Magnetic properties of electrons in metal–ammonia solutions

V. K. Mukhomorov

Agrophysical Scientific-Research Institute, 195220 St. Petersburg, Russia

A theoretical approach is proposed to describe the concentration and temperature dependence of the static magnetic susceptibility of a metal–ammonia solution. A mechanism is also suggested to explain the transition of the system from the paramagnetic to the diamagnetic state. The theory is based on the assumption that both single-particle and two-electron bound singlet formations of the bipolar type exist in the solution. It is shown that diamagnetism is due to the electron orbital motion and to the relative motion of the quasiparticles. © *1997 American Institute of Physics.* [S1063-7842(97)00108-6]

Studies of the concentration dependence of the magnetic susceptibility of metal–ammonia solutions have shown[1,2] that the static magnetic susceptibility of the systems varies substantially depending on the concentration of dissolved alkali metal, ranging from the paramagnetic, purely spin susceptibility typical of the electronic subsystem of noninteracting spins at low concentrations to a diamagnetic state at concentrations of the order of 10^{20} cm^{-3} and temperatures below 200 K. Further experimental studies[3] of the spin susceptibility of metal–ammonia solutions confirmed that as the electron concentration increases, a compensation of the spin angular momenta of the electrons does in fact occur. Numerous attempts have been made to explain this behavior of the static susceptibility of metal–ammonia systems. In Ref. 4, the experimentally observed change in susceptibility was analyzed assuming that solvated electrons and also bound electron states in the form of F color centers and double F_2 color centers may coexist in the solution. However, these bound states were not observed experimentally.[5] The binding energy of these formations[4] is considerably higher than the characteristic energy of a solvated electron, which should obviously lead to an appreciable short-wavelength shift of the optical absorption band maximum. In practice, however, the maximum is slightly shifted toward longer wavelengths.[6,7] An analysis of the absorption frequencies, performed in Ref. 8, showed that a similar shift of the optical absorption maximum may be attributed to the formation of bipolar two-electron bound singlet formations. It was noted in Refs. 6 and 9–16 that a singlet paired state of two electrons may be formed in ammonia. However, bound states of solvated electrons with metal cations do not form, as was demonstrated experimentally.[17,18] The valence electrons of dissociated metal atoms and metal cations are clearly spatially separated in the solution. This conclusion also agrees with the results of investigations of the optical properties of solvated electrons in ammonia, which proved to be independent of the nature of the dissolved alkaline-earth metals, including divalent ones.[19] The absence of any electron interactions with positive centers in the solution is also evidenced by the extremely narrow spin resonance line with a g factor of 2.0012 ∓ 0.0002 (Refs. 3 and 20), which is almost the same as the free-electron g factor of 2.0023.

In the present paper the concentration dependence and temperature dependence of the static magnetic susceptibility of metal–ammonia solutions are analyzed on the assumption that solvated electrons and bipolar two-electron bound singlet formations coexist in the solution. A rigorous, translationally invariant theory of continuous bipolarons was constructed in Ref. 21 using the adiabatic approximation. The conditions for the existence of spin-paired electron states (bipolarons) in polar dielectric media were discussed in detail in Refs. 22 and 23, where the authors determined typical parameters of bipolarons, criteria for stability of a two-electron bound state, and the dependence of the binding energy of the two-electron formation on the dielectric properties of the polar medium, including for the case of ammonia. The interelectron pair potential was also determined as a function of the distance between self-trapped electrons.

The states of solvated (self-trapped) electrons are described using a continuous model, which presupposes strong interaction between the electrons and the longitudinal branch of polarization oscillations of the medium. Various investigations[4,24–27] have shown that many properties of electrons solvated in ammonia may be described using the model of continuous polarons. The criteria for validity of the theory reduce to the following inequality: $\hbar\omega_f < \hbar\omega_e < \hbar\omega_m$. For an electron solvated in ammonia, $\hbar\omega_e = 0.885$ eV is the energy of the most active optical transition of a self-trapped electron,[5] $\hbar\omega_m \simeq 6$ eV is the excitation energy of electrons of the main substance, and $\hbar\omega_f \simeq 0.4$ eV is the energy of the longitudinal polarization oscillations of the medium. The orientational oscillations of molecules about their equilibrium position in a polar liquid form elastic waves which may be treated as in a crystal. As a result of the directionality and saturation of the intermolecular hydrogen bonds for ammonia, the ''quasicrystallinity'' of the structure is comparatively well defined. Far from the critical point, the thermal vibrations of the molecules may be reduced to a set of Debye waves, as in a crystal, where the spectrum of collective oscillations in the liquid has a cutoff at longer wavelengths than in crystals[28] on account of the translational motion of the particles. The elastic continuum approximation does not generally allow for anistropy and is far better applicable to a liquid than to a crystal.[29]

We first derive dynamical equations to describe the behavior of self-trapped (solvated) electrons in a static mag-

netic field with allowance for the interaction between the electrons and the longitudinal polarization oscillations of a polar medium. The author has previously studied the equations of motion for noninteracting electrons in a magnetic field in a polarizable medium[30] and has established that this motion is of a complex character, consisting of electron oscillations in a deep potential well together with oscillations of the center of inertia of the entire system in a magnetic field. Thus in the present study, particular attention will be devoted to analyzing the equations of motion for a bound two-electron system in a magnetic field. The influence of an external uniform magnetic field on self-trapped single-electron states in polar media was discussed in detail in Ref. 31, and its influence on bipolaron states was considered using a Feynmann path-integral formalism in Ref. 32. Here the effect of an external uniform magnetic field on two-electron bound states will be investigated by canonical transformations of Bogolyubov–Tyablikov coordinates.[33,34]

The complete Hamiltonian describing the state of two interacting electrons in a uniform, isotropic dielectric continuum in the presence of an external uniform magnetic field specified by the vector potential $\mathbf{A}(\mathbf{r})$ is written as follows:

$$
H = \frac{1}{m^*} \sum_{j=1,2} \left(\frac{\hbar}{i} \nabla_j - \frac{e}{c} \mathbf{A}(\mathbf{r}_j) \right)^2
$$
$$
+ \sum_{j=1,2} \sum_f [V_f \exp(i\mathbf{f}\cdot\mathbf{r}_j) b_f + V^*_{-f} \exp(-i\mathbf{f}\cdot\mathbf{r}_j) b_f^+]
$$
$$
+ \sum_f \hbar\omega_f(b_f^+ b_f + b_f b_f^+) + \frac{e^2}{\varepsilon_\infty |\mathbf{r}_1 - \mathbf{r}_2|}. \tag{1}
$$

The Fourier coefficients $V_f = (i/f)\hbar\omega_f(4\pi\alpha_c/V)^{1/2} \times (\hbar/2m^*\omega_f)^{1/4}$ satisfy the reality condition $V_f = V^*_{-f}$; the dimensionless electron–phonon coupling constant $\alpha_c = (e^2/2\varepsilon^*\hbar\omega_f)(2m^*\omega_f/\hbar)^{1/2}$ characterizes the ratio of the effective electron energy $\mathrm{Ry}^* = e^4 m^*/\varepsilon^{*2}\hbar^2$ to the energy quantum $\hbar\omega_f$ of a longitudinal polarization oscillation with quasimomentum $\hbar\mathbf{f}$ involving librational oscillations of the dipolar ammonia molecules; the system is situated in a certain finite volume V and obeys periodic boundary conditions; $\varepsilon^* = \varepsilon_\infty \varepsilon_s/(\varepsilon_s - \varepsilon_\infty)$ is the effective permittivity of the medium and corresponds to a polarization wherein the dipolar molecules follow the field of the electrons with an inertial delay, while their electron shells adiabatically follow the configuration of the dipolar environment. The following theoretical parameters were taken for electrons solvated in ammonia: static permittivity $\varepsilon_s = 22.8$, high-frequency permittivity $\varepsilon_\infty = 1.756$, and isotropic effective electron mass at the bottom of conduction band $m^* = 1.73m$ (Ref. 23), which was determined by comparing the theoretical and experimental positions of the absorption band maximum of a solvated electron in ammonia. The quantum amplitudes b_f and b_f^+ of the phonon field satisfy the commutation relations for the discrete spectrum $[b_f b_{f'}^+]_- = \delta_{ff'}$, $[b_f^+ b_{f'}^+]_- = [b_f b_{f'}]_- = 0$. The vector potential of the external magnetic field is given by $\mathbf{A}(\mathbf{r}_j) = (-Hy_j/2, Hx_j/2, 0)$ and satisfies the Coulomb gauge, i.e., $\mathrm{div}\mathbf{A} = 0$, where $\mathbf{r}_j$ is the position vector of the jth electron. An external static uniform magnetic field $\mathbf{H}$ is directed along the z axis. It is postulated that the magnetic field is a weak perturbation and that the effective electron mass, refractive index, and static permittivity of the polar medium do not depend on this field. Zero energy is taken as the energy of the system when all the alkali metal atoms are dissociated and the electrons are located at the bottom of the conduction band, infinitely separated, and do not interact with phonons.

