic density modulation disappears first only in the direction of shear. If this is indeed the case, then the melting has a peculiar one-dimensional character. Whether such one-dimensional melting actually takes place in the above experiments can be found out experimentally by studying the orientation of the atomic layers after shear-induced melting and the subsequent solidification of the layer when the load is reduced.

Let us show that the mentioned partial one-dimensional melting can occur as a second-order phase transition. Below we follow Landau's argument [14], making such modifications as necessary for an anisotropic system.

Let $\delta\rho$ be the change in microscopic density of the layer as a result of transition from the crystalline state to the shear-melted state with x as the coordinate in the direction of shear. $\delta\rho$ as a function of x is equal to zero in the molten state and is nonzero in the crystalline state. Let us represent $\delta\rho$ as an expansion in plane waves:

$$\delta\rho = \sum_k b_k e^{ikx}, \tag{77}$$

where $k = 2\pi n/a$, n is an integer, a is the lattice constant in the x direction, and the term with $k = 0$ is excluded from the summation. Since $\delta\rho$ is real, there should be

$$b_k = b_{-k}^*, \tag{78}$$

where $*$ denotes the complex conjugation.

Expansion of the free energy in powers of $\delta\rho$ contains terms of the form

$$b_{k_1}, b_{k_2}, b_{k_3}, \dots \tag{79}$$

For the free energy to be invariant with respect to translations, it is necessary that $k_1 + k_2 + k_3 + \dots = 0$. Hence, the second-order terms contain only the products $b_k b_{-k}$ or, according to Eq. (78), $|b_k|^2$. The free energy expansion, therefore, has the form

$$f = \sum_k A_k |b_k|^2. \tag{80}$$

Here, A_k are constants depending on temperature and elastic deformation. Above the melting point, f has a minimum for all A_k ; i.e., all A_k are positive. At the point of the continuous phase transition, the second-order term should become zero at nonzero values of $\delta\rho$. This means that one of the A_k coefficients should turn zero at the transition point; i.e., the curve $A(k)$ is tangential to

the abscissa axis. Thus, at the transition point, the density modulation arises at a certain single wavelength determined by the k_0 value at which A_k becomes zero.

In accordance with Landau's theory [14], the higher order terms can now be written in form (79), where the summations are carried out over only the wave vectors in which the instability appears. It is clear that the free energy expansion contains no third-order terms since no sum of three wave vectors from the set $\{k_0, -k_0\}$ is equal to zero. But the condition for the continuous phase transition is just that there must be no third-order terms in the expansion of the free energy. The disappearance of the microscopic density modulation in the direction normal to the sliding (at larger shear deformations) is also the one-dimensional phase transition and can proceed as a second-order phase transition.

We come to the conclusion that the specific symmetry of thin layers subjected to strong shear deformations, in contrast to melting by increasing temperature, results in the possibility of two melting transitions, both of which can proceed as a second-order phase transition. The first transition results in the disappearance of the density modulation only in the direction of shear (the one-dimensional melting) and the second transition results in the disappearance of the density modulation in the normal direction as well (total melting). Although we do not know of experiments that directly support the conclusions made, the computer simulation shows that the transition from the molten to the solid state indeed often occurs as two consecutive phase transitions [6].

CONCLUSIONS

The concepts developed in the present study naturally explain the experimentally observed peculiarities of the kinetics of shear-induced melting and solidification of thin layers of lubricants. Of fundamental importance is the introduction of, first, the nonequilibrium shear modulus and, second, the nonequilibrium order parameter, the latter being a more fundamental quantity than the shear modulus. It is for the order parameter that the analytical kinetic equations at the transition point can be written. The introduction of these two quantities makes possible a natural interpretation of the linear response of the system under fast alternating stress and of the occurrence of the phase transition as the amplitude of the alternating stress increases. The agreement between the obtained variations with time of the shear modulus in direct and reverse transitions, on the one hand, and the experimental data, on the other,

shows that the experimentally observed peculiarities of the melting kinetics of thin layers can be described assuming that shear-induced melting is a result of a second-order phase transition taking place as elastic deformation of the layer increases.

ACKNOWLEDGMENTS

The author is grateful to B. Persson for discussion and the invitation to conduct research at Forschungszentrum Jülich (Germany) during which time this study was done.

This study was supported by the German Ministry of Science and Technology (BMBF) in the framework of the joint German-Israeli project "Novel Tribological Strategies from Nano- to Meso-Scales."

REFERENCES

1. R. Erlandson, G. Hadziioannou, C. M. Mate, *et al.*, *J. Chem. Phys.* **89**, 5190 (1988).
2. J. Krim and A. Widom, *Phys. Rev. B* **38**, 12184 (1988).
3. J. N. Israelashvili, *Surf. Sci. Rep.* **14**, 109 (1992).
4. B. N. J. Persson, *Sliding Friction. Physical Principles and Applications* (Springer-Verlag, Berlin, 1998).
5. *Fundamentals of Friction: Macroscopic and Microscopic Processes*, Ed. by I. L. Singer and H. M. Pollock (Kluwer, Dordrecht, 1992).
6. *Physics of Sliding Friction*, Ed. by B. N. J. Persson and E. Tosatti (Kluwer, Dordrecht, 1996).
7. G. Reiter, A. L. Demirel, J. Peanasky, *et al.*, in *Physics of Sliding Friction*, Ed. by B. N. J. Persson and E. Tosatti (Kluwer, Dordrecht, 1996), pp. 119–138.
8. A. L. Demirel and S. Granik, *J. Chem. Phys.* **109**, 6889 (1998).
9. P. A. Thomson, M. O. Robbins, and G. S. Grast, *Isr. J. Chem.* **35**, 93 (1995).
10. L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Nauka, Moscow, 1976; Pergamon, Oxford, 1980), Part 1.
11. A. Seeger, in *Handbuch der Physik*, Ed. by S. Flügge (Springer-Verlag, Berlin, 1958), Vol. VII/2, p. 1.
12. A. L. Demirel and S. Granik, *Phys. Rev. Lett.* **77**, 4330 (1996).
13. G. Reiter, A. L. Demirel, J. Peanasky, *et al.*, *J. Chem. Phys.* **101**, 2606 (1994).
14. L. D. Landau, *Zh. Éksp. Teor. Fiz.* **11**, 627 (1937); in *Collected Papers of L.D. Landau* (Gordon and Breach, New York, 1967), Paper No. 29, Part II.

Translated by M. Lebedev

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Designing and Testing a Compton Converter of Gamma Radiation and X-rays to Electric Current

V. N. Gryzun, G. G. Katraev, and A. K. Chernyshev

Russian Federal Nuclear Center, All-Russia Research Institute of Experimental Physics, Sarov,
Nizhegorodsk oblast, 607190 Russia

e-mail: otd4@expd.vniief.ru

Received March 7, 2000; in final form, August 8, 2000

Abstract—The design and testing of a Compton current generator are briefly described. Its current response to a γ quantum (i.e., sensitivity) is estimated at $\eta = 7.0 \times 10^{-21}$ (A cm² s)/ γ . At relatively high gamma fluxes, the generator can be employed as a battery charger. When it is fed by a pulsed gamma flux, the resulting current pulse is capable of triggering the equipment and evaporating a metal wire synchronously with the application of the gamma pulse. The generator can also be used as a direct-charge radiation detector for determining the dose and dose rate of a radiation pulse, as well as its shape. © 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

Interaction of high-energy directional gamma fluxes with matter generates the directional motion of high-energy charged particles, i.e., a photo-Compton electron current I_C (with a density of \mathbf{j}_C), and also causes charge separation in the matter with the formation of a current of low-energy charged particles, i.e., conduction (or leakage) current I_σ (with a density \mathbf{j}_σ). In our case, the currents are oppositely sensed [1].

In practice, these phenomena adversely affect the operation of devices in the radiation environment. At the same time, they can have a beneficial effect in a number of specific cases [2].

Our aim was to take advantage of a directional flux of high-energy charged particles in converters of X-rays to electric current. These converters are of two types: current generators and radiation detectors that are charged directly by the action of a gamma flux (hence, the term “direct-charge radiation detectors”) [3]. Their designs are similar (see Eq. (6)). In the radiation energy range considered in this article (0.5–10 MeV), the Compton scattering of gamma quanta by bound electrons of insulator atoms prevails; therefore, both devices will be called Compton current generators (CCGs). Structurally, a CCG represents an insulating gap of arbitrary geometry with two metal electrodes on its boundaries or a set of such gaps.

In this work, the device was made in the form of a planar disk that had d -spaced 50- μ m-thick copper foil plates. The gap was filled by polyethylene (CH₂)_{*n*} or fluoroplastic (CF₂)_{*n*}. The energies of gamma quanta, chemical composition of the device, and size d were such that the pair-generation effect and the contribution of the thin plates into I_C and I_σ could be ignored.

ELECTRODYNAMIC PROBLEM AS APPLIED TO A GAMMA-IRRADIATED GAP WITH A CONDENSED INSULATOR

The electric field in a CCG insulating gap exposed to radiation is determined from the expression

$$\frac{\partial \rho}{\partial t} = \operatorname{div}(\mathbf{j}_C - \mathbf{j}_\sigma) = \operatorname{div}(\mathbf{j}_C - \sigma \mathbf{E}), \quad (1)$$

$$\operatorname{div} \mathbf{D} = \operatorname{div}(\epsilon_0 \epsilon \mathbf{E}) = \rho.$$

Here, σ is the radiation-induced conductivity of the insulator, \mathbf{E} is the electric field strength vector in the insulator, \mathbf{D} is the dielectric displacement vector, ϵ_0 is the dielectric constant, ϵ is the relative permittivity of the insulator, ρ is the space charge density in the insulator, and t is time. Problem (1) is solvable because the dependences of the material parameters (σ , ϵ , \mathbf{j}_C) of a specific insulator on radiation-field parameters (spectrum, dose, dose rate, irradiation time, and postirradiation relaxation), as well as on attendant external effects, are known. On the condition that the gamma quantum range in the insulator [4] is much greater than d even for the lower limit of the energy interval, the solution of system (1) over the space (without considering the explicit dependences of the insulator parameters on the acting radiation field) is given by

$$I_g(t) = \frac{d}{dt}(CV) + \left(\frac{1}{R_\sigma(t)} + \frac{1}{R_n} \right) V = I_c + I_\sigma + I_n. \quad (2)$$

Here, $I_g(t)$ is the generator current as a function of $\mathbf{j}_C(t)$; V , radiation-induced potential difference between the generator plates; C , capacitance of the gap; R_σ , leakage resistance of the gap; R_n , load resistance; I_n , current passing in the external circuit; and I_c , charging current in the gap. In (2), $I_g(t)$ is the only material parameter;

however, the radiation-induced electric field in the gap (the gap specifies the power and current capabilities of CCGs) also depends on the other material parameters of the insulator. Therefore, the complete design of the CCG also implies the calibration of the three $(\sigma, \epsilon, \mathbf{j}_C)$ material parameters of the insulator according to their dependences on the radiation-field parameters and device geometry. Note, however, that when insulating gaps are used as a current generator driven by gamma radiation, our goal is to obtain significant values of I_n ; therefore, the ideal case would be the short-circuit mode $I_n = I_g(t)$, according to (2). In practice, designers are making efforts to provide conditions as close to the short-circuit mode as possible. In this case, the leakage current is much smaller than $I_g(t)$ and is ignored; hence, $I_n \cong I_g(t) - I_c$. By choosing the CCG design and load such that the setting time of the potential difference between the plates is much less than the exposure time, we eventually obtain $I_g(t) \cong I_n$. Therefore, in this work, we calculate only the generator current $I_g A(t)$ and its sensitivity to radiation and verify our calculation experimentally; the calculation and calibration of the conductivity and permittivity of the insulator are beyond the scope of our report.

Calculation of the CCG Current and Sensitivity

Usually, gamma sources (especially, impulse sources) have a continuous spectrum of gamma quanta. Under these conditions, it is impossible to determine the CCG current and power capacities by calibration, since an appropriate set of calibrated high-energy sources that cover the continuous gamma spectrum is lacking. Therefore, one cannot get by with analytical methods in this case. Designers usually first develop an adequate procedure for determining the sensitivity of a CCG and then verify it with respect to the amplitude with experimental techniques using available calibrated monochromatic gamma sources. In addition, they observe how the CCG current reproduces the shape of incident gamma radiation by means of an impulse source with a known or a still unknown shape. Good agreement between the analytical and experimentally measured sensitivity values means that both the design and the fabrication technology of the generator are acceptable.

Let a gamma flux $N_0 \gamma/\text{cm}^2$ be incident on a planar CCG normally to its surface. We assume that the gamma quanta are distributed with some spectral density. Then, the normalized-to-unity group spectrum of the quanta is given by

$$f(E)\Delta E = \frac{\Delta N_E}{N_0}, \tag{3}$$

where E is the energy of the quanta [MeV], ΔN_E is the number of quanta with an energy in the interval $\Delta E =$

$E_2 - E_1$, and $f(E)$ is the density of the energy distribution of the quanta [MeV⁻¹].

Next, let the normalized-to-unity time distribution of the quanta have the form

$$\varphi(t)\Delta t = \frac{\Delta N_t}{N_0}. \tag{4}$$

Here, ΔN_t is the number of quanta per unit time interval $\Delta t = t_2 - t_1$ and $\varphi(t)$ is the density of the time distribution [s⁻¹], which describes the radiation pulse shape.

Let $\mu(E)$ be the attenuation coefficient of gamma radiation in the insulator, $\delta(E)$ be the mean projection of the recoil electron path in the insulator onto the direction of incidence of the radiation, and q be the charge of an electron. Then, the magnitude of the radiation-induced current in the insulator can be expressed as

$$j_C(x, t) = -qN_0\varphi(t) \int_0^{E_{gr}} \mu(E) \exp(-\mu(E)x) \delta(E) f(E) dE, \tag{5}$$

where E_{gr} is the greatest energy of the gamma quanta and x is the space coordinate that coincides with the direction of gamma flux incidence on the CCG and is reckoned from its front plate.

Equation (5) shows that the radiation-induced current reproduces (in time) the shape of the acting radiation. In Eq. (2), $I_g(t)$ is expressed through $\mathbf{j}_C(x, t)$; namely, $I_g(t)$ is the convolution of $\mathbf{j}_C(x, t)$ over the space with consideration for the distribution of recoil electrons in the insulator and with regard for the transport of their charge to the CCG plate (the transport ratio depends on the insulating gap geometry).

For a planar CCG, a steady-state monochromatic gamma source, normal incidence of gamma quanta on the CCG plane, uniform irradiation of the plate surface area S , and an insulator thickness much less than the gamma quantum range in the insulator, we have

$$I_g(t) = 0.5SqN_0\varphi(t)\delta\mu^2d \tag{6}$$

(the intermediate mathematics is omitted).