Here the action of the magnetic field on two-electron bound states will be analyzed using adiabatic perturbation theory. We formally introduce the small parameter ξ, assuming $\omega_f = \xi\nu_f$. Using the small parameter ξ allows us to investigate the Hamiltonian (1) by the method of canonical transformations of the Bogolyubov–Tyablikov coordinates,[33,34] which separates the internal, translationally invariant degrees of freedom from the motion of the system as a whole. This then opens up the possibility of constructing a self-consistent system of successive approximations to the energy and wave function of a composite system, while the individuality of its interacting components is still conserved.

In the Hamiltonian (1) we convert from the occupation numbers b_f and b_f^+ to the complex variables of the field coordinates q_f and their canonically conjugate momenta p_f:

$$
b_f^+ = (q_{-f}/\xi - i\xi p_f)/\sqrt{2},
$$
$$
b_f = (q_f/\xi + i\xi p_{-f})/\sqrt{2}. \tag{2}
$$

The variables q_f and p_f satisfy the commutation relation $[q_f, p_{f'}]_- = i\delta_{ff'}$. Using the transformation (2), the Hamiltonian (1) can be reduced to:

$$
H = \frac{1}{m^*} \sum_{j=1,2} \left(\frac{\hbar}{i} \nabla_j - \frac{e}{c} \mathbf{A}(\mathbf{r}_j) \right)^2
$$
$$
+ \sum_{j=1,2} \sum_f [W_f \exp(i\mathbf{f}\cdot\mathbf{r}_j) q_f + W^*_{-f} \exp(-i\mathbf{f}\cdot\mathbf{r}_j) q^*_{-f}]
$$
$$
+ \sum_f \hbar\nu_f(q_f^+ q_{-f} + \xi^4 p_f p_{-f}^+) + \frac{e^2}{\varepsilon_\infty |\mathbf{r}_1 - \mathbf{r}_2|},
$$
$$
W_f = V_f/\xi\sqrt{2}. \tag{3}
$$

Since the term proportional to the kinetic energy of the phonon field is small, to a first approximation the Heisenberg equations of motion have the form $q_f(t) = $ const and the principal effect of the particle–field interaction can be reduced to the creation of a deep potential well for each particle, whose motion is kinematically independent for large interparticle distances. The position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ for each particle can then by given as:

$$
\mathbf{r}_1 = \mathbf{R}_1 + \boldsymbol{\rho}_1, \quad \mathbf{r}_2 = \mathbf{R}_2 + \boldsymbol{\rho}_2. \tag{4}
$$

The vectors $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ are translationally invariant and describe the high-frequency oscillations of the electrons in the polarization potential wells, and the vectors $\mathbf{R}_1$ and $\mathbf{R}_2$ have the meaning of the coordinates of the center of mass of the first and second electrons.

It is known that as a result of electron–phonon interactions, the equilibrium positions of the field oscillators are

shifted and their amplitude changes. The new phonon coordinates Q_f can then be related to the old coordinates q_f by:

$$q_f \exp(i\mathbf{f}\cdot\mathbf{R}_1) = u_f(1) + \xi Q_f,$$

$$q_f \exp(i\mathbf{f}\cdot\mathbf{R}_2) = u_f(2) + \xi Q_f, \tag{5}$$

where $u_f(1)$ and $u_f(2)$ are the two kinematically independent classical components of the phonon field for the first and second quasiparticles.

The variable Q_f describes the quantum fluctuations of the field about their classical values. Using Eq. (5), we determine the old

$$q_f = [(u_f(1) + \xi Q_f)\exp(-i\mathbf{f}\cdot\mathbf{R}_1)$$
$$+ (u_f(2) + \xi Q_f)\exp(-i\mathbf{f}\cdot\mathbf{R}_2)]/2 \tag{6}$$

and new

$$Q_f = \{q_f[\exp(i\mathbf{f}\cdot\mathbf{R}_1) + \exp(i\mathbf{f}\cdot\mathbf{R}_2)] - u_f(1) - u_f(2)\}/2\xi \tag{7}$$

amplitudes of the phonon field for a two-particle system. In the absence of external fields, the Hamiltonian (3) is translationally invariant with respect to the transformations (4)–(7).

It is convenient to change from the variables $\mathbf{R}_1$ and $\mathbf{R}_2$ to the coordinates of the center of inertia of the system and the relative motion

$$\mathbf{r}_1 = \mathbf{R} + b\boldsymbol{\rho} + \boldsymbol{\rho}_1, \quad \mathbf{r}_2 = \mathbf{R} - a\boldsymbol{\rho} + \boldsymbol{\rho}_2,$$

$$a = M_1/(M_1 + M_2), \quad b = M_2/(M_1 + M_2), \tag{8}$$

where M_1 and M_2 are the effective translational masses of the first and second self-trapped electrons, separated by infinity, $\mathbf{R}$ is the position vector of the center of inertia of the two-electron system, and $\boldsymbol{\rho}$ is the coordinate of the relative motion of the quasiparticles.

Since the transformations (7) and (8) increase the number of independent variables compared with the initial variables $\mathbf{r}_1$, $\mathbf{r}_2$, and q_f by six, the same number of additional conditions must be introduced. The additional conditions are given in the following simple form:

$$\sum_f f_\alpha v_f^*(1)\{q_f \exp[i\mathbf{f}\cdot(\mathbf{R}+b\boldsymbol{\rho})] - u_f(1)\} = 0,$$

$$\sum_f f_\alpha v_f^*(2)\{q_f \exp[i\mathbf{f}\cdot(\mathbf{R}-a\boldsymbol{\rho})] - u_f(2)\} = 0,$$

$$\alpha = x, y, z. \tag{9}$$

The new variables satisfy the reality condition $u_f^*(i) = u_{-f}(i)$, $v_f^*(i) = v_{-f}(i)$, $Q_f^+ = Q_{-f}$, where $i = 1,2$. Without restricting the generality, it may be assumed that the complex numbers $v_f(i)$ and $u_f(i)$ also satisfy the following orthogonality relations:

$$\sum_f f_\alpha f_\beta v_f^*(1)[u_f(1) + u_f(2)\exp(i\mathbf{f}\cdot\boldsymbol{\rho})] = \delta_{\alpha\beta},$$

$$\sum_f f_\alpha f_\beta v_f^*(2)[u_f(2) + u_f(1)\exp(-i\mathbf{f}\cdot\boldsymbol{\rho})] = \delta_{\alpha\beta},$$

$$\alpha, \beta = x, y, z. \tag{10}$$

In the Hamiltonian (3) we change from the old to the new variables (7) and (8), for which we determine the operator

$$\frac{\partial}{\partial q_k} = \sum_f \frac{\partial Q_f}{\partial q_k}\frac{\partial}{\partial Q_f} + \frac{\partial \mathbf{R}}{\partial q_k}\frac{\partial}{\partial \mathbf{R}}$$
$$+ \frac{\partial \boldsymbol{\rho}}{\partial q_k}\frac{\partial}{\partial \boldsymbol{\rho}} + \frac{\partial \boldsymbol{\rho}_1}{\partial q_k}\frac{\partial}{\partial \boldsymbol{\rho}_1} + \frac{\partial \boldsymbol{\rho}_2}{\partial q_k}\frac{\partial}{\partial \boldsymbol{\rho}_2}. \tag{11}$$

The derivatives $\partial Q_f/\partial q_k$ are easy to determine from relation (7). In order to find the derivatives $\partial \mathbf{R}/\partial q_k$ and $\partial\boldsymbol{\rho}/\partial q_k$, we differentiate the additional condition (9) with respect to the field coordinates q_k and take into account the orthogonality condition (10). We then obtain two equations to determine the unknown derivatives:

$$\mathbf{k}v_k^*(1)\exp[i\mathbf{k}\cdot(\mathbf{R}+b\boldsymbol{\rho})] + (i/2)\partial(\mathbf{R}+b\boldsymbol{\rho})/\partial q_k + (i\xi/2)$$

$$\times \sum_f (\mathbf{f}\cdot\mathbf{f})v_f^*(1)Q_f[1 + \exp(i\mathbf{f}\cdot\boldsymbol{\rho})]\partial(\mathbf{R}+b\boldsymbol{\rho})/\partial q_k = 0,$$

$$\mathbf{k}v_k^*(2)\exp[i\mathbf{k}\cdot(\mathbf{R}-a\boldsymbol{\rho})] + (i/2)\partial(\mathbf{R}-a\boldsymbol{\rho})/\partial q_k + (i\xi/2)$$

$$\times \sum_f (\mathbf{f}\cdot\mathbf{f})v_f^*(2)Q_f[1 + \exp(-i\mathbf{f}\cdot\boldsymbol{\rho})]\partial(\mathbf{R}-a\boldsymbol{\rho})/\partial q_k = 0. \tag{12}$$

The solutions of system (12) may be sought by the method of successive approximations. As a result, we obtain the following iterative solutions in the form of a series expansion in terms of the small parameter ξ:

$$\partial(\mathbf{R}+b\boldsymbol{\rho})/\partial q_k = 2i\mathbf{k}v_k^*(1)\exp[i\mathbf{k}\cdot(\mathbf{R}+b\boldsymbol{\rho})]$$
$$- 2i\xi\mathbf{k}v_k^*(1)\exp[i\mathbf{k}\cdot(\mathbf{R}+b\boldsymbol{\rho})]$$
$$\times \sum_f (\mathbf{f}\cdot\mathbf{f})v_f^*(1)Q_f[1 + \exp(i\mathbf{f}\cdot\boldsymbol{\rho})] + \ldots,$$

$$\partial(\mathbf{R}-a\boldsymbol{\rho})/\partial q_k = 2i\mathbf{k}v_k^*(2)\exp[i\mathbf{k}\cdot(\mathbf{R}+a\boldsymbol{\rho})]$$
$$- 2i\xi\mathbf{k}v_k^*(2)\exp[i\mathbf{k}\cdot(\mathbf{R}-a\boldsymbol{\rho})]$$
$$\times \sum_f (\mathbf{f}\cdot\mathbf{f})v_f^*(2)Q_f[1 + \exp(-i\mathbf{f}\cdot\boldsymbol{\rho})] + \ldots \tag{13}$$

It is then easy to find the derivatives $\partial\boldsymbol{\rho}/\partial q_k$ and $\partial\boldsymbol{\rho}/\partial q_k$. In accordance with the definition of the coordinates $\boldsymbol{\rho}$, $\boldsymbol{\rho}_1$, $\boldsymbol{\rho}_2$, and $\mathbf{R}$, we find from Eq. (8)

$$\partial\boldsymbol{\rho}_1/\partial q_k = -\partial\mathbf{R}/\partial q_k - b\partial\boldsymbol{\rho}/\partial q_k,$$

$$\partial\boldsymbol{\rho}_2/\partial q_k = -\partial\mathbf{R}/\partial q_k + a\partial\boldsymbol{\rho}/\partial q_k. \tag{14}$$

Taking into account the iterative values of the derivatives (13), we obtain the derivatives

$$\partial\boldsymbol{\rho}_1/\partial q_k = -2i\mathbf{k}\exp(i\mathbf{k}\cdot\mathbf{R})v_k^*(1)\exp(ib\mathbf{k}\cdot\boldsymbol{\rho}) + \ldots,$$

$$\partial\boldsymbol{\rho}_2/\partial q_k = -2i\mathbf{k}\exp(i\mathbf{k}\cdot\mathbf{R})v_k^*(2)\exp(-ia\mathbf{k}\cdot\boldsymbol{\rho}) + \ldots. \tag{15}$$

Here we have confined ourselves only to the leading terms of the expansion as a series in powers of the parameter ξ. Differentiating Eq. (6) with respect to q_k allowing for the transformation (13) and (15), we express the operator $\partial/\partial q_k$ as a series to terms of zeroth order in the parameter ξ

$$\partial/\partial q_k = \left\{ \exp(i\mathbf{k}\cdot\mathbf{R})[\exp(ib\mathbf{k}\cdot\boldsymbol{\rho}) + \exp(-ia\mathbf{k}\cdot\boldsymbol{\rho})]\partial/\partial Q_k \right.$$

$$+ (i/2)\sum_f \mathbf{f}[(u_f(1) + \xi Q_f)\exp(-ib\mathbf{f}\cdot\boldsymbol{\rho})$$

$$+ (u_f(2) + \xi Q_f)\exp(ia\mathbf{f}\cdot\boldsymbol{\rho})]$$

$$\times \left[2i\mathbf{k}\exp(ib\mathbf{f}\cdot\boldsymbol{\rho})v_k^*(1)\exp(i\mathbf{k}\cdot(\mathbf{R}+b\boldsymbol{\rho})) \right]$$

$$\times \left[1 - \xi\sum_s \mathbf{s}\cdot\mathbf{s}v_s^*(1)Q_s(1+\exp(i\mathbf{s}\cdot\boldsymbol{\rho})) + \dots \right]$$

$$+ 2i\mathbf{k}\exp(-ia\mathbf{f}\cdot\boldsymbol{\rho})v_k^*(2)\exp(i\mathbf{k}\cdot(\mathbf{R}-a\boldsymbol{\rho}))$$

$$\times \left[1 - \xi\sum_s \mathbf{s}\cdot\mathbf{s}v_s^*(2)Q_s(1+\exp(-i\mathbf{s}\cdot\boldsymbol{\rho})) \right.$$

$$\left. + \dots \right]\partial/\partial Q_f\Big\}/2\xi$$

$$+ 2i\mathbf{k}\exp(i\mathbf{k}\cdot\mathbf{R})[av_k^*(1)\exp(ib\mathbf{k}\cdot\boldsymbol{\rho}) + bv_k^*(2)$$

$$\times \exp(-ia\mathbf{k}\cdot\boldsymbol{\rho})]\partial/\partial\mathbf{R} + 2i\mathbf{k}\exp(i\mathbf{k}\cdot\mathbf{R})$$

$$\times [v_k^*(1)\exp(ib\mathbf{k}\cdot\boldsymbol{\rho}) - v_k^*(2)\exp(-ia\mathbf{k}\cdot\boldsymbol{\rho})]\partial/\partial\boldsymbol{\rho}$$

$$- 2i\mathbf{k}\exp(i\mathbf{k}\cdot\mathbf{R})[v_k^*(1)\exp(ib\mathbf{k}\cdot\boldsymbol{\rho})\partial/\partial\boldsymbol{\rho}_1$$

$$+ v_k^*(2)\exp(-ia\mathbf{k}\cdot\boldsymbol{\rho})\partial/\partial\boldsymbol{\rho}_2] + \dots \quad (16)$$