Note that $N_0\varphi(t)$ in expressions (5) and (6) is the gamma flux density and represents the instantaneous intensity of the acting flux. This quantity specifies to what extent the CCG current reproduces the shape of incident radiation in time. For a static gamma source, $N_0\varphi(t)$ is the time-invariable gamma flux density; we designated it as Φ_0 . Let us define the CCG current response (sensitivity) as the CCG current per unit gamma flux density. Then, for an Étalon-3 static Co_{27}^{60} gamma source with an effective gamma energy of 1.253 MeV and for the CCG geometry mentioned

above, the rated sensitivity η is

$$\eta = \frac{I_g}{\Phi_0} = 0.5Sq\delta\mu^2d = 6.71 \times 10^{-21} \frac{\text{A cm}^2 \text{ s}}{\gamma}.$$

CCG CALIBRATION ON AN ÉTALON-3 GAMMA RADIATION SOURCE

Calibration was aimed at measuring the CCG sensitivity to Co_{27}^{60} gamma quanta: the CCG current is divided by the density of a gamma flux passing through the generator.

Metering Equipment

The Étalon-3 source was certified at the Mendeleev Research Institute of Materials as a secondary standard of exposure dose (ED) and exposure dose rate (EDR) (certificate no. 39/91). Other devices (U5-11 amplifier, M2300 dose meter, Shch1516 voltmeter, and D3-28 computer) were certified at the Mendeleev Research Institute of Materials and the Specialized Laboratory of Metering Equipment.

On calibration, the radionuclide activity was ≈ 100 Ci and the gamma beam axis passed perpendicular to the working space of the CCG through its center. The distance between the front wall of the CCG and the center of the source was 57 cm.

Results of Measurement

A uniform (within $\pm 3\%$) collimated gamma flux uniformly irradiated the CCG surface. With the M2300 dose meter, we measured the relative differences in the EDR values for various collimator diameters and determined the EDR at $l = 57$ cm using the known EDR value at the certified point ($l = 150$ cm). The validity of the EDR values measured (the accuracy 2.5%) was confirmed by the fulfillment of the condition $\text{EDR}l^2 = \text{const}$. We took the mean value.

The gamma flux density $\Phi_\gamma(\gamma/(\text{cm}^2 \text{ s}))$ is related to the EDR ($\mu\text{r/s}$) as $1.70 \times 10^3 \gamma/(\text{cm}^2 \text{ s}) = 1 \mu\text{r/s}$. Using the known EDR at a distance of 57 cm to the source center, we obtained $\Phi_\gamma = 1.51 \times 10^8 \gamma/(\text{cm}^2 \text{ s})$.

The CCG sensitivity is found by the formula

$$\eta_\gamma = \frac{U}{\Phi_\gamma R_n} = 7.0 \times 10^{-21} \frac{\text{A cm}^2 \text{ s}}{\gamma}.$$

Here, $R_n = 1.01 \times 10^{11} \Omega$ is the metering resistance of the U5-11 amplifier and $U(V) = U_2 - U_1$, where U_2 and U_1 are the mean voltages across the plates under and without irradiation, respectively.

The mean voltage values were obtained from several tens of observations, each observation being 100 to 300 instantaneous readings of the Shch1516 voltmeter that were averaged with the DZ-28.

The systematic errors of measurement were as follows: $\pm 3\%$ for the EDR at a distance of 57 cm to the source, $\pm 4\%$ for the coefficient of EDR-to- Φ_γ conversion, and $\pm 2\%$ for R_n . The relative errors introduced by the Shch1516 and M2300 were negligible. The random error of measurement is the sum of random errors of the voltages and does not exceed $\pm 6\%$ at a confidence probability of 0.95.

Thus, the overall relative error of the CCG sensitivity, measured according to State Standard 8.207, is no more than 9% at a confidence probability of 0.95. The reproducibility of the results was 96.5% or higher. The difference between the calculated and experimental sensitivities, 4.1%, lies within the calibration accuracy and is due to the known systematic error of calculation that can be eliminated in practice.

TESTING OF THE GENERATOR ON A FAST PULSE GRAPHITE REACTOR

The aim of experiments with the fast pulse graphite reactor [5] was to record the shape of the CCG current pulse when the generator is fed by gamma radiation from the reactor and to compare it with the reactor pulse recorded by the standard technique [5]. The qualitative parameter of comparison is the pulse shape itself, and the quantitative parameter is the pulse half-width.

The Compton generator was 84 cm distant from the surface of the active zone of the reactor.

RESULTS

The CCG pulse halfwidth measured from the oscillogram was found to be 3.25×10^{-3} s. This value differs from the reactor pulse halfwidth measured by the standard technique by 0.31%. Such a discrepancy falls within a measurement accuracy of 10% for both values.

The shapes of both pulses were totally coincident.

CONCLUSIONS

(1) The electrodynamic properties of an insulating gap exposed to continuous-spectrum gamma radiation cannot be determined by calibration because calibrated steady-state monochromatic high-power gamma sources covering the entire spectrum of impulse gamma sources are lacking. Therefore, in this work, the characteristics we are interested in were obtained analytically, i.e., by solving the boundary-value problem in the gap for a CCG of planar geometry. The solution, which includes the spectrum and density of a gamma flux, relates the current and potential difference between the generator plates to radiation-field parameters.

(2) Our analytic method was verified by the example of the Co_{27}^{60} radionuclide spectrum and the gamma flux

intensity from a pulse nuclear reactor. It turned out that the calculated and experimental CCG sensitivities differ by 4.1%. Also, the current shape and duration of the CCG differ from those of the nuclear reactor that were determined by standard technique by 0.31%. These discrepancies are well within the measurement accuracy, counting in favor of the analytic method and the CCG fabrication technology.

(3) The coincidence of the analytic and experimental CCG sensitivities means that the analytic method is valid and can be applied to calculating not only the current of a specific generator but also, more generally, the radiation-induced current in the insulator.

(4) It was shown that the insulating gap converts the energy of high-frequency (2×10^{18} – 2×10^{21} Hz) electromagnetic field to a current pulse of a duration equal to the irradiation time. The current can charge batteries, trigger equipment, or evaporate a metal wire. In addition, the current energy can be converted to the energy of an electromagnetic pulse with a frequency corresponding to the irradiation time.

(5) A CCG can be used as a direct-charging detector for determining the shape of gamma pulses with a con-

tinuous spectrum. It can also be applied to measuring the dose and intensity when the spectrum of gamma radiation is (i) monochromatic; (ii) continuous and either known prior to experiments or is to be measured experimentally; and (iii) continuous, unknown, and weakly ($\approx 10\%$) variable from run to run.

REFERENCES

1. B. J. Gross, *Appl. Phys.* **36**, 1635 (1965).
2. B. J. Gross, *Radiat. Res.* **14**, 117 (1961).
3. V. I. Ivanov, *Dosimetry of Ionizing Radiations* (Atomizdat, Moscow, 1964).
4. E. Storm and H. Israel, *Photon Cross Sections—from 0.001 to 100 meV for Elements 1 through 100* (Los Alamos Scientific Laboratory, New México, 1967; Atomizdat, Moscow, 1973).
5. M. I. Kuvshinov, V. F. Kolesov, A. M. Voinov, and I. G. Smirnov, *Vopr. At. Nauki Tekh., Ser. Impul. Reak. Prostye Kritich. Sborniki* **1**, 3 (1988).

Translated by V. Isaakyan

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Thermally Stimulated Depolarization of Human Blood

L. S. Pinchuk, A. G. Kravtsov, and S. V. Zotov

*Belyi Research Institute of Mechanics of Metal-Containing Polymer Systems,
Belarusian Academy of Sciences, ul. Kirova 32a, Gomel, 246050 Belarus*

e-mail: kravtsovag@usa.net

Received March 27, 2000

Abstract—Thermally stimulated current spectra of human blood preparations are investigated. The electret effect in blood is interpreted using the fundamental concepts of the electret state of matter and electrical polarization of biopolymers. It is speculated that the electret effect in blood is related to the biological cycles in the organism and associated biochemical processes. It is noted that the method of thermally stimulated current is appropriate for express diagnostics of biological objects. © 2001 MAIK “Nauka/Interperiodica”.

INTRODUCTION

The electret effect, i.e., the ability of materials to retain the electrically polarized state after an external effect responsible for polarization has been removed and to produce the electric field in the environment, is a fundamental property of living and nonliving matter [1]. To date, this effect has been discovered in many materials and biological objects [1–3]; therefore, it is logical to assume that it plays a vital role in the life activity of biological systems. Obviously, the study of the electret properties of living tissues is of great interest in biology and medicine.

The concept of electrical phenomena in biological objects dates back to the middle of the 19th century, when Du Bois-Ramond put forward the so-called electromolecular theory [2]. It is known that more than 90% of living matter is accounted for by polar molecules of proteins, nucleic acids, lipids, adipose matter, carbohydrates, and water, which bear coupled electric charges (dipoles). The essence of Du Bois-Ramond’s theory is that biological electrogenesis is governed largely by the dynamics of polar groups entering into the composition of the molecules listed. By the early 1970s, two fundamental properties of living matter had been established: natural electrical polarization as a general property of living tissues and the presence of a quasi-polar bioelectric field [1–3]. Today, thorough investigation into the bioelectret effect via modern physical instruments and available theoretical concepts of the electret state has become a burning issue.

The most effective way of studying the charge state of insulators is the method of thermally stimulated depolarization (current) (TSC). Perhaps, the term “thermally stimulated discharge” would more adequately reflect its basic idea. In this method, the relaxation of a charge responsible for the electret state is studied. Since charge relaxation at room temperature is a long-term process, the discharge is stimulated by heating at a constant heating rate. The sample placed

between two electrodes is heated with a constant rate, and the current passing in the circuit is measured. From its temperature dependence (TSC spectrum), one can judge mechanisms behind the electret effect. The advantages of the TSC method are its high sensitivity and high resolution. Also, it gives a chance to trace relaxation processes in a material that can be in different states and to record the transition temperatures.

Recently, the tendency to extend the application of the TSC method has appeared. In particular, it was used for food characterization [4] and in studying the depolarization of collagen, hemoglobin, mioglobin, etc. [1, 5, 6]. In addition, with the TSC method, the electret effect was found to be a common property of polypeptides, polynucleotides, and polysaccharides [7–9].

Unlike electrical polarization in “simple” chemical compounds, the electret effect in human body tissues has a more complex nature and is of greater practical value. There are reasons to believe that the parameters of this effect vary according to life conditions of the organism [9]. The electrophysical properties of blood, one of the most important components of the human organism, are particularly sensitive to changes in the environmental conditions.

It was the aim of this work to study the electret properties of blood of different groups with the TSC method and to treat the results in terms of the fundamental concepts of the electret effect.

EXPERIMENTAL METHODS

In experiments, we used Rh-positive peripheral blood of groups I–IV. From each donor, blood was taken five times at 1–2 h intervals during a day. Samples (1 ml) were placed into an aluminum electrode (degraded in ethyl alcohol) and covered by a sterile polytetrafluoroethylene (teflon) spacer on which the second electrode was applied. The sample was heated at a rate of 5 °C/min, and the interelectrode current was

measured (for the measuring circuit, see Fig. 1). The thermoclearance method [10] was used to accurately determine the energies of activation of polarization charge relaxation that correspond to peaks in the TSC spectra. The sample was first heated with a constant rate to obtain the first peak and then was rapidly cooled to room temperature. After repeat heating, another peak at a higher temperature was recorded, etc. The energy of activation for each of the peaks was evaluated from the ascending portion of the $I(T)$ curve [10].

RESULTS AND DISCUSSION

For all the blood preparations, the TSC spectra have three extrema (Fig. 2). In the low-temperature range ($T_{\max} = 40\text{--}50^\circ\text{C}$), a smooth peak of negative charge relaxation is observed. It is known [1] that the low-temperature depolarization peak for collagen and hemoglobin is due to so-called bound water. Human blood is a multicomponent system consisting of plasma (a water solution of mineral salts, amino acids, proteins, steroids, enzymes, etc.) and blood cells (erythrocytes, leukocytes, thrombocytes, etc.) [11]. Erythrocytes (or red blood cells), containing hemoglobin enclosed by a lipid sheath and other components responsible for oxygen transport in the organism, account for about 45 wt % of blood. The specific chemical structure of amino acids, proteins, and lipids (a great number of polar groups, such as NH, CO, and OH) define the high hydrophilicity of the blood components; hence, most of them are covered by hydrate sheaths [1, 12]. Thus, it might be expected that some amount of water in blood is in the bound (structured) state since hydrogen is bonded to polar fragments of various organic molecules. This appears to increase the dipole moment of the components and cause the electret effect, since the energy of dipole water molecules depends on their orientation. Thermally stimulated disintegration of such coordinated structures is, in essence, dipole disordering. It seems likely that this disordering shows up in the form of the small low-temperature TSC peak. For this peak, the energy of charge relaxation activation is $W_1 = 0.45$ eV. The mean energy of a hydrogen bond is known to vary between 0.10 and 0.25 eV [13]. With complexly configured organic compounds (containing polar groups) present in blood, water molecules may form coordinated structures with higher bond energies.

The results obtained indirectly corroborate data that bound water in biopolymers can be in the electret state [1, 14].

In the middle-temperature range ($T_{\max} = 70\text{--}90^\circ\text{C}$), the extremum is the superposition of several peaks whose intensities are several times greater than that of the low-temperature peak. This means that, in this temperature range, some coordinated structures break down and bound dipoles with relatively high bond energies relax. In our opinion, the middle-temperature peak can be attributed to the thermal denaturation of

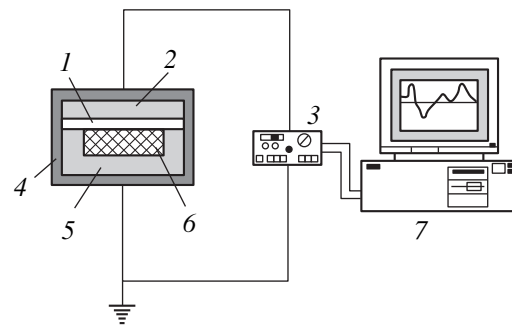


Fig. 1. Setup for measuring and recording TSCs: 1, teflon spacer; 2, upper (Al) electrode; 3, data amplifier; 4, dismountable screen; 5, lower (Al) electrode; 6, blood sample; and 7, PC with dedicated software.