Bearing in mind that $\partial Q_f/\partial\mathbf{r}_1 = 0$ and $\partial Q_f = \partial\mathbf{r}_2 = 0$, we have for the operators $\nabla_{\mathbf{r}_1} = \nabla_{\boldsymbol{\rho}_1}$ and $\nabla_{\mathbf{r}_2} = \nabla_{\boldsymbol{\rho}_2}$. Assuming the Coulomb gauge for the vector potential and also using formulas (15) and (16), we can rewrite the complete Hamiltonian of the system in terms of collective coordinates as a series in increasing powers of the small parameter up to quantities of order ξ^4:

$$H = -(\hbar^2/2m^*)(\nabla_{\boldsymbol{\rho}_1}^2 + \nabla_{\boldsymbol{\rho}_2}^2) + (1/2)$$

$$\times \sum_f \left\{ W_f[[u_f(1) + u_f(2)\exp(i\mathbf{f}\cdot\boldsymbol{\rho})]\exp(i\mathbf{f}\cdot\boldsymbol{\rho}_1) \right.$$

$$+ [u_f(2) + u_f(1)\exp(-i\mathbf{f}\cdot\boldsymbol{\rho})]\exp(i\mathbf{f}\cdot\boldsymbol{\rho}_2)]$$

$$+ W_{-f}^*[[u_{-f}(1) + u_{-f}(2)\exp(-i\mathbf{f}\cdot\boldsymbol{\rho})]\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}_1)$$

$$+ [u_{-f}(2) + u_{-f}(1)\exp(i\mathbf{f}\cdot\boldsymbol{\rho})]\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}_2)]\}$$

$$+ e^2/\varepsilon_\infty|\boldsymbol{\rho} + \boldsymbol{\rho}_1 - \boldsymbol{\rho}_2| + \sum_f \hbar\nu_f\{u_f(1)[u_{-f}(1)$$

$$+ u_{-f}(2)\exp(i\mathbf{f}\cdot\boldsymbol{\rho})] + u_{-f}(2)[u_f(2) + u_f(1)$$

$$\times \exp(-i\mathbf{f}\cdot\boldsymbol{\rho})]\}/2 + i\xi(eH\hbar/2m^*c)$$

$$\times \left\{ \sum_{i=1,2} [(R_y\partial/\partial\rho_{ix} - R_x\partial/\partial\rho_{iy}) \right.$$

$$+ (\rho_{iy}\partial/\partial\rho_{ix} - \rho_{ix}\partial/\partial\rho_{iy})] + \rho_y(b\partial/\partial\rho_{1x} - a\partial/\partial\rho_{2x})$$

$$\left. - \rho_x(b\partial/\partial\rho_{1y} - a\partial/\partial\rho_{2y}) \right\} + \xi^2(e^2H^2/8m^*c^2)$$

$$\times \{(R_x^2 + R_y^2) + 2(bR_x\rho_x - aR_y\rho_y) + b^2\rho_x^2 + a^2\rho_y^2\}$$

$$+ \sum_{i=1,2} [(\rho_{ix}^2 + \rho_{iy}^2) + 2(R_x\rho_{ix} + R_y\rho_{iy})$$

$$+ 2(b\rho_x\rho_{ix} - a\rho_y\rho_{iy})] + (\xi/2)\sum_f \{Q_f W_f[(1$$

$$+ \exp(i\mathbf{f}\cdot\boldsymbol{\rho}))\exp(i\mathbf{f}\cdot\boldsymbol{\rho}_1) + (1 + \exp(-i\mathbf{f}\cdot\boldsymbol{\rho}))\exp(i\mathbf{f}\cdot\boldsymbol{\rho}_2)]$$

$$+ Q_{-f}W_{-f}^*[(1 + \exp(-i\mathbf{f}\cdot\boldsymbol{\rho}))\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}_1)$$

$$+ (1 + \exp(i\mathbf{f}\cdot\boldsymbol{\rho}))\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}_2)] + \hbar\nu_f Q_f[u_{-f}(1)$$

$$\times (1 + \exp(i\mathbf{f}\cdot\boldsymbol{\rho})) + u_f(2)(1 + \exp(-i\mathbf{f}\cdot\boldsymbol{\rho}))]$$

$$+ \hbar\nu_f Q_{-f}[u_f(1)(1 + \exp(-i\mathbf{f}\cdot\boldsymbol{\rho})) + u_{-f}(2)$$

$$\times (1 + \exp(i\mathbf{f}\cdot\boldsymbol{\rho}))]\} + (\xi^2/2)\left\{\sum_f \hbar\nu_f Q_f Q_{-f} \right.$$

$$\times [2 + \exp(i\mathbf{f}\cdot\boldsymbol{\rho}) + \exp(-i\mathbf{f}\cdot\boldsymbol{\rho})] + [P_f' + (2i\mathbf{f}/\hbar)\cdot\mathbf{F}_f]$$

$$\left. \times [P_{-f}' - (2i\mathbf{f}/\hbar)\cdot\mathbf{F}_{-f}] \right\}/4 + (i\xi^3/2)$$

$$\times \sum_f \hbar\nu_f\mathbf{f}\cdot[(P_f' + (2i\mathbf{f}/\hbar)\cdot\mathbf{F}_f)\mathbf{G}_{-f} - \mathbf{G}_f(P_{-f}'$$

$$- (2i\mathbf{f}/\hbar)\cdot\mathbf{F}_{-f})] + 2\xi^4\sum_f (\mathbf{f}\cdot\mathbf{f}/\hbar^2)\mathbf{G}_f\cdot\mathbf{G}_{-f} + \dots$$

$$= H_0 + \xi H_1 + \xi^2 H_2 + \xi^3 H_3 + \xi^4 H_4 + \dots, \quad (17)$$

where $H_0 = H_0^{(0)} + \xi H_0^{(1)} + \xi^2 H_0^{(2)}$.

Here we have introduced the additional small parameter ξ, whereby the effect of the external magnetic field can be taken into account as a weak perturbation. For the magnetic fields usually used experimentally, the magnetic energy is of the order of 10^{-5} eV and is much lower than the binding energy of a solvated electron, which is 0.16 Ry*. The following notation was adopted in the expansion (17):

$$\mathbf{F}_f = S_f^*\mathbf{I} + Z_f^*\mathbf{J}, \quad \mathbf{G}_f = 2\xi(Y_f^*\mathbf{P}_{\boldsymbol{\rho}_1} + X_f^*\mathbf{P}_{\boldsymbol{\rho}_2}),$$

$$\mathbf{I} = 2\xi\mathbf{P}_\mathbf{R}, \quad \mathbf{J} = 2\xi\mathbf{P}_{\boldsymbol{\rho}},$$

$$\mathbf{P}_\mathbf{R} = -i\hbar\partial/\partial\mathbf{R}, \quad \mathbf{P}_{\boldsymbol{\rho}} = -i\hbar\partial/\partial\boldsymbol{\rho},$$

$$\mathbf{P}_{\boldsymbol{\rho}_1} = -i\hbar\partial/\partial\boldsymbol{\rho}_1, \quad \mathbf{P}_{\boldsymbol{\rho}_2} = -i\hbar\partial/\partial\boldsymbol{\rho}_2,$$

$$P_f = -i\hbar\partial/\partial Q_f,$$

$$S_f^* = av_f^*(1)\exp(-ib\mathbf{f}\cdot\boldsymbol{\rho}) + bv_f^*(2)\exp(ia\mathbf{f}\cdot\boldsymbol{\rho}),$$

$$Y_f^* = v_f^*(1)\exp(-ib\mathbf{f}\cdot\boldsymbol{\rho}),$$

$$Z_f^* = v_f^*(1)\exp(-ib\mathbf{f}\cdot\boldsymbol{\rho}) - v_f^*(2)\exp(ia\mathbf{f}\cdot\boldsymbol{\rho}),$$

$$X_f^* = v_f^*(2)\exp(ia\mathbf{f}\cdot\boldsymbol{\rho}),$$

$$P_f' = \exp(ib\mathbf{f}\cdot\boldsymbol{\rho})\left\{ P_f - \mathbf{f}v_f^*(1)\sum_k \mathbf{k}[u_k(1) \right.$$

$$\left. + u_k(2)\exp(i\mathbf{k}\cdot\boldsymbol{\rho})]P_k \right\} + \exp(-ia\mathbf{f}\cdot\boldsymbol{\rho})\left\{ P_f - \mathbf{f}v_f^*(2) \right.$$

$$\times \sum_k \mathbf{k}[u_k(2)+u_k(1)\exp(-i\mathbf{k}\cdot\boldsymbol{\rho})]P_k\Big\} + \dots$$

$$= \exp(ib\mathbf{f}\cdot\boldsymbol{\rho})P_f'(1)+\exp(-ia\mathbf{f}\cdot\boldsymbol{\rho})P_f'(2)+\dots \quad (18)$$