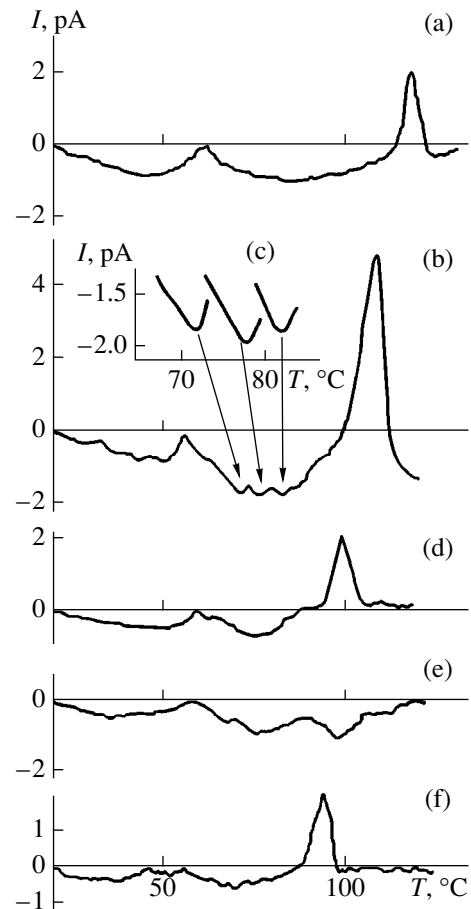


Fig. 2. TSC spectra of (a) group-I, (b, c) group-II, (d, e) group-III, and (f) group-IV human blood. In (d, e), the high-temperature peaks are of different polarity.

protein compounds in blood, i.e., to an irreversible modification of the protein structure without breaking the polypeptide chain. As is known, the secondary structure of proteins is formed by intramolecular hydrogen bonds and the tertiary structure, largely by Van der Waals forces [13]. The thermal relaxation of charges of ionized and protonated groups, which are

responsible for the electrostatic attraction between fragments of a protein molecule, and that of Van der Waals interactions cause the denaturation of blood proteins [15]. The wavy middle-temperature peak in Fig. 2b (which was resolved by thermoclearance, Fig. 2c) seems to correspond to the three values of the bond energies (W_2 between 0.5 and 0.7 eV) responsible for the polarization of proteins of group-II blood.

The high-temperature ($T_{\max} = 105\text{--}120^\circ\text{C}$) peak is the most intense. At these temperatures, one can expect the phase transition with the formation of thromboid mass. Intense thermal oxidative destruction of organic compounds entering into the blood composition is accompanied by the sticking together of lipid sheaths of erythrocytes, denaturated proteins, fragments of polypeptide chains, etc. The thermal decomposition of newly formed structures results in the relaxation of their polarization charge, which is associated mainly with Maxwell–Wagner polarization [1].

The TSC spectra for different blood groups are similar; however, some of the peaks shift along the temperature scale, vary in intensity, and even change sign (Figs. 2a–2f). For example, the high-temperature peak shifts toward lower temperatures with increasing group number. It lies at 115–118 (group I), 108–110 (II), 98–100 (III), and 95–97°C (IV).

The above effects are likely to correspond to specific life cycles of the organism. Reversible biochemical processes involve positive and negative ions, as well as positive and negative potentials in the double electrical layer around blood cells [16]. According to physiological factors (blood pressure, muscular activity, the presence of stimulators in the organism, etc.), the concentrations of various ions and the potential distributions may change, thereby affecting the parameters responsible for the electret effect. Small peak intensities may also indicate variations in the dipole moments of the blood components when a number of biochemical reactions fluctuate about equilibrium. A change in the high-temperature peak polarity for blood samples taken from the same donor during a day has been noted (Figs. 2d, 2e)—this phenomenon is yet to be understood. Note also that, in each of the experiments, the intensities and polarities of the TSC peaks vary in a different manner.

CONCLUSION

Thus, the TSC method provides information on such an important biological object as blood. With this method, it was established that human blood exhibits the electret effect due to the coordination of the supermolecular structure of its components. Our interpretation of the effect is based on the ability of complex organic molecules, including biopolymers, to show the electret properties, which follow from their specific complex chemical configuration. The data obtained may indicate that the electret effect in blood is both a

function of biochemical processes in the human organism and a factor that initiates them.

We believe that the TSC method might become a simple yet informative blood diagnostics tool. To fully exploit its potentialities, it is necessary to gain more experimental findings and perform complex medical and physical investigation in the fields of hematology, cardiology, physics, physical and biologic chemistry, etc. Therefore, it will be the further aim of the authors to refine the behavior of the parameters characterizing the electret effect, in particular, to relate the effect to the living conditions of a donor, its Rh factor, blood group, sex, age, etc., which is of great practical interest.

ACKNOWLEDGMENTS

We thank G.N. Tomina and V.A. Shapovalov for the assistance in the experiments and valuable discussions.

REFERENCES

1. *Electrets*, Ed. by G. M. Sessler (Laplacian Press, Morgan-Hill, 1999), Vol. 1.
2. A. I. Lakomkin and I. F. Myagkov, *Electrophysiology* (Khimiya, Moscow, 1977).
3. E. T. Kulin, *Bioelectret Effect* (Nauka i Tekhnika, Minsk, 1980).
4. Marc Galop, <http://ww.file:///ftpfiller/1/food/htm>.
5. M. G. Bridelli, R. Capeletti, S. Losi, *et al.*, in *Proceedings of the 8th International Symposium on Electrets, Paris, 1994*, p. 864.
6. R. Capeletti and M. G. Bridelli, in *Proceedings of the 10th International Symposium on Electrets, Delphi–Athens, 1999*, p. 213.
7. S. Mascarenhas, *J. Electrostat.* **1**, 141 (1975).
8. A. Reiche, N. Nedeskam, and A. Mayer, *J. Phys. Chem.* **74**, 2659 (1970).
9. S. Mascarenhas, *Ann. (N.Y.) Acad. Sci.* **238**, 36 (1974).
10. I. van Turnhout, *Thermally Discharge of Polymer Electrets* (Elsevier, Amsterdam, 1975).
11. *Human Physiology*, Ed. by G. I. Kositskiĭ (Meditsina, Moscow, 1979).
12. V. M. Stepanov, *Molecular Biology. Structure and Function of Albumen*, Ed. by A. S. Spirin (Vysshaya Shkola, Moscow, 1996).
13. S. Hauptmann, J. Graefe, and H. Remane, *Lehrbuch der Organischen Chemie* (VEB Deutscher Verlag für Grundstoffindustrie, Leipzig, 1976).
14. V. Tomaselli and M. Shamos, *Biopolymers* **12**, 353 (1973).
15. D. G. Knorre and S. D. Myzina, *Biological Chemistry* (Vysshaya Shkola, Moscow, 1998).
16. A. V. Makarevich, L. S. Pinchuk, and V. A. Gol'dade, *Electric Fields and Electro-active Materials in Biology and Medicine* (IMMS NANB, Gomel, 1998).

Translated by V. Isaakyan

Improved Sensitivity of Refractive Interface Detection in Capillary Methods of Separation

A. I. Gorshkov, O. V. Oshurkova, and V. B. Konstantinov

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

Received July 17, 2000

Abstract—It is demonstrated that the sensitivity of the refractive method for detecting solution–solution interfaces can be increased by optimally choosing the ratio of the outer and inner capillary diameters. © 2001 MAIK “Nauka/Interperiodica”.

The method for detecting electrolyte interfaces in transparent glass or quartz capillaries of a constant cross section (outer diameter 1.0–1.6 mm, and inner diameter, 0.5–0.8 mm) dates back to early in the 1970s [1]. A capillary with solutions was centrally located in a parallel beam of white light cut by a 0.3- to 1.8-mm-wide slit with sharp rectilinear edges. The “image” (resulting optical field) observed on the screen placed at a distance of 4–20 cm behind the capillary represents a series of rectangular contours. Their width is proportional to the refractive index of the solution in a given site of the capillary [2]. Each rectangle is filled by straight light lines parallel to the outer contour and concentrating near the axis. The outer contour and the inner lines consist of colored bands. A narrow-band transmission filter makes it possible to observe the pattern in the corresponding color. As the capillary–screen distance increases, the rectangles widen, the inner line spacing grows, and the image quality degrades. Figure 1 shows an example of the “image” of a solution–solution interface.

Thus, the method makes it possible to visualize the interface in a transparent capillary provided that the solutions have different refractive indices. There are many alternative methods for detecting the interfaces in capillaries [3]. The advantages of the method proposed are as follows. First, the method is sensitive to the refractive index—the property inherent in any solution (hence, its general applicability). Second, the method is “extended” in that it simultaneously detects all the interfaces inside a capillary and gives a chance to observe the process of separation, unlike point methods. However, its sensitivity is low. If the width of the capillary “image” is to be measured with an accuracy of 0.1 mm, our optical method is applicable for solutions with a concentration ranging from 0.1–0.2 mol/l to saturation. Most point methods for detecting interfaces in capillaries are good at concentrations of 0.001 mol/l or even lower. Therefore, it is expedient to improve the

sensitivity of the refractive method, which was the purpose of this work.

An LGN-208A laser with a wavelength of 0.63 μm was used as a light source in our experimental setup. The beam was telescoped in the horizontal direction by a combination of short- and long-focus cylindrical lenses. The width of the horizontal optical slit could manually be varied from 0.5 to 2.0 mm. A micrometer screw allowed the vertical adjustment of the optical knife, placed in front of the capillary at a distance of 0.3 mm. The distance from the capillary to the screen was no less than 5 cm. Substituting a laser for a white light source (with or without a color filter) is natural for to day’s experimentation. Having passed through the second lens, the laser beam represents a ~20-mm-wide stripe with the intensity decreasing at the edges.

The optical-knife screening of the upper half of the capillary makes only the lower part of the image accessible for observation. This means that imaging beams cross the optical axis between the capillary and the screen. We used a graphical approach in terms of geometrical optics to construct an image. Assume that the capillary is made of glass with a refractive index of 1.50 and filled with distilled water with a refractive index of 1.330. The refractive index of surrounding air is 1.000.

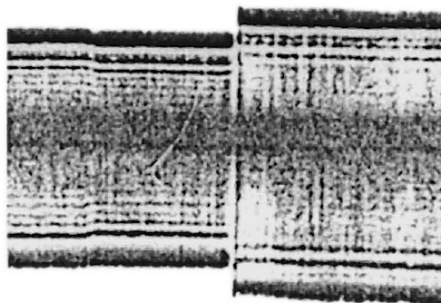


Fig. 1. Negative image of a capillary filled with solutions with different refractive indices.

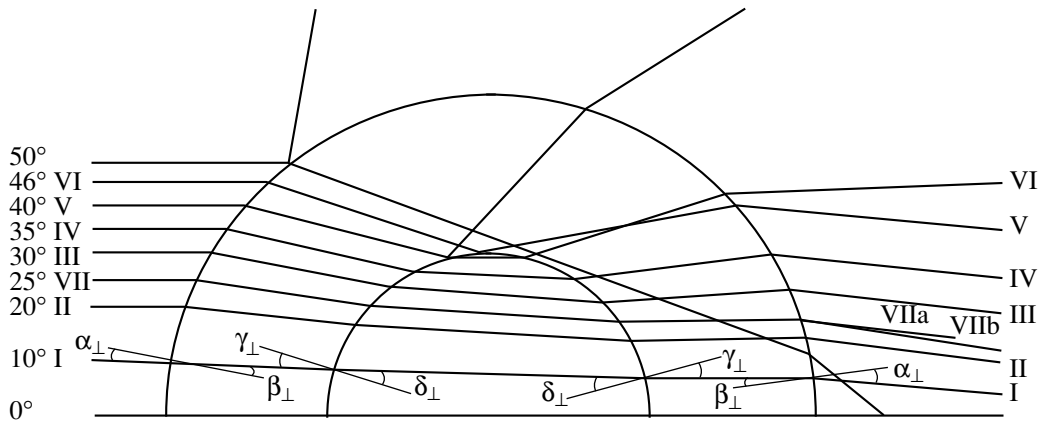


Fig. 2. Beam trajectories in the capillary filled with distilled water and in surrounding air. Shown are the angles of incidence on the outer wall of the capillary.

Beams I–VII in Fig. 2 cross the axis after leaving the capillary and, hence, form the image. The ratio of the outer diameter D to the channel diameter d , $D/d = 2.00$ (see Fig. 2), is typical of earlier practical applications. The plot of the angle of departure of the beams versus their angle of incidence exhibits a maximum (curve *a* in Fig. 3). It is seen that the maximum of 7.1° is achieved at the angle of incidence 25° (beam VII). Thus, it is clear that the beam with the highest refraction (VII) forms the outer contour of the “image,” whereas the

other beams with angles of incidence of 0° – 38° fill the inner area up to the axis. The beam with an angle of incidence of 38° (not shown in Fig. 2) leaves the capillary parallel to the axis. There are two groups of beams with angles of incidence of 0° – 25° and 25° – 38° . The beams of either group occupy the entire space within the half-image under consideration. Thus, what we observe within the image contour is interference fringes.

To be certain that the above mechanism is valid, let us analyze the trajectories of the other beams. The beams with angles of incidence of 38° – 44° (the extreme beam is beam V in Fig. 2) are deflected from the axis and do not take part in imaging. The beams with angles of incidence of 45° – 49° experience total internal reflection at the glass–water interface. The beams with angles of incidence of 50° – 90° travel only inside the glass and, being deflected by angles exceeding 38° , do not take part in imaging.

Note that, in addition to the beams refracted, we must consider the reflected ones, since there are four interfaces between the media with different refractive indices. The beams reflected from the outer wall of the capillary (one of them is shown in Fig. 2 for an angle of incidence of 50°) do not form the image. Let us turn to the beams reflected from the second (glass–water) interface to the glass. If the angle of incidence of such a beam ranges from 46° to 48° , the beam leaves the capillary at the “appropriate” angle and takes part in imaging (e.g., beam VII in Fig. 2). However, the image is formed mainly by the beams refracted into the water channel. To prove this, we carried out experiments in which the capillary was filled by water with a high concentration of black ink instead of by distilled water. The image did not form.

Now let us consider the effect of the refractive index of the capillary liquid on the width of the image. The beam that forms the outer contour of the image leaves the capillary filled with water at the maximum angle of 7.1° (beam VIIa in Fig. 2). If the capillary is filled by a

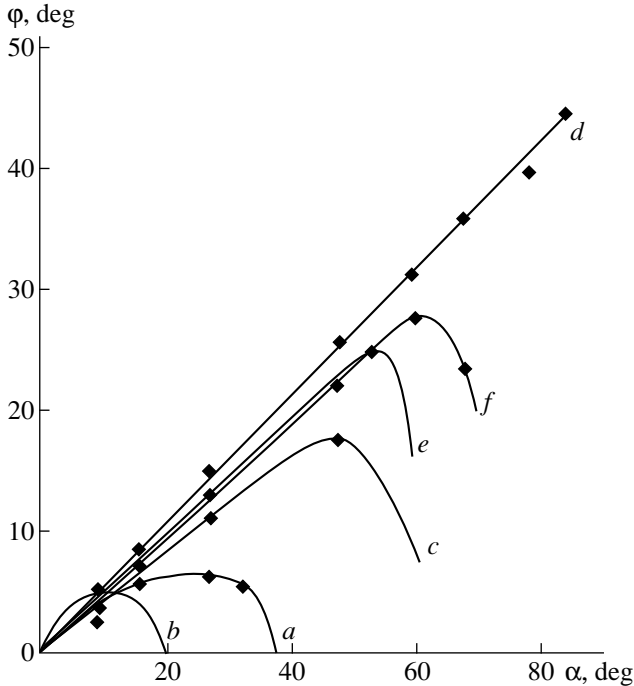


Fig. 3. Graphically obtained plots of the angle of departure (φ) versus angle of incidence (α) at the air–capillary interface for distilled-water-filled capillaries: (a) $D/d = 2.0$, $n = 1.5$; (b) $D/d = 2.857$, $n = 1.5$; (c) $D/d = 1.538$, $n = 1.5$; (d) $D/d = 1.33$, $n = 1.5$; (e) $D/d = 1.538$, $n = 1.4$; and (f) $D/d = 1.396$, $n = 1.5$.

transparent liquid with a refractive index of 1.340, this angle increases to 8.0° (beam VIIb). Experiments indicate that the image width does increase in the latter case [2].

Let us introduce the notion of the sensitivity of the refractive method (s). We define the sensitivity as $s = \Delta n / \Delta h$, i.e., as the difference between the refractive indices of two liquids filling the capillary (Δn) when the difference in the widths of the corresponding images (Δh) equals 1 mm at a capillary–screen distance of 5 cm. We can easily obtain the sensitivity from the above graphical data: $s = n/2L(\tan 8.0^\circ - \tan 7.1^\circ) = 0.0065$ n/mm. Here, $L = 5$ cm is the distance to the screen. The sensitivity of a conventional detector with $D/d = 1.8$ is $s = 0.007$ 1/mm [2] (for simplicity, s will be expressed in 1/mm).

A decrease in the ratio D/d increases the sensitivity of the refraction method. The angle of departure versus angle of incidence plots for distilled-water-filled capillaries made of the material with a refractive index of 1.50 retain the peaks up to $D/d = 1.39$, and the height of the peak grows (Fig. 3). Specifically, the maximum angles of departure are 17.5° and 27° for $D/d = 1.538$ and 1.396, respectively. The graphical determination of the sensitivity for the latter ratio of the diameters yields the value 0.001 1/mm, which is 3.5 times higher than that for $D/d = 2.0$. The peak in the plot disappears when D/d decreases to 1.33 (curve d in Fig. 3). In this case, the image of the capillary forms in a different way. Experiments show that the outer contour of the image represents a 10-mm-wide noisy fringe at a distance of 5 cm to the screen. Therefore, such a capillary is intrinsically inapplicable for our detection method.

It is evident that there must be a curve between curves f and d in Fig. 3 that has no peak and tends to saturation on the right of the figure. Here, the sensitivity is the highest provided that the ratio D/d is properly chosen.

It was found that the variation of the refractive index of the capillary from 1.4 to 1.6 leads to minor changes in the ratio D/d at which the peak disappears. This is illustrated with curves c and e in Fig. 3, which were obtained for the same ratio $D/d = 1.538$ but different refractive indices of the capillary (1.50 and 1.40, respectively).

Chernov *et al.* have elaborated the approximate theory that accounts for the peak in the angular dependence [4].

Capillaries used for optical detection of the interfaces are produced by rapidly drawing a heated tube having the required ratio of the diameters (apparently, some similarity principle sets in). No additional processing of the capillary walls is needed.

It was mentioned that the beam that is deflected through the maximal angle forms the outer contour of the image. Therefore, a greater distance from the capillary to the screen could increase the sensitivity of the

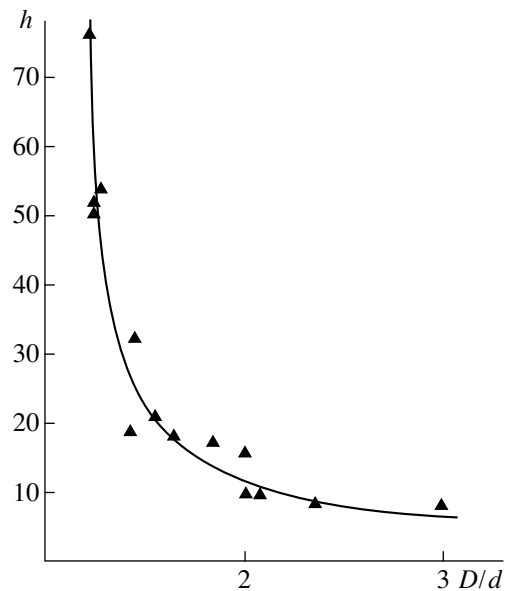


Fig. 4. Experimental plot of the capillary image width versus ratio D/d . The capillary is made of glass and filled by water. The distance from the capillary to the screen is 5 cm.

detection method. In practice, however, the image quality degrades at large distances. Moreover, when the screen is 70 cm distant from the capillary, the outer contour and the adjacent fringes by themselves consist of more closely spaced interference fringes. Thus, we must choose the distance optimal in terms of sensitivity and a method of measuring the image width. This problem calls for further investigation.

The following experimental fact has also found explanation. The image of a water-filled capillary disappears if it is placed into a water-filled plane-parallel transparent cuvette for thermostatic control. The graphical analysis similar to that for the capillary with $D/d = 2.0$ shows that the beams refracted at the capillary–thermostat water interface are deflected up from the optical axis and give no peak in the associated curve. Consequently, the image of the capillary with the solution on the screen is absent. This holds for the refractive index of the capillary wall from 1.4 to 1.6. Thus, thermostatic control in capillary separation methods using the detection technique proposed should be provided by a gas flow (see, i.e., [5]).

The results of the graphical analysis agree well with experimental data. Figure 4 shows an experimental plot of the capillary image width versus D/d for distilled-water-filled capillaries and the capillary–screen distance 5 cm. The scatter in the data points is apparently related to a possible difference in the refractive indices of the capillary glasses.

In the above considerations on the image formation, we dealt only with the ratio D/d without regard for the smallness of the capillary diameter. This means that the peak in the angular dependence does not depend on the

outer diameter provided that the ratio D/d is appropriately chosen and the beam width is larger than D . For this effect to be observed in large-diameter tubes, the optical uniformity of the material must be high. In drawn capillaries, high uniformity is provided in a natural way, as demonstrated with specially selected glass tubes with a diameter of about 2.0 mm.

In the case of spherical symmetry, we would have a thick-wall sphere filled by water. Illumination of the sphere with an appropriate ratio of the diameters by a plane monochromatic beam will allow the observation of Newton rings obtained in such an unusual manner.

CONCLUSIONS

(1) The outer contour of the "image" of a water- or solution-filled capillary is formed by the beam deflected from the axis by the largest angle. Therefore, the width of the capillary image is proportional to the refractive index of the filling liquid.

(2) The sensitivity of the method can be improved four- or fivefold (with the image width measured with the same accuracy) if the ratio D/d is optimal.

(3) The sensitivity of the method can additionally be increased by using a capillary material with a refractive index of 1.4.

(4) Refraction plays the decisive role in forming the capillary image. Therefore, the term "refractive detec-

tor" (instead of the earlier "diffraction detector") seems to be more appropriate for the associated device.

Recall that our method is a versatile and "extended" technique to observe interfaces capillary separation. It will be competitive with the alternative techniques if its sensitivity is improved.

ACKNOWLEDGMENTS

We are grateful to E.B. Shadrin for valuable discussions

REFERENCES

1. B. P. Konstantinov and O. V. Oshurkova, Dokl. Akad. Nauk SSSR **148**, 1110 (1963).
2. B. P. Konstantinov, N. S. Lyadov, and O. V. Oshurkova, Zh. Tekh. Fiz. **38**, 2117 (1968) [Sov. Phys. Tech. Phys. **13**, 1702 (1969)].
3. O. V. Oshurkova and A. I. Gorshkov, Élektrokimiya **26**, 532 (1990).
4. S. M. Chernov, K. K. Zhilik, and P. G. Rabzonov, Zh. Prikl. Spektrosk. **37**, 455 (1982).
5. P. I. Bresler, I. A. Ivanova, O. V. Oshurkova, and G. A. Shtilerman, Zh. Anal. Khim. **36**, 593 (1981).

Translated by A. Chikishev

BRIEF COMMUNICATIONS

Computer Simulation of Crystal Growth from Solution

Yu. Yu. Tarasevich

Astrakhan State Pedagogical University, ul. Tatishcheva 20a, Astrakhan, 414056 Russia

Received June 29, 2000

Abstract—In [1], for computer simulations of salt crystal growth from biological fluids, the so-called τ -model was put forward. This model was used to describe the crystal growth under different conditions [1–5]. However, our independent investigation indicates that the conclusions obtained with the τ -model are not confirmed. The effects observed are related to mistakes in the algorithm. © 2001 MAIK “Nauka/Interperiodica”.

Methods for investigating structures formed by salt crystallization from biological fluids (plasma, lymph, saliva, etc.) have found increased application in medical diagnostics over the past few decades, because they are simple and provide much information. In addition, they are suitable for proximate examination of large groups of population.

In diagnostics, rapid (for 2–4 h) drying of samples is usually applied. In this case, salt crystals precipitate as grains or dendrites, and the biological fluid itself forms a complicated lobed structure.

In the absence of biological components, the growth of salt crystals is simulated in the framework of the τ -model [1, 2]. The model was also employed to describe two-component media. It characterizes cyclic changes in dendrite shape (kinetic recurring phase transitions) [3, 4]. A simplified version of the τ -model was used to describe self-similarity [5].

However, our investigations did not confirm the conclusions obtained in its framework. The effects observed are related to mistakes in the algorithm.

Consider a thin layer of a supersaturated common-salt solution on a substrate. In the direction perpendicular to the substrate, the variation in the salt concentration is assumed to be negligible. Then, the salt concentration obeys the diffusion equation

$$\frac{\partial C}{\partial t} = D \left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} \right), \quad (1)$$

where C is the salt concentration in the solution and D is the diffusion coefficient.

To solve this equation numerically, it is necessary to set initial conditions, boundary conditions, and a computational scheme. At the initial time instant, the solution concentration in the interior is assumed to be constant and equal to $1.1C_0$, where C_0 is the saturation concentration of the solution.

There exist a number of ways to set boundary conditions. Since in [1] boundary conditions are not considered, we will briefly analyze them. If the size of a

region for which calculations are performed were equal to that of crystallizing droplet, then the zeroth solution concentration at the boundary would be a natural boundary condition. In [1], simulation is carried out in the region with linear dimensions of 10^{-2} cm, which is two orders of magnitude smaller than the actual system being simulated. If we assume that this region is isolated, then, for a given supersaturation of the solution, the volume of a grown crystal will be about 2% of the entire solution volume. Since our region constitutes only a small part of the entire system, it is reasonable to set a fixed concentration or a fixed flow at the boundary of the region. The latter condition means that the surroundings have a higher solution concentration and, with time, would lead to a rise in the concentration in the region under consideration. Therefore, the most adequate boundary condition is a constant solution concentration equal to that in the interior at the initial time instant.

When a crystal nucleates in the solution, the mass conservation law must be fulfilled at the liquid–solid interface. This results in the following boundary conditions [6]:

$$D \frac{\partial C}{\partial n} = (\rho - C)V,$$

where $\partial C/\partial n$ is the derivative along the normal to the interface, ρ is the salt density, and V is the crystal growth rate.

Taking into account the expression for growth rate [6], we have

$$D \frac{\partial C}{\partial n} = \frac{(\rho - C)(C - C_0)}{\rho} \beta, \quad (2)$$

where β is the kinetic coefficient of crystallization.

To numerically solve Eq. (1), we apply the finite-difference method and an explicit two-layer symmetric

Table

β , cm/s	Crystal growth rate $V = \frac{C - C_0}{\rho} \beta$, cm/s	Time (L/V) the liquid–solid interface takes to travel a distance L , s	Growth rate V_1 in the $\langle 100 \rangle$ direction, cm/s		L/V_1 , s	
			data from [1]	computer experiment	data from [1]	computer experiment
1	1.6×10^{-2}	6.1×10^{-3}		1.1×10^{-3}		8.9×10^{-3}
10^{-1}	1.6×10^{-3}	6.1×10^{-2}		7.5×10^{-4}		1.3×10^{-1}
10^{-2}	1.6×10^{-4}	6.1×10^{-1}		3.1×10^{-4}		3.2×10^{-1}
10^{-4}	1.6×10^{-6}	6.1×10^1	9.26×10^{-4}	4.8×10^{-6}	1.108	2.1×10^1
10^{-5}	1.6×10^{-7}	6.1×10^2	4.06×10^{-4}	4.9×10^{-7}	0.246	2×10^2
10^{-6}	1.6×10^{-8}	6.1×10^3	8.9×10^{-5}	4.9×10^{-8}	1.124	2×10^3

scheme with equal steps L along the spatial coordinates and a time step τ :

$$\frac{C_{i,j}^{\tau+1} - C_{i,j}^{\tau}}{\tau} = D \left(\frac{C_{i+1,j}^{\tau} - 2C_{i,j}^{\tau} + C_{i-1,j}^{\tau}}{L^2} + \frac{C_{i,j+1}^{\tau} - 2C_{i,j}^{\tau} + C_{i,j-1}^{\tau}}{L^2} \right).$$

Having rearranged the terms and multiplied both sides by L , we obtain the expression for flow of matter inside the liquid:

$$I = \frac{m}{L^2 \tau} = \frac{D}{L} ((C_{i+1,j}^{\tau} - C_{i,j}^{\tau}) + (C_{i-1,j}^{\tau} - C_{i,j}^{\tau}) + (C_{i,j+1}^{\tau} - C_{i,j}^{\tau}) + (C_{i,j-1}^{\tau} - C_{i,j}^{\tau})).$$

At the interface, the expression for the flows takes

the form

$$I = \frac{m}{L^2 \tau} = \frac{(\rho - C_{i,j}^{\tau})(C_{i,j}^{\tau} - C_0)}{\rho} \beta.$$

The above difference scheme is completely coincident with that used in [1] in view of the corrections made in [3]. The only difference is that, in [3], it is referred to as a method of diffusion fluxes. This difference scheme is stable only under the condition [7, 8]

$$\tau \leq \frac{L^2}{4D},$$

which renders it unfeasible. In practice, such problems are solved with implicit schemes. Nevertheless, we applied just this scheme to be able to compare our results and those obtained in [1]. The time step in [1] is twice as large as the limiting value, which casts doubt on the results obtained in that article. In addition, when calculating the variation of the salt concentration in a mesh, the authors of [1] mistakenly assume that if the mesh has three neighbors with the same or a greater concentration, the salt mass lost by any of them for a period τ is equal to $IL\tau/3$. However, it is easy to check that such a quantity of the salt leaves the mesh for $\tau/3$. If any mesh at any time step had the same number of neighbors with equal concentrations, this error and the improper choice of the time step would merely compensate each other, making the computational scheme stable. Actually, however, different meshes have a different number of neighbors at various time steps. Therefore, once one step of the calculation has been completed, the states of different meshes will correspond to different time steps.