The wave equation with the Hamiltonian (17) will be solved by the perturbation theory method for which we write the complete wave function and the energy as a series expansion in terms of the small parameter ξ

$$\Psi = \Psi_0 + \xi\Psi_1 + \xi^2\Psi_2 + \dots ,$$

$$E = E_0 + \xi E_1 + \xi^2 E_2 + \dots . \quad (19)$$

We substitute the series (19) into the Schrödinger equation with the Hamiltonian (17) and collect terms with the same powers of ξ. We then obtain the following chain of coupled equations:

$$(H_0-E_0)\Psi_0=0,(H_1-E_1)\Psi_0+(H_0-E_0)\Psi_1=0,$$

$$(H_2-E_2)\Psi_0+(H_1-E_1)\Psi_1+(H_0-E_0)\Psi_2=0 \dots . \quad (20)$$

Since the operator H_0 does not depend on the field variables Q_f and the vectors $\boldsymbol{\rho}$ and $\mathbf{R}$ appear in parametric form, the zeroth-order approximation function, Ψ_0, can thus be expressed as a product: $\Psi_0(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\boldsymbol{\rho},\mathbf{R},Q_f)$ $=\psi(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\boldsymbol{\rho},\mathbf{R})\Phi(Q_f)$. The wave function $\psi(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\boldsymbol{\rho},\mathbf{R})$ can then be expanded as a series in terms of the small parameter ξ. The leading term of the expansion (17) which contains nontrivial information on the system is the Hamiltonian $H_0^{(0)}$. In the zeroth-order approximation with respect to the magnetic field, the variables $\boldsymbol{\rho}_1$, $\boldsymbol{\rho}_2$, $\mathbf{R}$, and $\boldsymbol{\rho}$ in the Hamiltonian $H_0^{(0)}$ separate. Thus, the wave function ψ in the zeroth approximation can be expressed as the product $\psi_0=\chi_0(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2,\boldsymbol{\rho})\eta_0(\boldsymbol{\rho})\varphi_0(\mathbf{R})$, where the function χ_0 depends parametrically on the distance ρ. We can then find the unknown complex numbers $u_f(i)$ which determine the classical components of the polarization field. To do this, we average the Hamiltonian $H_0^{(0)}$ over the wave function χ_0 and then minimize the total energy functional $F=\langle\chi_0|H_0^{(0)}|\chi_0\rangle$ with respect to $(u_{-f}(1)+u_{-f}(2)\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}))$ and $(u_{-f}(1)+u_{-f}(2)\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}))$. We finally obtain

$$u_f(1) = -4W_{-f}^*\langle\chi_0|\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}_1)|\chi_0\rangle/\hbar\nu_f,$$

$$u_f(2) = -4W_{-f}^*\langle\chi_0|\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}_2)|\chi_0\rangle/\hbar\nu_f. \quad (21)$$

The extreme values (21) cause the linear Hamiltonian H_1 with respect to Q_f to vanish and thereby satisfy the condition for regularity of the solution in terms of the variable Q_f, as is required in the Bogolyubov–Tyablikov method.[33,34] Using formulas (21), the Hamiltonian $H_0^{(0)}$ may be rewritten as:

$$H_0^{(0)} = -(\hbar^2/2m^*)(\nabla_{\boldsymbol{\rho}_1}^2+\nabla_{\boldsymbol{\rho}_2}^2)-2\sum_f W_fW_{-f}^*$$

$$\times\{[J_f(1)+J_f(2)\exp(i\mathbf{f}\cdot\boldsymbol{\rho})]\exp(i\mathbf{f}\cdot\boldsymbol{\rho}_1)$$

$$+[J_f(2)+J_f(1)\exp(-i\mathbf{f}\cdot\boldsymbol{\rho})]\exp(i\mathbf{f}\cdot\boldsymbol{\rho}_2)$$

$$+[J_{-f}(1)+J_{-f}(2)\exp(-i\mathbf{f}\cdot\boldsymbol{\rho})]\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}_1)$$

$$+[J_{-f}(2)+J_{-f}(1)\exp(i\mathbf{f}\cdot\boldsymbol{\rho})]\exp(-i\mathbf{f}\boldsymbol{\rho}_2)\}/\hbar\nu_f$$

$$+2\sum_f W_fW_{-f}^*\{J_f(1)[J_{-f}(1)+J_{-f}(2)\exp$$

$$\times(-i\mathbf{f}\cdot\boldsymbol{\rho})]+J_f(2)[J_{-f}(2)+J_{-f}(1)$$

$$\times\exp(i\mathbf{f}\cdot\boldsymbol{\rho})]\}/\hbar\nu_f+e^2/\varepsilon_\infty|\boldsymbol{\rho}+\boldsymbol{\rho}_1-\boldsymbol{\rho}_2|, \quad (22)$$

where the following notation is used:

$$J_f(1)=\langle\chi_0|\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}_1)|\chi_0\rangle,$$

$$J_f(2)=\langle\chi_0|\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}_2)|\chi_0\rangle. \quad (23)$$

The electronic states of the two-electron formation in the zeroth approximation with respect to the magnetic field can be determined either by solving an integrodifferential equation with the operator $H_0^{(0)}$ or by using a direct variational method, i.e., by minimizing the total energy functional with an additional normalization condition imposed on the wave function χ_0. Changing from the discrete momentum representation with respect to the quasiwave vector $\mathbf{f}$ to integration, we derive the following expression for the functional of the total self-consistent electron energy in the coordinate representation:

$$E_0^{(0)}(\rho) = -(\hbar^2/2m^*)\int\chi_0(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2)(\nabla_{\boldsymbol{\rho}_1}^2+\nabla_{\boldsymbol{\rho}_2}^2)$$

$$\times\chi_0(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2)d\tau_1 d\tau_2-(e^2/2\varepsilon^*)$$

$$\times\int\chi_0^2(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2)\chi_0^2(\boldsymbol{\rho}_1',\boldsymbol{\rho}_2')\{|\boldsymbol{\rho}_1-\boldsymbol{\rho}_1'|^{-1}$$

$$+|\boldsymbol{\rho}_1-\boldsymbol{\rho}-\boldsymbol{\rho}_1'|^{-1}+|\boldsymbol{\rho}_2-\boldsymbol{\rho}_2'|^{-1}$$

$$+|\boldsymbol{\rho}_2-\boldsymbol{\rho}-\boldsymbol{\rho}_2'|^{-1}\}d\tau_1 d\tau_2 d\tau_1' d\tau_2'+(e^2/\varepsilon_\infty)$$

$$\times\int\chi_0^2(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2)|\boldsymbol{\rho}_1-\boldsymbol{\rho}-\boldsymbol{\rho}_2|^{-1}d\tau_1 d\tau_2$$

$$= T_1-U_1+U_2. \quad (24)$$

The energy $E_0^{(0)}(\rho)$ is the total energy of two interacting self-trapped electrons, which depends parametrically on the distance ρ. When the relative distance ρ in Eq. (24) increases without bound, one gets twice the result of the single-particle problem. Equation (24) was derived in the long-wavelength ($\mathbf{f}\to0$) approximation $\omega_f=\omega_0$. For ammonia, the limiting frequency ω_0 lies in the range $(2.3–6.3)\times10^{13}$ s^{-1} (Refs. 4 and 35).

A detailed analysis of the functional (24), with allowance for exchange forces and dynamic short-range interelectronic correlations, was performed in Refs. 8 and 23, where it was shown that the two-electron bound state is an axisymmetric dimer. It was also established that for a fixed center of mass, the dominant singlet term $E_0^{(0)}(\rho)$ for electrons in ammonia has a broad potential well with a depth of 0.15 eV, at an equilibrium interparticle spacing $\rho_0=6.2a_0^*$, $a_0^*=\varepsilon^*(m/m^*)a_0$, where a_0 is the Bohr radius. The particles may undergo periodic oscillations described by the relative coordinate ρ about the minimum of the potential

well. For $\rho=0$ the leading term has a maximum, remaining finite, but at large distances this term has Coulomb asymptotic behavior.