For the parameters used in [1], the crystallization process is quasi-equilibrium; i.e., the solution concentration has time to fit a new position of the interface. Indeed, the relaxation time, $\tau_0 = L^2/D$ is on the order of 10^{-3} s, which is at least two orders of magnitude smaller than the characteristic time of crystal growth (see table). Thus, the statement that the crystal growth is simulated under nonequilibrium conditions has little

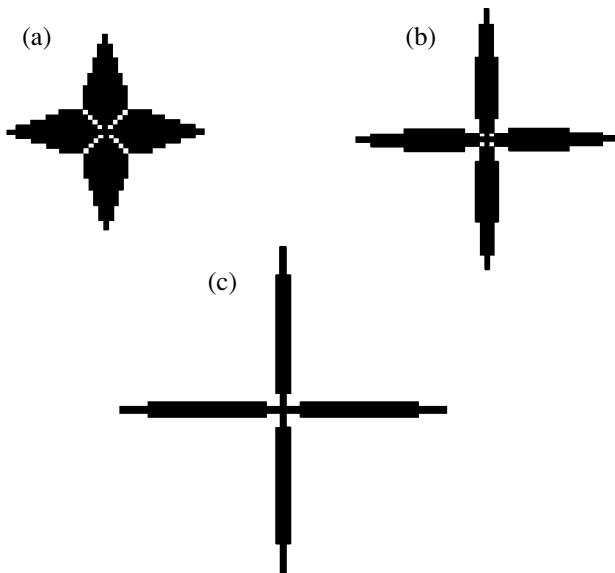


Fig. 1. Crystal shapes at the kinetic coefficients $\beta =$ (a) 10^{-2} (quasi-equilibrium regime; the crystal shape does not change at lesser β), (b) 10^{-1} , and (c) 10 cm/s.

force. The quasi-equilibrium growth of crystals is characterized by the skeleton rather than by the dendritic form [6]. The formation of dendrites is due to growing instabilities when a crystal deviates at random from its equilibrium shape. Since a random deviation in the framework of this model is impossible and instability mechanisms are not involved, there are no apparent reasons for dendrites to appear. Our calculations are shown in Fig. 1. Both in the kinetic (quasi-static) and in the diffusion regimes, only the skeleton crystals are observed.

The data in the table are again evidence that the results obtained in [1] are invalid. The table compares our theoretical estimates of the crystal growth rate with the results from [1]. The crystal growth rates obtained in [1] exceed those of the theoretical estimations and numerical simulation by several orders of magnitude. Our calculations were performed for the same parameters as in [1]: $C_0 = 0.36 \text{ g/cm}^3$, $\rho = 2.2 \text{ g/cm}^3$, $D = 1.5 \times 10^{-5} \text{ cm}^2/\text{s}$, and $L = 10^{-4} \text{ cm}$.

In [3], salt crystal growth in the presence of a protein (albumin) is simulated. The kinetic coefficient of crystallization and the relative supersaturation are assumed to be equal to $2.2 \times 10^{-3} \text{ cm/s}$ and 0.9, respectively. With reference to [1], it is argued that it is these values of the parameters that correspond to the growth of skeleton salt crystals in the absence of the protein. In [1], however, the relative supersaturation used was 0.1 and the skeleton shapes were observed at the kinetic coefficient of crystallization 10^{-6} cm/s . At greater values of the coefficient, the crystal had the dendrite shape. Thus, the results obtained by the same authors in the framework of the same model are in obvious conflict. In addition, the value 0.9 of the relative supersaturation is open to question. In [3], it is supposed that the concentration of the protein and its transition to the gel phase are governed by diffusion equation (1) and boundary condition (2) with the appropriate diffusion

coefficient. If there were two different salts in the solution, such a transition would be possible. However, it can hardly occur in a salt-protein system. Nowadays, the percolation theory is commonly accepted in describing the process of gelatination (see, for example, [9]).

Thus, our independent verification does not confirm the main data obtained in [1–5], which resulted from the improper computational algorithm.

REFERENCES

1. L. M. Martyushev, V. D. Seleznev, and S. A. Skopinov, *Pis'ma Zh. Tekh. Fiz.* **22** (4), 28 (1996) [*Tech. Phys. Lett.* **22**, 146 (1996)].
2. L. M. Martiouchev, V. D. Seleznev, *et al.*, *J. Stat. Phys.* **90**, 1413 (1998).
3. L. M. Martyushev, V. D. Seleznev, and S. A. Skopinov, *Pis'ma Zh. Tekh. Fiz.* **22** (16), 12 (1996) [*Tech. Phys. Lett.* **22**, 648 (1996)].
4. L. M. Martyushev, V. D. Seleznev, and S. A. Skopinov, *Pis'ma Zh. Tekh. Fiz.* **23** (13), 1 (1997) [*Tech. Phys. Lett.* **23**, 495 (1997)].
5. L. M. Martyushev and V. D. Seleznev, *Pis'ma Zh. Tekh. Fiz.* **25** (20), 71 (1999) [*Tech. Phys. Lett.* **25**, 833 (1999)].
6. A. A. Chernov, E. I. Givargizov, Kh. S. Bagdasarov, *et al.*, *Modern Crystallography*, Ed. by B. K. Vaĭnshteĭn, A. A. Chernov, and L. A. Shuvalov (Nauka, Moscow, 1980; Springer-Verlag, Berlin, 1984).
7. J. M. Ortega and W. G. Poole, Jr., *An Introduction to Numerical Methods for Differential Equations* (Pitman, Marshfield, 1981; Nauka, Moscow, 1986).
8. A. A. Samarskiĭ and A. V. Gulin, *Stability of Difference Schemes* (Nauka, Moscow, 1973).
9. D. Stauffer and A. Aharony, *Introduction to Percolation Theory* (Taylor and Francis, London, 1992).

Translated by Yu. Vishnyakov

A Correlation Method of Search for Higher-Order Angular Focusing

A. A. Trubitsyn

Ryazan State Radiotechnical Academy, Ryazan, 390000 Russia

Received March 21, 2000; in final form, August 8, 2000

Abstract—A numerical method for estimating the order of angular focusing, which is a most important parameter of electron-optical systems, is proposed. The method is based on the determination of the value of the correlation between a power function and a function of charged-particle entry angle, the latter being constructed by analysis of trajectories. The method is verified for systems admitting analytical solutions. © 2001 MAIK “Nauka/Interperiodica”.

The order of focusing is a parameter of major importance for electron-optical systems. For instance, this parameter determines the value of spherical aberrations for electron lenses, as well as specifies the degree of contradiction lying in the requirement of simultaneously achieving high resolution and transmission in mass and energy dispersion analyzers.

In contrast to analytical methods, numerical techniques make it possible to study systems that are very close to real ones. Our numerical method of search for higher-order angular focusing may be considered as an extension of the method reported in [1, 2]. In these papers, it is shown that the condition of N th-order focusing for the central angle of entry α_0 is given by

$$\delta_n(\alpha_0) = R^{(n)}(\alpha_0)t'(\alpha_0) - R'(\alpha_0)t^{(n)}(\alpha_0) = 0, \quad (1)$$

where $n = 2, 3, \dots, N$; α is the initial angle of motion; $R(\alpha) = y_c(\alpha) + x_c(\alpha)t(\alpha)$; $t(\alpha) = \tan(\gamma)$; and $x_c(\alpha)$, $y_c(\alpha)$, and $\gamma(\alpha)$ are the coordinates and the angle at which a particle escapes from the field gradient region.

However, numerical simulations of higher-order derivatives ($n > 2$) cause large errors. Therefore, the function $\delta_n(\alpha)$ cannot be calculated with a required accuracy.

In practice, this method has proven to be effective in searching for second-order focusing and inappropriate for higher-order focusing. Thus, the problem of adapting the method to search for conditions of higher-than-second order focusing arises. Having introduced the function

$$F(\alpha) = R''(\alpha)t'(\alpha) - R'(\alpha)t''(\alpha), \quad (2)$$

we can recast the condition for $(N+2)$ th-order focusing as a set of the equalities

$$F^{(n)}(\alpha_0) = 0, \quad n = 0, 1, 2, \dots, N. \quad (3)$$

The Taylor expansion of the function $F(\alpha)$ in the vicinity of α_0 is written as

$$F(\alpha) = F(\alpha_0) + F'(\alpha_0)(\alpha - \alpha_0) + (1/2)F''(\alpha_0) \times (\alpha - \alpha_0)^2 + \dots + (1/N!)F^{(N)}(\alpha_0)(\alpha - \alpha_0)^N + \dots$$

In view of (3), for $(N+2)$ th-order focusing, we have

$$F(\alpha) = 1/(N+1)!F^{(N+1)}(\alpha_0)(\alpha - \alpha_0)^{N+1} + 1/(N+2)!F^{(N+2)}(\alpha_0)(\alpha - \alpha_0)^{N+2} + \dots \quad (4)$$

Since $F(\alpha)$ given by (2) is a result of numerical simulations, this function is the superposition of the true values and noise. Therefore, it is reasonable to consider it as a random function of the nonrandom argument α . Let $S^*(\alpha)$ denote the function defined by (4). It is clear that

$$S^*(\alpha) \approx 1/(N+1)!F^{(N+1)}(\alpha_0)(\alpha - \alpha_0)^{N+1}.$$

We assume that $S^*(\alpha)$ is also a random function with negligible noise and $1/(N+1)!F^{(N+1)}(\alpha_0)$, a nonrandom factor. It is clear enough that the functions $F(\alpha)$ and $S^*(\alpha)$ are correlated in the case of $(N+2)$ th-order focusing.

The normalized mutual correlation function [3, 4] is used to estimate the dependence between the sections of two random functions. Since the normalized mutual correlation function does not change when a random function is multiplied by nonrandom factors, it is sufficient to study the correlation between $F(\alpha)$ and $S(\alpha) = (\alpha - \alpha_0)^{N+1}$ in order to estimate the correlation dependence between $F(\alpha)$ and $S^*(\alpha)$.

For a discrete change in the argument $\alpha = \alpha_1, \alpha_2, \dots, \alpha_L$ and the zero shift between the functions $F_i = F(\alpha_i)$ and $S_i = S(\alpha_i) = (\alpha_i - \alpha_0)^{N+1}$ ($i = 1, 2, \dots, L$), the normalized mutual correlation function is given by

$$\rho_0(N) = R_{FS}/(K_F K_S)^{1/2}, \quad (5)$$

where

$$R_{FS} = 1/L \sum_{i=1}^L (F_i - \bar{F})(S_i - \bar{S}),$$

$$K_F = 1/L \sum_{i=1}^L (F_i - \bar{F})^2, \quad K_S = 1/L \sum_{i=1}^L (S_i - \bar{S})^2,$$

$$\bar{F} = 1/L \sum_{i=1}^L F_i, \quad \bar{S} = 1/L \sum_{i=1}^L S_i.$$

The higher the degree of correlation between the functions F_i and S_i , the closer ρ_0 to unity. Here, we consider ρ_0 as a function of N , which defines the degree of the corresponding polynomial.

The technique of searching for higher-order focusing is as follows. In accordance with the method in [1, 2], the angle of second-order focusing α_0 is determined by solving the equation $F(\alpha) = 0$. Then, from formula (5), the correlation between the function $F(\alpha)$ (see (2)) and the power function $S(\alpha) = (\alpha - \alpha_0)^{m+1}$ is estimated for consecutively varying $m = 0, 1, \dots, M$, where M is the upper bound of search that meets a particular application. Finally, we determine the value of $N(0 \leq N \leq M)$ for which $\rho_0(N) = \max\{\rho_0(0), \rho_0(1), \dots, \rho_0(M)\}$. The closeness of $\rho_0(N)$ to unity will indicate a correlation between $F(\alpha)$ and $S(\alpha) = (\alpha - \alpha_0)^{N+1}$, i.e., the existence of $(N + 2)$ th-order focusing. Practically, the values of $M = 10\text{--}20$ are found to suffice.

In order to avoid large-number manipulation, one should normalize the function $S(\alpha) = (\alpha - \alpha_0)^{N+1}$ by dividing by the factor $(\alpha_{\max} - \alpha_{\min})^{N+1}$.

Formula (4) implies that, for high orders of focusing, $F(\alpha)$ is close to zero in a wide range of α . Therefore, spurious roots of the equation $F(\alpha) = 0$ may appear, since $F(\alpha)$ is noisy. In such cases, in order to refine α_0 , one should estimate the maximum value of the mutual correlation function $\rho_0(\alpha_0 N)$ with respect to two variables: central focusing angle and degree of polynomial. Below, our method is verified for models admitting analytical solutions.

As is known [5], a cylindrical reflector has second-order angular focusing at $\alpha_0 \approx 42^\circ$. The trajectory analysis of a cylindrical reflector by the technique described in [6] shows that second-order angular focusing at $\alpha_0 \approx 42^\circ$ takes place also in real devices with the edge field corrected by three pairs of adjusting rings. The corre-

sponding value of the normalized function of correlation between $F(\alpha)$ and $(\alpha - \alpha_0)^{m+1}$ takes the maximum value $\rho_0 = 0.99$ at $m = 0$.

An energy analyzer with three cylindrical electrodes [7] provides third-order focusing with the central angle $\alpha_0 \approx 40^\circ$. The normalized correlation function ρ_0 reaches its maximum value of 0.97 at $m = 1$ when the charged particle trajectories in an analytically defined electrostatic field are calculated by the Runge–Kutta method. The focusing parameters coincide with those obtained in [7] within the error of numerical integration of the equations of motion.

The trajectory analysis of a spherical mirror, which provides perfect angular focusing at $\alpha \approx 90^\circ$ [8], and the estimate of the corresponding correlation function, $\rho_0 > 0.95$ for the highest degree of the polynomial used in this paper ($N + 1 = 25$) allow us to conclude that electron-optical systems with high-order focusing can be studied by the technique described.

CONCLUSION

A correlation method for estimating the order of angular focusing in numerical analysis of electron-optical systems is proposed. It is tested for systems admitting analytical solutions.

REFERENCES

1. V. A. Gorelik, in *Proceedings of the VI All-Union Symposium on Secondary Electron and Photoelectron Emission and Spectroscopy of Solid Surfaces* (Radio, Moscow, 1986), p. 190.
2. V. A. Gorelik, O. D. Protopopov, and A. A. Trubitsyn, *Zh. Tekh. Fiz.* **58**, 1531 (1988) [*Sov. Phys. Tech. Phys.* **33**, 917 (1988)].
3. *Handbook of Mathematical Functions*, Ed. by M. Abramowitz and I. A. Stegun (National Bureau of Standards, Washington, 1964; Nauka, Moscow, 1979).
4. J. S. Bendat and A. G. Piersol, *Engineering Applications of Correlation and Spectral Analysis* (Wiley, New York, 1980; Mir, Moscow, 1983).
5. V. V. Zashkvara, M. I. Korsunskii, and O. S. Kosmachev, *Zh. Tekh. Fiz.* **36**, 132 (1966) [*Sov. Phys. Tech. Phys.* **11**, 96 (1966)].
6. A. A. Trubitsyn, *J. Electron Spectrosc. Relat. Phenom.* **73**, 305 (1995).
7. W. Franzen and J. Taaffe, US Patent 4,367,406, 1983.
8. H. Z. Sar-El, *Nucl. Instrum. Methods* **42**, 71 (1966).

Translated by M. Fofanov

Self-Organization of the Channel Structure of a Nanosecond Diffuse Discharge in a Wire–Plane Electrode System

P. B. Repin and A. G. Rep'ev

All-Russia Research Institute of Experimental Physics, Sarov, Nizhni Novgorod oblast, 607188 Russia

Received August 28, 2000

Abstract—The results of an experimental study of the spatial structure of a high-voltage diffuse discharge in a wire–plane electrode system are presented. Self-organization of the discharge current channels into regular cells is observed in the plane perpendicular to the electric field vector. The dependences of the structural parameters of the discharge in centimeter-sized gaps on the interelectrode distance are studied at air pressures within the range 220–760 torr. Self-organization of the discharge structure is explained in terms of the electric interaction among charges of the diffuse channel heads during bridging of the gap. © 2001 MAIK “Nauka/Interperiodica”.

High-voltage nanosecond diffuse discharge in centimeter-sized interelectrode gaps with a highly nonuniform distribution of the electric field is accompanied by the generation of fast electron fluxes and associated X-ray bremsstrahlung [1]. That is why this type of discharge is widely used to preionize the active medium of gas lasers [2] and efficiently clean air of organic pollutants [3]. The generation of ionizing radiation can be related to the channel or microchannel spatial structure of the discharge [1, 4]. This stimulates further investigations of the discharge structure.

Studies of the discharge showed that it had a regular multichannel structure. The main dependences of the structural parameters and mechanisms for the development of the discharge on the discharge initiation mode were determined [1, 4, 5]. However, the spatial structure of the discharge was detected in [4, 5] only by photographing the discharge from one side. Thus, only current channel projections on the plane passing through the wire cathode perpendicular to the anode were detected. In addition, the dependence of the geometric parameters of the current channels on the gas pressure remained unknown.

The goal of this work was to describe the results of studying the spatial structure of a nanosecond diffuse discharge in the anode plane. The discharge was initiated in air at a pressure of $p = 220\text{--}760$ torr.

A wire–plane electrode system was used in the experiments. The wire diameter was 0.2 mm and the wire length was 120 mm. The length h of the interelectrode gap ranged from 50 to 105 mm. Negative-polarity voltage pulses with a pulse rise time of 8 ns and total duration of 180 ns were generated at the wire cathode using a power source. The voltage amplitude ranged from 80 to 100 kV depending on the discharge initiation conditions. The discharge current amplitude increased from 20 to 860 A as the pressure and the gap

length decreased. The spatial structure of the discharge was detected by photographing the discharge as a whole and its individual regions. Another method of the discharge structure imaging was based on the traces left by the current channels on the anode coating. The discharge was photographed using a Zenit E camera with a Gelios 44-2 objective. The viewing angle was chosen so as to provide simultaneous detection of the spatial structure of the discharge and the distribution of the discharge channels over the anode surface. The photographs were taken at a distance of 0.2–0.4 m from the discharge. An isochromatic film with the spectral sensitivity boundary at 640 nm was used as the photosensitive material.

The anode was coated with a fine-dispersion foam material, carbon soot, or a commercial photoemulsion impregnated with an electrolyte. It was found experimentally that neither the general view of the discharge nor its electrical characteristics were modified by the presence of a detection coating on the anode.

Experiments show that carbon soot coating (with a thickness of ~ 0.05 mm and dispersion of ≤ 4 μm) provides the most detailed information when detecting the traces of discharge current channels. The channel trace on soot had the shape of a round imprint slightly different in hue from soot. The imprint could be seen in the reflected light. The silvery hue of the imprint was caused by the large number of microchannels constituting the imprint [1]. However, the processing of traces on the soot coating was rather difficult. Taking into account that the experiments were repeated many times because of the statistical scatter in the experimental data, soot coating was used mostly in control experiments and for studying the microstructural parameters of the discharge. The results of these studies will be published in a separate paper.

A fine-dispersion foam material was used instead of soot for studying the discharge microstructure. The main advantage of this material over soot is the significantly higher sensitivity to discharge current channels. In addition to the main current channels, a foam layer approximately 1 mm in thickness provided the detection of even hardly visible fragments. The resolution of the method depended on the material dispersion ($\leq 200 \mu\text{m}$). We note that the material dispersion increases gradually as the pressure decreases. For this reason, at reduced pressure, this method was used only to determine the number of current channels and their distribution over the anode surface.

Experiments showed that the discharge consisted of several separate diffuse channels distributed uniformly along the wire cathode (Fig. 1). Such a pattern is typical of high-voltage nanosecond discharges in the wire-plane gap [4, 5].

The mean number of the discharge channels decreased from 36 to 14 as the gap length increased from 50 to 105 mm. The number N_c of diffuse discharge channels was determined using the three methods mentioned above. Photographing of the discharge allowed N_c to be determined within the whole pressure range at small values of the interelectrode gap length h or within the whole range of h at a reduced pressure at which the discharge glow intensity was sufficient for detection. The number N_c of diffuse discharge channels determined photographically was somewhat smaller than the number N_c determined by the trace method. A special experiment combining the two methods showed that this discrepancy was caused by the coincidence of projections of some channels on the photofilm. In some cases, insufficient focusing of the channel images caused an additional measuring error. The total error in the number of channels determined by photographing was 20–40%. The values of N_c determined by the trace methods on the soot and the fine-dispersion foam layer were usually close to each other. The total error in the number of channels determined by the trace method did not exceed $\pm 5\%$; this error was caused by errors in the identification of traces. It was shown experimentally that, at most, one in every ten channel traces on the anode coating could be due to artifacts (e.g., a dust grain, a flaw on the soot surface, or a burst foam bubble) rather than the discharge.

Experiments showed that the regular spatial structure of the discharge manifested itself not only as the uniform distribution of the current channels along the wire cathode, but also as self-organization of the discharge channels into regular cells in the anode plane. The traces of channels of a single discharge pulse on a fine-dispersion foam layer are shown in Fig. 2. Analysis of photographs of the channel structure at various h showed that the channels in the anode plane were arranged in hexagonal order in several rows symmetrically about the cathode projection. An idealized pattern

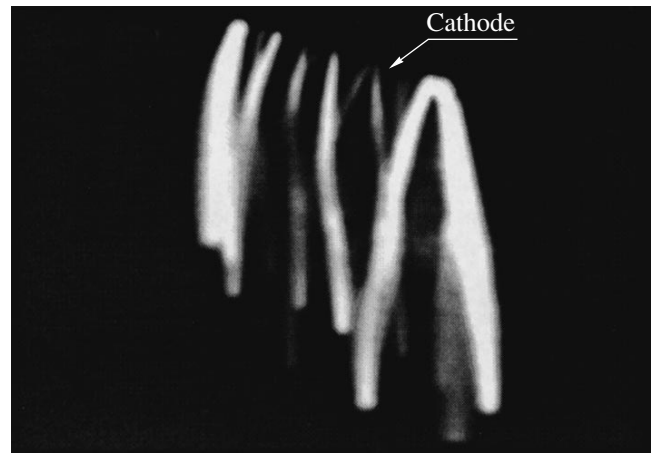


Fig. 1. General view of the discharge at $h = 74 \text{ mm}$ and $p = 310 \text{ torr}$.

of the channel distribution over the anode surface is shown in Fig. 3.

As the gap length h decreased from 100 to 60 mm, the trace density increased. The number of rows remained the same, but the number of channels in each row increased. The distance between the axes of the outermost rows remained almost the same (25 mm). As the gap length was further reduced to 50 mm, two additional outer rows of traces occurred; the distance between the rows adjacent to the central row decreased from 25 to 15 mm, while the distance between the axes of the outermost rows became equal to 33 mm. Within the gap length range $50 < h < 60 \text{ mm}$, the number of detected channels increased with decreasing h faster than within the range from 100 to 60 mm. We note that the observed pattern of the discharge structure organization is consistent with the results reported in [4], where a nonlinear dependence of the number of discharge channels on the interelectrode gap length was also observed.

As the gap length h increased from 50 to 80 mm, the mean diameter of the channel decreased only slightly from 1.1 to 0.9 mm. However, as the gap length increased to 105 mm, the diameter decreased to 0.4 mm. The diameters of the channel traces in the central rows of the pattern were usually larger than in the outer rows. The reason could be that the channel length increased with increasing distance from the cathode projection on the anode (the channel length corresponds to the length of the external field line [5]). This caused an increase in the electrical resistance of the channel and, therefore, a decrease in the electric current through the channel.

The number of channels and their distribution over the anode were almost independent of the pressure. A decrease in the number of discharge channels was observed only near the lower limit of the pressure range under study. The reason could be that the discharge current flowed partly through the visible regions of a dif-

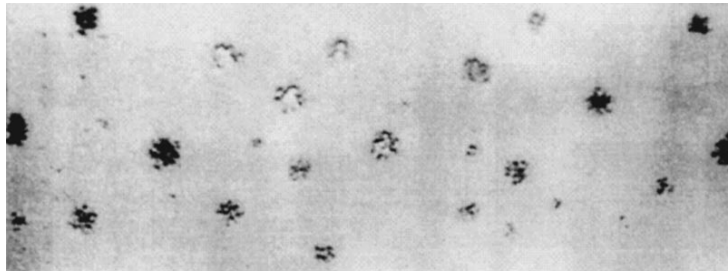


Fig. 2. Distribution of the discharge traces over the surface of an anode coated with a thin layer of a fine-dispersion foam material ($h = 74$ mm and $p = 400$ torr).

fuse glow observed at reduced pressure. Another reason could be that the sensitivity of the methods used to determine the number of channels became insufficient as the current density decreased. For $h = 74$ mm, three rows of traces were observed at all pressures. The distance between the axes of the outermost rows increased from 25 to 40 mm as the pressure decreased from 1 atm to $p \approx 300$ torr, whereas the mean channel diameter increased from 1 to 4.5 mm.

The process of self-organization of the discharge structure in the plane perpendicular to the electric field is best explained in terms of the electric interaction between the charges of the channel heads during the bridging of the gap [5].

The heads of the channels emerging from the cathode are at different distances from the anode because of the statistical scatter of the start times and axisymmetric distribution of the electric field near the wire (this distribution determines the initial direction of the channels). As the distance from the cathode increases and the field strength determined by the electrode geometry decreases, the superposition of the external electric field and the field produced by the adjacent channels becomes the main factor determining the propagation direction of each channel. The electric interaction

between the heads of the delayed channels and quasineutral columns of the adjacent channels of greater length results in the merging of delayed channels with adjacent channels [5]. Thus, the number of channels is reduced, while the mean distance between the channels increases and the distances from the anode to the channel heads equalize. This slows down the process of channel merging. The dependence of the number of channel traces on the anode coating on the inter-electrode distance becomes nonlinear. The propagation of the channels whose ends are approximately at the same distance from the anode is accompanied by the repulsion of their likely charged heads. As a result, the distribution of channels in the plane perpendicular to the external electric field becomes regular.

REFERENCES

1. S. N. Buranov, V. V. Gorokhov, V. I. Karelin, *et al.*, in *Research in Plasma Physics*, Ed. by V. D. Selemir and A. E. Dubinov (RFYaTs-VNIIEF, Sarov, 1998), p. 39.
2. S. N. Buranov, V. V. Gorokhov, V. I. Karelin, *et al.*, *Kvantovaya Élektron. (Moscow)* **18**, 891 (1991).
3. S. N. Buranov, S. V. Voevodin, I. A. Voevodina, *et al.*, in *Research in Plasma Physics*, Ed. by V. D. Selemir and A. E. Dubinov (RFYaTs-VNIIEF, Sarov, 1998), p. 339.
4. A. I. Pavlovskii, M. A. Voinov, V. V. Gorokhov, *et al.*, *Zh. Tekh. Fiz.* **60** (1), 64 (1990) [*Sov. Phys. Tech. Phys.* **35**, 37 (1990)].
5. S. N. Buranov, V. V. Gorokhov, V. I. Karelin, and P. B. Repin, *Teplofiz. Vys. Temp.* **29**, 383 (1991).

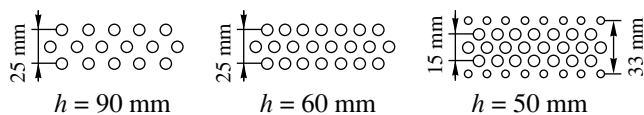


Fig. 3. Idealized pattern of the distribution of channel traces over the anode surface.

Translated by K. Chamorovskii

On the Dependence of the Free Carrier Concentration on Light Intensity in Photorefractive Crystals

N. A. Gusak and N. S. Petrov

Interbranch Institute of Advanced Training in New Areas of Technology,
State Polytechnical Academy, Minsk, 220107 Belarus

Received September 3, 1999; in final form, July 17, 2000

Abstract—The dependence of the free charge carrier concentration on light intensity is analyzed in two types of crystals where photorefractive centers are either traps or donors. Conditions under which the dependence becomes sublinear at relatively low light intensities are established. © 2001 MAIK “Nauka/Interperiodica”.

Although studies on photorefractive crystals date back to the 1970s, kinetic phenomena occurring in them are still not fully understood. In particular, the concentration of free carriers as a function of light intensity in photorefractive crystals remains unclear.

Numerical calculations showed [1] that the reciprocal of the characteristic time of space charge grating formation depends sublinearly on light intensity when it exceeds a certain critical value. Note that, in $\text{Bi}_{12}\text{SiO}_{20}$ (BSO) silicosillenite crystals, the sublinear dependence arises even at small intensities. The authors of [1] argue that their numerical data make it possible to revise the current theory of the photorefractive effect.

This circumstance makes us consider again the dependence of the free carrier concentration on the light intensity, since it is this dependence that causes the sublinearity. The aim of this study is to compare this dependence for two types of crystals in order to directly (without numerical calculations) find the sublinearity conditions in the operating range of light intensities.

We will use the basic model of a photorefractive medium [2]. This so-called one-center model assumes that a single energy level lies in the midgap of a solid. Electrons are excited into the conduction band from filled traps C^- either thermally or under the action of light of a certain frequency. Free electrons from the conduction band may recombine with the empty traps C^0 .

The time variation of the filled trap concentration is described by the kinetic equation

$$\frac{\partial N^-}{\partial t} = -(\beta + SI)N^- + \gamma N^0 N_e, \quad (1)$$

which was first applied to analyze the photorefractive effect in [3]. Here, N^- and N^0 are the concentrations of the C^- and C^0 centers, respectively; β is the probability of the thermal generation of free electrons; S is the cross section of optical absorption; I is the light intensity in the crystal; γ is the recombination coefficient;

and N_e is the electron concentration in the conduction band. The total concentration N of photorefractive centers during the process remains constant:

$$N = N^- + N^0. \quad (2)$$

From Eq. (1), it follows that the concentration of free carriers in the stationary state ($\partial N^-/\partial t = 0$) is determined by the expression

$$N_e = \frac{N^-}{N^0} n \quad \left(n = \frac{\beta + SI}{\gamma} \right). \quad (3)$$

Since N^- denotes the concentration of the bound negative charge, the electroneutrality condition requires the positive charge to be present in the medium. Consequently, a possible physical model of a photorefractive crystal should involve the existence of photorefractive centers with a concentration N and other positively charged centers with a concentration N_c (type-1 crystals). The latter are inactive and do not participate in light-induced kinetic phenomena.

At $\beta = I = 0$, the free charge is absent. In this case, $N^- = N_c$ and $N^0 = N - N_c$. Since the quantity N_c does not change, the appearance of N_e under illumination ($I \neq 0$) means that N^- decreases and N^0 simultaneously rises exactly by the same quantity N_e . Substituting $N^- = N_c - N_e$ and $N^0 = N - N_c + N_e$ into (3) yields the quadratic equation for N_e :

$$N_e^2 + (N - N_c + n)N_e - nN_c = 0. \quad (4)$$

Hence, we obtain

$$N_e = \frac{N_c}{N - N_c} n, \quad (5)$$

which is valid at small n satisfying the inequality

$$n \ll \frac{(N - N_c)^2}{2(N + N_c)}. \quad (6)$$

At large n obeying the condition

$$n \gg 2(N + N_c), \quad (7)$$

Eq. (4) provides

$$N_e^{\max} = N_c. \quad (8)$$

The N_e vs. I graph is a curve originating from some value of N_e^0 (at $I = 0$) and linearly rising in a certain interval of I . Further, the rise slows down and the curve asymptotically tends to N_e^{\max} .