The variational parameters minimizing the functional $E_0^{(0)}(\rho)$ were determined numerically and only those solutions satisfying the virial theorem[8,22,23] will be physically meaningful

$$2T(\rho)-U_1(\rho)+\rho dE_0^{(0)}(\rho)/d\rho+U_2(\rho)=0. \qquad (25)$$

Nonfulfillment of the constraints imposed by the virial theorem leads to exaggerated binding energies and thus causes an unjustified expansion of the region of dielectric media where bound two-electron formations may exist. It may also lead to inaccurate symmetry of the continuous strong-coupling bipolaron.

Collecting terms including the electron coordinates in the Hamiltonian H, we obtain the following Hamiltonian describing the electron motion:

$$\begin{aligned}H_0=H_0^{(0)}+i\xi(eH/2m^*c)&\Bigg\{\sum_{i=1,2}[(R_y\partial/\partial\rho_{ix}\\&-R_x\partial/\partial\rho_{iy})+(\rho_{iy}\partial/\partial\rho_{ix}-\rho_{ix}\partial/\partial\rho_{iy})]\\&+\rho_y(b\partial/\partial\rho_{1x}-a\partial/\partial\rho_{2x})-\rho_x(b\partial/\partial\rho_{1y}-a\partial/\partial\rho_{2y})\Bigg\}\end{aligned}$$

$$\begin{aligned}&+\xi^2(e^2H^2/8m^*c^2)\sum_{i=1,2}[(\rho_{ix}^2+\rho_{iy}^2)\\&+2(R_x\rho_{ix}+R_y\rho_{iy})+2(b\rho_x\rho_{ix}-a\rho_y\rho_{iy})].\end{aligned} \qquad (26)$$

The diamagnetic component of the dimer susceptibility due to the electron motion in the potential polarized well can then easily be determined by perturbation theory methods. Using the results[23] for the electron wave eigenfunction $\chi_0(\boldsymbol{\rho}_1,\boldsymbol{\rho}_2)$ of the Hamiltonian $H_0^{(0)}$, we obtain the following value for the diamagnetic contribution to the susceptibility in the multiplicative approximation:

$$\begin{aligned}\chi_e^{(d)}&=-(e^4/4m^*c^2)\sum_{i=1,2}\langle\chi_0|\rho_i^2\delta_{\alpha\beta}-\rho_{i\alpha}\rho_{i\beta}|\chi_0\rangle\\&=-39.52\mu_B^2a_0/e^2,\end{aligned} \qquad (27)$$

where μ_{B} is the Bohr magneton.

A paramagnetic contribution to the susceptibility also appears in the second order of perturbation theory. However, for axisymmetric systems the paramagnetic component can be reliably approximated by:[36]

$$\chi_e^{(p)}=2(\chi_{e,xx}^{(d)}-\chi_{e,zz}^{(d)})^2/|\chi_{e,yy}^{(d)}|=1.52\ \mu_B^2a_0/e^2. \qquad (28)$$

Thus in the selected coordinate system with its origin tied to the center of inertia of the system, the main contribution of the electron component of the susceptibility is diamagnetic. In order to estimate the contributions made to the susceptibility by the other components of the Hamiltonian H, we need to find solutions of the wave equation with the Hamiltonian H_2. The additional unknown quantities $v_f(1)$ and $v_f(2)$ are selected so that the terms in the Hamiltonian H_2 which are linear in the momenta P_f vanish. The requirement that the linear operator with respect to P_f should vanish

derives from the condition for regularity of the wave function in terms of the phonon field variable Q_f. We write the phonon kinetic energy as follows:

$$(1/4)\sum_f\hbar\nu_f\{P_f'+2if S_f^*\mathbf{I}/\hbar+2if Z_f^*\mathbf{J}/\hbar\}$$
$$\times\{P_{-f}'-2if S_{-f}^*\mathbf{I}/\hbar-2if Z_{-f}^*\mathbf{J}/\hbar\}, \qquad (29)$$

where the following notation is used

$$P_f'=\Bigg[P_f-\mathbf{f}v_f^*(1)\sum_k\mathbf{k}(u_k(1)+u_k(2)\exp(i\mathbf{k}\cdot\boldsymbol{\rho}))P_k\Bigg]$$
$$\times\exp(ib\mathbf{f}\cdot\boldsymbol{\rho})+\Bigg[P_f-\mathbf{f}v_f^*(2)\sum_k\mathbf{k}(u_k(2)$$
$$+u_k(1)\exp(-i\mathbf{k}\cdot\boldsymbol{\rho}))P_k\Bigg]\exp(-ia\mathbf{f}\cdot\boldsymbol{\rho}). \qquad (30)$$

Setting $P_f=\pi_f+a_f$, where a_f are some arbitrary numbers, one can easily show[34] that the condition for vanishing of the linear terms with respect to the momenta P_f in the Hamiltonian H_2 may be expressed in the form:

$$\mathbf{f}v_f[\exp(ib\mathbf{f}\cdot\boldsymbol{\rho})+\exp(-ia\mathbf{f}\cdot\boldsymbol{\rho})](S_{-f}^*\mathbf{I}+Z_{-f}^*\mathbf{J})$$
$$=\mathbf{f}\Bigg\{[u_f(1)+u_f(2)\exp(i\mathbf{f}\cdot\boldsymbol{\rho})]$$
$$\times\sum_k\mathbf{k}\cdot\mathbf{k}(S_{-k}^*\mathbf{I}+Z_{-k}^*\mathbf{J})\nu_k v_k^*(1)\exp(ib\mathbf{k}\cdot\boldsymbol{\rho})$$
$$+[u_f(2)+u_f(1)\exp(-i\mathbf{f}\cdot\boldsymbol{\rho})]$$
$$\times\sum_k\mathbf{k}\cdot\mathbf{k}(S_{-k}^*\mathbf{I}+Z_{-k}^*\mathbf{J})\nu_k v_k^*(2)\exp(-ia\mathbf{k}\cdot\boldsymbol{\rho})\Bigg\}. \qquad (31)$$

Taking into account the definition of the complex quantities S_f and Z_f (18), we rewrite Eq. (31) as follows:

$$\mathbf{f}v_f\{(a\mathbf{I}+\mathbf{J})[1+\exp(-i\mathbf{f}\cdot\boldsymbol{\rho})]v_{-f}^*(1)+(b\mathbf{I}-\mathbf{J})$$
$$\times[1+\exp(i\mathbf{f}\cdot\boldsymbol{\rho})v_{-f}^*(2)]\}$$
$$=\mathbf{f}\Bigg\{\sum_k(\mathbf{k}\cdot\mathbf{k})\nu_k(a\mathbf{I}+\mathbf{J})$$
$$\times[(u_f(1)+u_f(2)\exp(i\mathbf{f}\cdot\boldsymbol{\rho}))v_k^*(1)v_{-k}^*(1)+(u_f(2)$$
$$+u_f(1)\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}))v_k^*(2)v_{-k}^*(1)\exp(i\mathbf{k}\cdot\boldsymbol{\rho})]$$
$$+\sum_k(\mathbf{k}\cdot\mathbf{k})\nu_k(b\mathbf{I}-\mathbf{J})[(u_f(1)+u_f(2)$$
$$\times\exp(i\mathbf{f}\cdot\boldsymbol{\rho}))v_k^*(1)v_{-k}^*(2)+(u_f(2)+u_f(1)$$
$$\times\exp(-i\mathbf{f}\cdot\boldsymbol{\rho}))v_k^*(2)v_{-k}^*(2)\exp(i\mathbf{k}\cdot\boldsymbol{\rho})]\Bigg\}. \qquad (32)$$