There exists another possible physical model of a photorefractive crystal. Instead of traps, this model deals with photorefractive donor-type centers with a concentration N . In such a crystal (type-2 crystal), a fixed negative charge N_c is present. It appears when some donors donate their electrons to acceptors. The equal positive charge N^+ is on ionized donors. Under illumination, thermally excited free carriers with a concentration N_e appear. For this model, the kinetic equation can be written in the form

$$\frac{\partial N^+}{\partial t} = (\beta + SI)(N - N^+) - \gamma N^+ N_e. \quad (9)$$

From (9), we obtain the expression for the concentration of free carriers in the stationary state:

$$N_e = \frac{N - N^+}{N^+} n. \quad (10)$$

Substituting the quantity $N^+ = N_c + N_e$ into (10), we come to the quadratic equation for N_e . Its coefficients are expressed through the parameters of the medium and light intensity:

$$N_e^2 + (N_c + n)N_e - n(N - N_c) = 0. \quad (11)$$

It is seen that, at small values of n satisfying the inequality

$$n \ll \frac{N_c^2}{2(2N - N_c)}, \quad (12)$$

N_e is proportional to the parameter n :

$$N_e = \frac{N - N_c}{N_c} n. \quad (13)$$

In the case of large n , i.e., at

$$n \gg (2N - N_c), \quad (14)$$

the concentration N_e tends to its maximum:

$$N_e^{\max} = (N - N_c). \quad (15)$$

Equations (4) and (11) coincide only if $N_c = 1/2N$. In this particular case, the behavior of N_e as a function of I is totally independent of whether the photorefractive centers are traps or donors. An example of such a

crystal is a lithium niobate crystal doped by iron to $\sim 10^{-2}\%$. This crystal has the following parameters [4]: $N = 6.6 \times 10^{24} \text{ m}^{-3}$, $N_c = 3.3 \times 10^{24} \text{ m}^{-3}$, $\beta = 1 \text{ s}^{-1}$, $S = 6.2 \times 10^{-5} \text{ m}^2/\text{J}$, and $\gamma = 10^{-15} \text{ m}^3/\text{s}$. For this case, inequalities (6) and (12) yield

$$I \ll 10^{13} \text{ W/m}^2. \quad (16)$$

Hence, we see that the range of I where the linear behavior of N_e takes place is much wider than the operating range of light intensities, i.e., $I \sim 10^4 \text{ W/m}^2$.

At $N_c \neq 1/2N$, Eqs. (4) and (11) do no longer coincide and their solutions may significantly differ. From (12), it follows that, at $N_c \ll N$, the bound of inequality (16) shifts to much lower values, whereas (6) raises it only insignificantly. As N_c decreases as compared to N (the concentration of photorefractive centers is usually in the range $10^{24} - 10^{25} \text{ m}^{-3}$), this boundary may approach the operating range of I for the type-2 crystals. The type-1 crystals remain linear under these conditions.

To illustrate the aforesaid, consider type-2 BSO crystals, for which typically [5] $N = 10^{25} \text{ m}^{-3}$, $N_c = 0.95 \times 10^{22} \text{ m}^{-3}$, $\beta = 1 \text{ s}^{-1}$, $S = 1.06 \times 10^{-5} \text{ m}^2/\text{J}$, and $\gamma = 1.65 \times 10^{-17} \text{ m}^3/\text{s}$. Let us find a range of I where the N_e vs. I dependence is linear. According to (12), n must be much less than $2 \times 10^{18} \text{ m}^{-3}$. This means that $(\beta + SI) \ll 33 \text{ s}^{-1}$. The value $I = 10^5 \text{ W/m}^2$ does not satisfy this inequality. Strictly speaking, even the first term in the above inequality does not meet it. Having refined inequality (12) in view of this, we find that the N_e vs. I dependence is linear at $I < 10^4 \text{ W/m}^2$. This is in exact agreement with numerical results in [1].

It is important to find a quantitative test for the deviation of the concentration N_e from its linear approximation. For this purpose, the quantity

$$\eta = \frac{N_e' - N_e}{N_e' - N_e^0}, \quad (17)$$

where N_e' is the linear approximation of N_e , is convenient.

We will proceed from the rigorous solution of Eq. (11):

$$N_e = -\frac{N_c + n}{2} + \frac{N_c}{2} \sqrt{1 + \frac{2(2N - N_c)n}{N_c^2} \left[1 + \frac{n}{2(2N - N_c)} \right]}. \quad (18)$$

At small values of n , the term involving n and standing in brackets is negligible. Then, this solution can be represented as

$$N_e = -\frac{N_c + n}{2} + \frac{N_c}{2} \sqrt{\frac{p+1}{p}} \sqrt{1 + \frac{1}{p+1} \frac{n_1}{n_0}}, \quad (19)$$

where

$$p = \frac{N_c^2}{2(2N - N_c)n_0}. \quad (20)$$

Here, n_0 and n_1 are the constant and linear (in I) components of n , respectively.

We are interested in the range of small I , where the sublinearity occurs. Expanding, Eq. (19) in terms of a small parameter proportional to I , we obtain, to second-order accuracy,

$$\eta = \frac{1}{4(p+1)(1-m)n_0} \left(m = 2\sqrt{p(p+1)} \frac{n_0}{N_c} \right). \quad (21)$$

It is seen that the difference between the actual and linear dependences is proportional to the light intensity itself.

It is convenient to represent relationship (21) in the form of the dependence $I(\eta)$. Taking into account the smallness of m (for BSO, $m = 5 \times 10^{-4}$), we have

$$I = \frac{4(p+1)\beta}{S} \eta. \quad (22)$$

This expression makes it possible to directly determine the light intensity that corresponds to a preset value of η for any type-2 crystal.

According to (22), the light intensity is a parabolic function of N_c . It is minimum, $I_{\min} = 4\beta S^{-1}\eta$, at $p = 0$. For small p , a rise in I is slightly affected by an increase in N_c . For example, at $N_c = 2\sqrt{Nn_0}$, only the twofold growth of I_{\min} is observed. Further increasing N_c , however, results in a sharp rise in I .

Let us turn to the BSO crystal again. According to (20), to it, $p = 37$ and $I = 1.4 \times 10^5 \text{ W/m}^2$, there corresponds $\eta = 0.01$ (for such I , the difference between N_e and its linear approximation is 1%). Note that expression (22) is also appropriate for greater values of η (however, in this case, the associated value of I may be inaccurate). For example, at $\eta = 0.1$, $I = 1.4 \times 10^6 \text{ W/m}^2$,

while the true value, $I = 1.75 \times 10^6 \text{ W/m}^2$, calculated from exact expression (19) is 25% greater.

Expression (22) also allows us to make quantitative estimations when the values of N and N_c differ from sample to sample. From (20), a decrease in N_c by an order of magnitude (N remains the same) leads to $p = 0.37$ in the case of BSO. For such a sample, the value of $\eta = 0.01$ would be attained at $I = 5 \times 10^3 \text{ W/m}^2$, which is higher than I_{\min} only by 37%. However, an increase in N_c by one order of magnitude causes I to rise by two orders of magnitude. For example, if $N_c \approx N$ (as for lithium niobate), I increases by six orders of magnitude.

If the photorefractive centers in BSO were traps rather than donors, then, according to (6), the linear behavior of N_e would be observed in an extremely wide range of I that satisfies inequality (16).

It is of interest that, at $N_c = N - \alpha$ (where $\alpha \ll N$), a situation may arise when the type-1 crystals rather than the type-2 ones exhibit a sublinear N_e - I dependence at comparatively small light intensities.

ACKNOWLEDGMENTS

We thank V.N. Belov for valuable discussions.

REFERENCES

1. Nagendra Singh, S. P. Nadar, and Partha P. Banerjee, *Opt. Commun.* **136**, 487 (1997).
2. K. Buse, *Appl. Phys. B* **B74**, 273 (1997).
3. N. V. Kukhtarev, *Pis'ma Zh. Tekh. Fiz.* **2**, 1114 (1976) [*Sov. Tech. Phys. Lett.* **2**, 438 (1976)].
4. G. C. Valley, *IEEE J. Quantum Electron.* **19**, 1637 (1983).
5. P. M. Johansen, *IEEE J. Quantum Electron.* **25**, 530 (1989).

Translated by Yu. Vishnyakov

BRIEF COMMUNICATIONS

Carbon Flows from an Arc Discharge in Regimes Optimum for Production of Fullerenes

D. V. Afanas'ev, G. A. Dyuzhev, and A. A. Kruglikov

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

Received September 12, 2000

Abstract—Radial flows of carbon particles from the interelectrode gap of an arc discharge in modes optimum for production of fullerenes are measured. From the experimental data conclusions are made concerning transformations of the carbon particles in a gas-plasma jet. © 2001 MAIK “Nauka/Interperiodica”.

1. Arc discharge between graphite electrodes in a helium atmosphere is the most simple and effective means of producing fullerenes, but the scale of experimental studies of this kind of discharge is still inadequate [1–5]. It is known that in a usual high-current arc discharge at pressures close to normal, a strong axial plasma flow arises as a result of both thermal processes and the inherent magnetic field of the arc [6]. If the separation between electrodes is small, this flow can transform into a fanlike jet exiting the interelectrode gap [7]. In the case of a fullerene arc carbon atoms vaporized from the anode are present in the jet. In the exiting carbon vapor, transformations of carbon atoms into fullerenes, associates, nanoclusters, aggregates, and soot particles take place [8]. In this work, we attempted to obtain data on these transformations using carbon collecting probes.

2. The arc was burning between two vertical graphite electrodes 6 mm in diameter in a cylindrical vacuum chamber 180 mm in diameter. Most of the measurements were carried out in a mode optimum for producing fullerenes: an arc current of 70 A, a helium pressure of 60 torr, and an interelectrode separation of 4 mm. Around the arc, at different distances r from its axis, carbon-collecting probes consisting of tungsten wires of a diameter of $2a = 1.5$ mm and a length of $L = 60$ mm were arranged parallel to the arc axis. After the termination of the arc, the weight $m(r)$ of the soot deposited on the probes and the total weight m_0 of the carbon that exited the interelectrode gap of the arc discharge were determined.

3. Deposition of a material on the probe in a gas flow is usually characterized by a capture parameter $\epsilon(r)$ [9], which is the ratio of the material flow to the flow of material particles in a rectilinear path. The capture parameter of a probe is given by formula $\epsilon(r) = (m/m_0)/a/\pi r$, where a is the probe radius.

As carbon from the arc is deposited on the probe, its effective diameter increases. Figure 1a shows variations with time of the maximum probe deposit diameter D_{\max} , which was determined from probe pictures taken

in the burning arc. It is seen that for times of $t < 3$ min, the experimental data can be approximated with a straight line. The lack of coincidence of the probe diameter extrapolated to $t = 0$ with the initial probe diameter (for $r = 12$ mm) is possibly due to the high initial erosion in the arc. Variation of the shape of deposited carbon layer with time is insignificant (Fig. 1b). If the jet boundary is defined as a half-width of the distribution of the deposit over the probe length, then the jet divergence will be $\sim 100^\circ$, at least for $r < 30$ mm.

The quantity m_0 is affected by many, often uncontrollable factors, and the function $m(r)$ should be determined in one experiment. Therefore, in all experiments an arc burning time of $t = 3$ min was used, which ensured that at small r the wire diameter increased almost linearly and at large r the deposited material could be reliably weighed. In calculating the capture parameter, the initial probe diameter was increased by a quantity $\delta = 1/4(D_{\max} - 2a)$ (linear averaging over time and probe length) and the quantity D_{\max} was measured after termination of discharge.

The dependences of the capture parameter on r are shown in Fig. 2. Data points in this figure are the averages of four experiments. Though m_0 values in these experiments differed by a factor of up to 1.5, the ratio

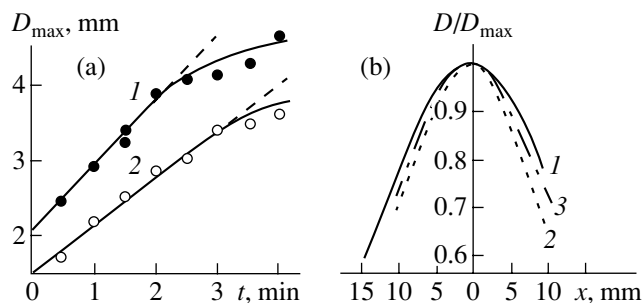


Fig. 1. (a) Dependence of the maximum probe diameter on the arc burning time at $r = 12$ (1) and 20 mm (2) and (b) the relative distribution of deposited carbon along the probe length at $r = 12$ mm; $t = 1$ (1), 2 (2), and 2.5 min (3).

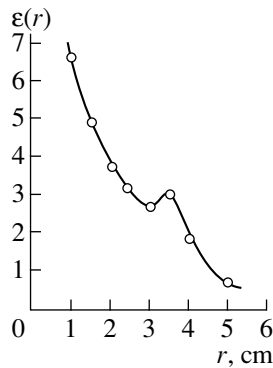


Fig. 2. Dependence of the dimensionless parameter $\varepsilon(r)$ on distance r from the discharge axis.

(m/m_0) is constant to within less than 10%. This is evidence that the jet properties depend weakly on the content of carbon added.

It is seen in Fig. 2 that at small r , the capture parameter $\varepsilon(r) > 1$. With increasing r , the capture parameter $\varepsilon(r)$ drops and, at $r > 50$, mm, becomes less than unity. The experiments have shown that $\varepsilon(r)$ diminishes as helium pressure increases. For example, at $r = 15$ mm, $\varepsilon(r) \sim 5$ at a helium pressure of 30 torr and ~ 2 at 180 torr.

4. The simplest estimates show that under the experimental conditions any carbon particles are completely entrained by the helium flow. In this case, if the flow about the probe is laminar or noneddying, the particles can get to the probe only by diffusion across the flow tubes.

Let us make some estimations. The scattering cross section σ was determined using an approximation of the elastic collision of spheres $\sigma = \pi/4(D_p + D_{\text{He}})^2$, D_p being the diameter of a carbon particle and D_{He} , the diameter of a helium atom, $D_{\text{He}} = 2.3 \times 10^{-8}$ cm. The free path is $\lambda = 1/\sigma N$. The concentration of helium atoms is $N = P_{\text{He}}/kT$. The diffusion coefficient is $D = 1/3\lambda \bar{v}$, \bar{v} being the thermal velocity of helium atoms.

Let the diffusion time τ be the time it takes a particle to traverse the region of the probe, $\tau = 2a/v_0$, v_0 being

the velocity of the undisturbed flow; then the diffusion length will be $L = (D\tau)^{1/2}$. In this case, particles from tubes at a distance of L from the zero tube can be deposited on the probe and the capture parameter can be estimated as $\varepsilon \sim L/a$.