Equation (32) is easily reduced to a system of two equations for the vectors $\mathbf{I}+\mathbf{J}/a$ and $\mathbf{I}-\mathbf{J}/b$

$$\nu_f\mathbf{f}(\mathbf{I}+\mathbf{J}/a)[1+\exp(-i\mathbf{f}\cdot\boldsymbol{\rho})]v_f(1)$$

$$= \mathbf{f} \sum_k \ (\mathbf{k} \cdot \mathbf{k}) \nu_k (\mathbf{I} + \mathbf{J}/a) [(u_f(1) + u_f(2)$$

$$\times \exp(i\mathbf{f} \cdot \boldsymbol{\rho})) v_k^*(1) v_k(1) + (u_f(2) + u_f(1)$$

$$\times \exp(-i\mathbf{f} \cdot \boldsymbol{\rho})) v_k^*(2) v_k(1) \exp(-i\mathbf{k} \cdot \boldsymbol{\rho})],$$

$$\nu_f \mathbf{f} (\mathbf{I} - \mathbf{J}/b) [1 + \exp(i\mathbf{f} \cdot \boldsymbol{\rho})] v_f(2)$$

$$= \mathbf{f} \sum_k \ (\mathbf{k} \cdot \mathbf{k}) \nu_k (\mathbf{I} - \mathbf{J}/b) [(u_f(2) + u_f(1)$$

$$\times \exp(i\mathbf{f} \cdot \boldsymbol{\rho})) v_k^*(2) v_k(1) + (u_f(1) + u_f(2)$$

$$\times \exp(-i\mathbf{f} \cdot \boldsymbol{\rho})) v_k^*(1) v_k(2) \exp(i\mathbf{k} \cdot \boldsymbol{\rho})]. \tag{33}$$

Using the orthogonality conditions (11), the system (33) is identically satisfied if it is assumed that

$$\mathbf{f} \nu_f v_f(1)(\mathbf{I} + \mathbf{J}/a) = \mathbf{f} \hbar^2 [u_f(1) + u_f(2) \exp(i\mathbf{f} \cdot \boldsymbol{\rho})] (\mathbf{V} + \mathbf{W}),$$

$$\mathbf{f} \nu_f v_f(2)(\mathbf{I} - \mathbf{J}/b) = \mathbf{f} \hbar^2 [u_f(2) + u_f(1) \exp(-i\mathbf{f} \cdot \boldsymbol{\rho})] (\mathbf{V} - \mathbf{W}). \tag{34}$$

Here $\mathbf{V}$ is the velocity vector of the translational displacement of the center of inertia of the system, and $\mathbf{W}$ is the velocity vector of the relative motion of the electrons. We divide each of the equations (34) by the frequency ν_f and multiply the first equation by $\mathbf{f}[u_{-f}(1) + u_{-f}(2) \exp(-i\mathbf{f} \cdot \boldsymbol{\rho})]$ and the second equation by $\mathbf{f}[u_{-f}(2) + u_{-f}(1) \exp(-i\mathbf{f}\boldsymbol{\rho})]$, and we then sum over the wave vector $\mathbf{f}$. Using the orthogonality conditions (11), we obtain the system

$$\mathbf{I} + \mathbf{J}/a = (\mathbf{V} + \mathbf{W}) \sum_f \ \hbar^2 \mathbf{f} \cdot \mathbf{f} [u_{-f}(1) + u_{-f}(2) \exp(-i\mathbf{f} \cdot \boldsymbol{\rho})]$$

$$\times [u_f(1) + u_f(2) \exp(i\mathbf{f} \cdot \boldsymbol{\rho})] / \nu_f,$$

$$\mathbf{I} - \mathbf{J}/a = (\mathbf{V} - \mathbf{W}) \sum_f \ \hbar^2 \mathbf{f} \cdot \mathbf{f} [u_{-f}(2) + u_{-f}(1) \exp(i\mathbf{f} \cdot \boldsymbol{\rho})]$$

$$\times [u_f(2) + u_f(1) \exp(-i\mathbf{f} \cdot \boldsymbol{\rho})] / \nu_f. \tag{35}$$

Then, taking account of the physical meaning of the vectors $\mathbf{I}$, $\mathbf{J}$, $\mathbf{V}$, and $\mathbf{W}$, we can find the translational effective mass of a two-electron formation,

$$M_R = (\hbar^2/3) \sum_f \ \mathbf{f} \cdot \mathbf{f} [u_{-f}(2) + u_{-f}(1) \exp(i\mathbf{f} \cdot \boldsymbol{\rho})]$$

$$\times [u_{-f}(2) + u_{-f}(1) \exp(i\mathbf{f} \cdot \boldsymbol{\rho})] / \omega_f, \tag{36}$$

and also its reduced mass

$$M_\rho = M_1 M_2 M_R / (M_1 + M_2)^2, \tag{37}$$

where the effective mass of the ith solvated electron is given by

$$M_i = (\hbar^2/3) \sum_f \ \mathbf{f} \mathbf{f} u_{-f}(i) u_f(i) / \omega_f.$$

The Hamiltonian H_2 describing the relative motion of the particles, the translational displacement of the bound two-particle formation as a whole, and the motion of the renormalized phonons may be finally written in the following form:

$$\xi^2 H_2 = (1/2) \sum_f \ \hbar \omega_f (Q_f' Q_{-f}' + \Pi_f \Pi_{-f})$$

$$\times + P_\mathbf{R}^2 / 2M_R + P_\rho^2 / 2M_\rho + \mathbf{P}_\mathbf{R} \cdot \mathbf{P}_\rho / 2M_{R\rho}, \tag{38}$$

where

$$M_{R\rho} = (4/3) \sum_f \ \mathbf{f} \cdot \mathbf{f} \omega_f [2a v_f^*(1) v_{-f}^*(1)$$

$$- 2b v_f^*(2) v_{-f}^*(2)$$

$$+ (b - a) v_f^*(2) v_{-f}^*(1) \exp(i\mathbf{f} \cdot \boldsymbol{\rho})$$

$$+ (b - a) v_f^*(1) v_{-f}^*(2) \exp(-i\mathbf{f} \cdot \boldsymbol{\rho})] / \hbar,$$

$$Q_f' = Q_f [1 + \exp(i\mathbf{f} \cdot \boldsymbol{\rho})],$$

$$\Pi_f = [\pi_f'(1) \exp(ib\mathbf{f} \cdot \boldsymbol{\rho}) + \pi_f'(2) \exp(-ia\mathbf{f} \cdot \boldsymbol{\rho})]/2,$$

$$\pi_f'(1) = \pi_f - \mathbf{f} \cdot v_f^*(1) \sum_k \ \mathbf{k}[u_k(1) + u_k(2)$$

$$\times \exp(i\mathbf{k} \cdot \boldsymbol{\rho})] \pi_k,$$

$$\pi_f'(2) = \pi_f - \mathbf{f} \cdot v_f^*(2) \sum_k \ \mathbf{k}[u_k(2) + u_k(1)$$

$$\times \exp(-i\mathbf{k} \cdot \boldsymbol{\rho})] \pi_k.$$

The first sum in the Hamiltonian (38) corresponds to the energy of oscillators which do not interact with the sources of the polarization field. For a symmetric bound two-electron formation, $M_1 = M_2$, the parameters are $a = b = 1/2$, and therefore $M_{R\rho} = 0$. The second and third terms in Eq. (38) determine the kinetic energies of the translational and relative motion.