In the table, estimates of the capture parameters for three kinds of carbon particles at a helium pressure of $P_{\text{He}} = 60$ torr and a probe diameter of $2a = 1.5$ mm are given. The calculation was carried out for two values of the temperature T and the flow velocity v_0 . From the table and Fig. 2, it is seen that at small r only carbon atoms can fit the experimental values of $\varepsilon(r)$ and, thus, only for certain values of T and v_0 ; at large r , on the contrary, only for associates of $\varepsilon < 1$. Qualitatively, these relationships appear quite reasonable. Note that in one of the studies [10], it has been shown that fullerenes are produced in the region $10 < r < 30$ mm. However, in order to arrive at quantitative conclusions, it is necessary to make appropriate calculations.

ACKNOWLEDGMENTS

The authors thank A.A. Bogdanov for discussions.

This study was carried out within the program "Fullerenes and Atomic Clusters" (complex project no. 3) and supported in part by the Russian Foundation for Basic Research (project no. 00-02-16928).

REFERENCES

1. D. H. Parker, P. Wurz, K. Chatterjee, *et al.*, *J. Am. Chem. Soc.* **347**, 354 (1991).
2. D. V. Afanas'ev, I. L. Blinov, A. A. Bogdanova, *et al.*, *Zh. Tekh. Fiz.* **64** (10), 76 (1994) [*Tech. Phys.* **39**, 1017 (1994)].
3. V. P. Bubnov, I. S. Krainskiĭ, E. É. Laukhina, and É. Ya. Yagubskiĭ, *Izv. Akad. Nauk, Ser. Khim.*, No. 5, 805 (1994).
4. A. Huczko, H. Lange, P. Byszewski, *et al.*, *J. Phys. Chem.* **101**, 1267 (1997).
5. T. Belz, J. Find, D. Herein, *et al.*, *Ber. Bunsenges. Phys. Chem.* **101**, 712 (1997).
6. S. Ramakrishnan, A. Stokes, and J. Lowke, *J. Phys. D* **11**, 2267 (1978).
7. O. A. Nerushev and G. I. Sukhinin, *Zh. Tekh. Fiz.* **67** (2), 41 (1997) [*Tech. Phys.* **42**, 160 (1997)].
8. O. P. Gorelik, G. A. Dyuzhev, D. V. Novikov, *et al.*, *Zh. Tekh. Fiz.* **70** (11), 118 (2000) [*Tech. Phys.* **45**, 1489 (2000)].
9. G. L. Natanson, *Dokl. Akad. Nauk SSSR* **112**, 100 (1957).
10. G. A. Dyuzhev and V. I. Karataev, *Fiz. Tverd. Tela (St. Petersburg)* **36**, 2795 (1994) [*Phys. Solid State* **36**, 1528 (1994)].

Translated by B. Kalinin

Table

Particle	D_p , cm	$\varepsilon(r)$				
		T , K	10^3		10^4	
			v_0 , cm/s	10^3	10^4	10^3
Atom	3×10^{-8}		1.3	0.41	7.3	2.31
Fullerene	7×10^{-8}		0.72	0.23	4.1	1.30
Associate	3×10^{-6}		0.02	0.006	0.11	0.035

BRIEF COMMUNICATIONS

Estimation of the Longitudinal Electric Field Produced by a Relativistic Electron Beam in a Plasma

V. B. Vladyko

Russian Research Centre Kurchatov Institute, pl. Kurchatova 1, Moscow, Russia

Received October 26, 1999; in final form, July 28, 2000

Abstract—An axially symmetric model is used to analytically estimate the longitudinal electric field created in a plasma by an injected electron beam. The relationship can be used to estimate the electron beam energy loss in the range of parameters where theoretical studies rely only on numerical simulations. The results are of interest for creating relativistic electron beams, studying their relaxation in plasma, and developing new methods for charged particle acceleration. © 2001 MAIK “Nauka/Interperiodica”.

Systems consisting of a great number of particles are very difficult to study analytically. Therefore, theoretical investigations of the electron beam–plasma interaction rely primarily on numerical simulations, which can allow for a variety of processes evolving simultaneously. Such an approach is capable of accurately predicting the behavior of a complex system with given parameters. However, along with simulating particular facilities numerically, one should also analyze the physical processes (first of all, the dominant processes that govern the system behavior). The study of these processes with approximate models that ignore fine effects even gives a number of useful relationships. In this paper, we analyze the results of numerical simulations reported in [1–4] and use them to develop an analytical model and to obtain an equation for the longitudinal electric field produced by a relativistic electron beam (REB) injected into a plasma. This result is very important in applications, because the beam energy is transferred to the plasma mainly through the excitation of the longitudinal electric field.

We consider the REB–plasma interaction assuming that the problem is axially symmetric. The set of Maxwell’s equations in cylindrical coordinates (r, φ, z) under the condition $\partial/\partial\varphi = 0$ has the form

$$\begin{aligned} -\frac{\partial B_\varphi}{\partial z} &= \frac{1}{c} \frac{\partial E_r}{\partial t} + \frac{4\pi}{c} j_{pr}, \\ \frac{1}{r} \frac{\partial}{\partial r} r B_\varphi &= \frac{1}{c} \frac{\partial E_z}{\partial t} + \frac{4\pi}{c} (j_{pz} + j_b), \\ \frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} &= -\frac{1}{c} \frac{\partial B_\varphi}{\partial t}, \end{aligned} \quad (1)$$

where B_φ is the azimuthal magnetic field component, E_r and E_z are the radial and longitudinal electric field components, j_{pr} and j_{pz} are the radial and longitudinal components of the plasma electron current density, and j_b is the beam current density.

In Eqs. (1), the radial component of the beam current is neglected. Furthermore, to a first approximation, one can neglect the effect of electromagnetic fields on the beam electrons and assume that the beam is given and that the induced fields depend on the variable $\zeta = t - (z/V_b)$ (where V_b is the longitudinal velocity of the beam, z is the longitudinal coordinate, and t is the time), rather than on the two independent variables z and t . The above assumptions significantly simplify the problem and simultaneously allow us to study the physical processes that govern the behavior of the system under study. In this case, Eqs. (1) yield

$$\frac{\partial E_z}{\partial r} = -\frac{1}{\gamma_b^2 \beta_b^2 c} \frac{\partial B_\varphi}{\partial \zeta} + \frac{4\pi}{\beta_b c} j_{pr}, \quad (2)$$

where $\gamma_b = 1/\sqrt{1 - V_b^2/c^2}$.

Since all the above quantities depend on ζ rather than on z and t independently, the electromagnetic field and charges are carried with the beam. If the beam current grows slowly enough (so that the displacement currents can be neglected) and the longitudinal velocity of the plasma electrons is small (the longitudinal component of the plasma electron current density may be ignored), the continuity equation $\partial\rho/\partial t = \text{div } \mathbf{j} = 0$ takes the form

$$\frac{\partial \rho}{\partial \zeta} + \frac{1}{r} \frac{\partial}{\partial r} r j_{pr} - \frac{1}{4\pi\beta_b} \frac{1}{r} \frac{\partial}{\partial r} r \frac{\partial}{\partial \zeta} B_\varphi = 0. \quad (3)$$

Integrating Eq. (3) with respect to r , we obtain

$$\frac{\partial B_\varphi}{\partial \zeta} = \frac{2}{r} \beta_b \frac{\partial Q}{\partial \zeta} + 4\pi\beta_b j_{pr}, \quad (4)$$

where $Q = \int_0^r 2\pi r \rho dr$ is the charge contained in a straight unit-length cylinder of radius r aligned with the beam axis.

Substituting Eq. (4) into Eq. (2), we obtain

$$\frac{\partial E_z}{\partial r} = -\frac{1}{\gamma_b^2 \beta_b c} \frac{2\partial Q}{r \partial \zeta} + \frac{4\pi\beta_b}{c} j_{pr}. \quad (5)$$

The first term on the right-hand side of Eq. (5) is due to charge transport in the longitudinal direction. The denominator of this term contains γ_b^2 , because the vortex and Coulomb fields associated with the charge transport are directed oppositely and differ by a factor of β_b^2 . Even at a moderate relativistic factor of $\gamma_b > 3$, this term can be neglected if the linear density of the plasma channel satisfies the inequality $N_p > N_b \gamma_b^{-2}$. This condition is fulfilled with a margin in the majority of applications. In this case, Eq. (5) reduces to a simpler equation for the longitudinal electric field:

$$\frac{\partial E_z}{\partial r} = \frac{4\pi\beta_b}{c} j_{pr}. \quad (6)$$

Let us consider corollaries of Eq. (6) in more detail. This equation may be used to calculate the longitudinal electric field from the known radial dynamics of the plasma electrons. For example, when even a moderately relativistic beam ($\gamma_b > 3$) is transported through a plasma channel of radius b in which the plasma density is much higher than the beam density, the boundary condition for the longitudinal electric field component can be taken as

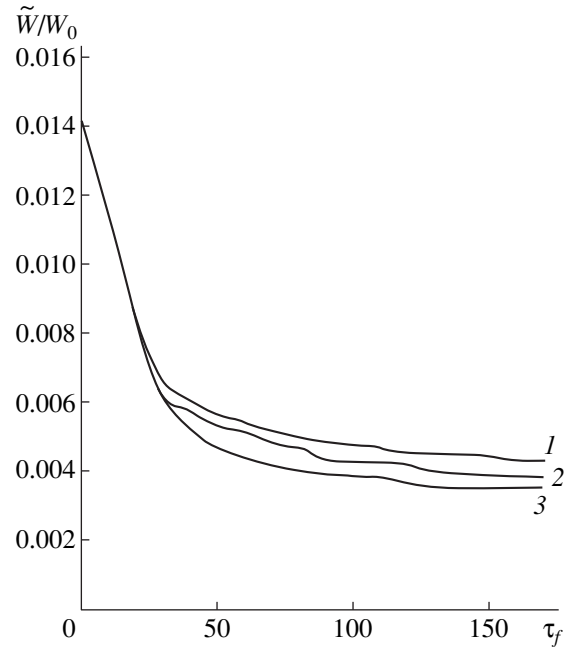
$$E_z(b) = 0. \quad (7)$$

This is an evident corollary of Eq. (6), because radial currents are only possible inside the channel. Boundary condition (7) is valid for a wide range of parameters and can be used to analyze the beam transport through dense plasma channels.

Let us use Eq. (6) to find the electron beam energy loss due to the displacement of plasma electrons. We consider the injection of an electron beam with the linear particle density N_b higher than the linear electron density N_p of the plasma channel. Therefore, it can be assumed that, after a sufficiently long period of time T , all plasma electrons will leave the channel and deposit on the wall of the drift tube with a radius R . Under these assumptions, the beam energy loss W_b associated with the displacement of electrons over the distance Δz can be estimated as

$$W_b = \Delta z \int_0^T I_b E_z(0) dt, \quad (8)$$

where I_b is the beam current and $E_z(0)$ is the field at the axis of the system.



Beam energy loss \tilde{W} as a function of τ_f for $N_b/N_p = 2$, $a = b = 0.5R$; $\tilde{W}_0 = mc^2 mc^2 / 16e^2 \Delta z$; and $\omega_p R/c = (1) 0.5$, (2) 1, and (3) 1.5.

The longitudinal electric field at the axis can be found from Eq. (6) as

$$E_z(0) = \int_R^0 \frac{4\pi\beta_b}{c} j_{pr} dr. \quad (9)$$

Assuming that the plasma channel is uniform, we obtain

$$\int_0^r j_{pr} dt = \begin{cases} \frac{N_p e}{2\pi r}, & r > b \\ \frac{N_p e r^2}{2\pi r b^2}, & r \leq b, \end{cases} \quad (10)$$

where e is the electron charge.

With (9) and (10), Eq. (8) yields

$$W_b = \frac{I_b N_p e \beta_b}{c} \Delta z (1 + 2 \ln R/b). \quad (11)$$

Relationship (11) completely agrees with the results of numerical calculations, as well as with the analytical results obtained from other models in [2–6]. It should be noted that this energy loss refers to unmagnetized plasma. In a magnetized plasma, $j_{pr} = 0$ and this loss is absent.

An important corollary of Eq. (6) is that the longitudinal electric field is governed by the radial component of the plasma electron current density. When the beam front is wide, the energy loss is significantly lower. The

validity of this assumption was verified by a numerical model based on Maxwell's equations and Vlasov's equation.

The figure shows the energy loss

$$\tilde{W} = W_b \frac{c^4 N_p}{\omega_p^4 R^4 N_b}$$

versus the dimensionless duration of the beam front $\tau_f = t_f \omega_p$ for different values of $\omega_p R/c$. The beam radius is $a = b$. It is seen that \tilde{W} is almost independent of $\omega_p R/c$, but it depends strongly on τ_f . For small $\tau_f (< 20)$, the beam energy loss is mainly determined by the kinetic energy gained by the plasma electrons in the electromagnetic field induced by the beam front. As the front duration increases, the electromagnetic field in the system decreases and the plasma electrons gain a lower kinetic energy. With these parameters, the energy loss \tilde{W} rapidly decreases with increasing τ_f . However, starting with $\tau_f > 100$, the energy loss becomes almost independent of τ_f , because the duration of the beam front becomes much longer than the oscillation period of the plasma electrons. Even though the front duration is large, the electron beam excites radial oscillations of the plasma electrons. It is the amplitude of these oscillations

(which weakly depends on the pulse front duration) that governs the major part of the beam energy loss in this case. Thus, we have shown that the energy loss can be made several times lower by increasing the duration of the pulse front. If the propagation distance is fairly long, the energy savings may be significant. It should be noted that at $\tau_f = 0$, the numerical results agree well with formula (11).

REFERENCES

1. T. W. L. Sanford, D. R. Welch, and R. C. Mock, *Phys. Plasmas* **1**, 404 (1994).
2. J. Krall, K. Nguyen, and G. Joyce, *Phys. Fluids B* **1**, 2099 (1989).
3. A. B. Filimonov, M. G. Nikulin, and S. V. Vinogradov, in *Proceedings of the XIV International Symposium on Discharges and Electrical Insulation in Vacuum, Santa-Fe, 1990*, p. 742.
4. V. B. Vladko and Yu. V. Rudyak, *Fiz. Plazmy* **20**, 973 (1994) [*Plasma Phys. Rep.* **20**, 872 (1994)].
5. H. L. Buchanan, *Phys. Fluids* **30**, 221 (1987).
6. H. S. Uhm and G. Joyce, *Phys. Fluids B* **3**, 1587 (1991).

Translated by A. Khzmalyan

ERRATA

**Errata: “On One Mechanism of Magnetization Reversal
in Crystals with Combined Anisotropy”
[*Technical Physics* 45 (11), 1441 (2000)]**

R. M. Vakhitov and V. E. Kucherov

e-mail: VakhitovRM@bsu.bashedu.ru

In formula (8) (p. 1443), ζ should be replaced by ξ .

Throughout the paper, “null domain wall” should read “zero-degree domain wall.”