Since for strong electron–phonon coupling, the kinetic energy of the relative and translational motion under normal experimental conditions is of the same order of magnitude, one must in general carry out a joint analysis of the motion in the coordinates ρ and R and in the coordinates Q_f. However, since the characteristic size of the region of localization of each particle is $\sqrt{\langle \rho^2 \rangle} \simeq \xi^{1/2}$ and the main contribution to the sum over the vector $\mathbf{f}$ is made by values $|\mathbf{f}||\boldsymbol{\rho}| \gg 1$, the exponential function $\exp(i\mathbf{f} \cdot \boldsymbol{\rho})$ is therefore rapidly oscillating. Then, in the zeroth approximation with respect to the magnetic field the variables $\boldsymbol{\rho}$, $\mathbf{R}$, and Q_f separate, and the wave eigenfunction $\Theta_0(\boldsymbol{\rho}, \mathbf{R}, Q_f)$ of the Hamiltonian H_2 may be expressed in the product form $\Theta_0(\boldsymbol{\rho}, \mathbf{R}, Q_f) = \varphi_0(\boldsymbol{\rho}) \eta_0(\mathbf{R}) \Phi(Q_f)$. Averaging the Hamiltonian H (17) over the ground-state electron wave function $\chi_0(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$ and bearing in mind that the Hamiltonian H_2 can be separated into independent parts, we obtain an equation describing the relative motion of the particles in the potential field $E_0^{(0)}(\rho)$:

$$[P_\rho^2 / 2M_\rho + E_0^{(0)}(\rho) + \xi^2 (e^2 H^2 / 8m^* c^2)$$

$$\times (\rho_x^2 + \rho_y^2)] \varphi(\boldsymbol{\rho}) = W \varphi(\boldsymbol{\rho}) \tag{39}$$

and the equation of motion for the center of mass as a whole

$$[P_\mathbf{R}^2 / 2M_R + \xi^2 (e^2 H^2 / 8m^* c^2)$$

$$(R_x^2 + R_y^2)]\,\eta(\mathbf{R}) = F\,\eta(\mathbf{R}). \tag{40}$$

Equation (40) leads to the appearance of Landau orbital diamagnetism whose magnitude per particle in this case is given by

$$\chi_L = -(m^2/m^*M_R)\mu_B^2/k_B T = -3.75\times10^{-5}\mu_B^2/k_B T. \tag{41}$$

The translational mass of a bound two-electron formation is not a sum of additive effective masses of the individual particles, and studies have shown[21] that it has the value $M_R = 9\times10^{-2}\alpha_c^4 m^*$.

The potential $E_0^{(0)}(\rho)$ must be defined analytically in order to determine the relative oscillation spectrum (39). The potential energy $E_0^{(0)}(\rho)$ for the ground singlet term was calculated by a direct variational method in Ref. 23. An analysis has shown[37] that the dependence of the leading term of the two-electron formation on the distance ρ may be approximated by the analytical expression

$$E_0^{(0)}(\rho) = -\mathrm{Ry}^*(1-\varepsilon^*/\varepsilon_\infty)a_0^*/\rho + V_{\mathrm{eff}}(\rho). \tag{42}$$

The first term in Eq. (42) describes the long-range screened Coulomb repulsion of the quasiparticles. The second term in Eq. (42) may be written as:

$$V_{\mathrm{eff}}(\rho) = \mathrm{Ry}^*\{D + C\rho^2$$
$$+ [A + B(\rho-\rho_1)^2][1-\exp(-g\rho)]$$
$$+ a_0^*(1-\varepsilon^*/\varepsilon_\infty)/\rho\}\exp(-\delta\rho), \tag{43}$$

which gives a short-range attraction of like-charged quasiparticles owing to the exchange of longitudinal quanta of the polarization field. A correlation analysis was used to obtain the following parameters of the approximation (43):

$$A = -0.01952, \quad B = 1.2689\times10^{-3}/a_0^*,$$
$$C = -4.5854\times10^{-4}/a_0^{*2}, \quad D = -9.5238\times10^{-3},$$
$$g = 0.195/a_0^*, \quad \delta = 0.345/a_0^*, \quad \rho_1 = 0.05a_0^*.$$

The energy origin in Eq. (43) is taken as the energy equal to the sum of two complete self-consistent energies of infinitely separated ground-state solvated electrons. The choice of potential in the form (42), (43) has been physically substantiated. At the point $\rho = 0$ the potential is finite, as is deduced from variational calculations,[8,23] and for $\rho\to\infty$ the approximation (42), (43) gives the correct Coulomb asymptotic form.

It is impossible to obtain an accurate analytic solution of the Schrödinger equation with the potential (39), but satisfactory approximations to the solution of the radial equation may be found by the shifted $1/N$ expansion method.[38] For smooth potentials the analytical technique of the shifted $1/N$ expansion can give the eigenspectrum with a high degree of accuracy, similar to the results of the exact analytical solution. The method involves using an expansion as a series in terms of the dimensionless parameter $k = N + 2l - a$, where N is the number of spatial measurements and $l(l+N)\hbar^2$ is the square of the eigenvalue of the N-dimensional orbital angular momentum. The shift parameter a will be determined subsequently.

To determine the spectrum of Eq. (39) with the potential (42) in the zeroth approximation with respect to the magnetic field, we reduce the Schrödinger equation (39) to the following form:

$$-\frac{\hbar^2}{2M_\rho}\left\{\frac{d^2}{d\rho^2} + \frac{N-1}{\rho}\frac{d}{d\rho}\right.$$
$$\left. + \frac{l(l+N-2)}{\rho^2} + E_0^{(0)}(\rho)\right\}\varphi^{(0)} = W\varphi^{(0)}. \tag{44}$$

The superscript in $E_0^{(0)}(\rho)$ will subsequently be omitted. The approximate wave function of the time-independent equation (44) with a centrosymmetric potential in the approximation of the shifted $1/N$ expansion may be written as $\varphi^{(0)} \times(\rho) = R_{nl}\rho^{-(N-1)/2}Y_{lm}(\theta,\psi)$, where $Y_{lm}(\theta,\psi)$ are spherical harmonics, n is the radial quantum number, and the principal quantum number is $n+l+1$. The radial component $R_{nl}(\rho)$ of the wave function satisfies the equation[38]

$$-\frac{\hbar^2}{2M_\rho}\left[\frac{d^2}{d\rho^2} - \frac{(N-1)(3-N)}{4\rho^2} + \frac{l(l+N-2)}{\rho^2} + E_0^{(0)}(\rho)\right]$$
$$\times R_{nl} = W_{nl}R_{nl}. \tag{45}$$

The $1/N$ expansion method usually only gives asymptotic convergence for eigenvalues. To overcome this difficulty, it has been suggested[38] to introduce an additional shift parameter a. The shift parameter a is selected so that the results of the shifted $1/N$ expansion technique agree with the exact analytical results for the eigenvalues of an N-dimensional harmonic oscillator. With allowance for the additional parameter a, the Schrödinger equation (53) may be rewritten as:

$$-\frac{\hbar^2}{2M_\rho}\frac{d^2R_{nl}}{d\rho^2} + \frac{\hbar^2}{8M_\rho}\frac{k^2}{\rho^2}[1-(1-a)/k]$$
$$\times[1-(3-a)/k]R_{nl} + E_0(\rho)R_{nl} = W_{nl}R_{nl}. \tag{46}$$

As in Ref. 38, we also introduce the scale factor Q and rewrite Eq. (46) in the form:

$$-\frac{\hbar^2}{2M_\rho}\frac{d^2R_{nl}}{d\rho^2} + k^2\left\{\frac{\hbar^2}{8M_\rho\rho^2}[1-(1-a)/k]\right.$$
$$\left.\times[1-(3-a)/k] + \frac{E_0(\rho)}{Q}\right\}R_{nl} = W_{nl}R_{nl}. \tag{47}$$

The main contribution to the energy in Eq. (49) is made by the effective potential

$$E_{\mathrm{eff}}^{(0)}(\rho) = \hbar^2/8M_\rho\rho^2 + E_0(\rho)/Q. \tag{48}$$

The scale factor Q is determined from the condition of minimum effective potential (48) in terms of the coordinate ρ

$$Q = 4M_\rho\rho^3 E^{(1)}(\rho_m)/\hbar^2, E^{(j)} = d^jE_0(\rho_m)/d\rho^j, \tag{49}$$

where ρ_m is the position of the local minimum of the effective potential (48).

It is subsequently convenient to convert to the dimensionless variable $x = k^{1/2}(\rho - \rho_m)/\rho_m$ in Eq. (47). Using the variable x, we expand the potential in Eq. (47) as a series in terms of x and $k^{-1/2}$. The Schrödinger equation (49) can then be rewritten as:

$$\{-(\hbar^2/2M_\rho)d^2/dx^2 + (k\hbar^2/8M_\rho)$$
$$\times(1 + 3x^2/k - 4x^3/k^{3/2} + 5x^4/k^2 - \dots)$$
$$-(2-a)\hbar^2(1 - 2x/k^{1/2} + 3x^2/k - \dots)/4M_\rho$$
$$+(1-a)(3-a)\hbar^2(1 - 2x/k^{1/2} + 3x^2/k - \dots)/8kM_\rho$$
$$+\rho_m^2 k[E_0(\rho_m) + E^{(2)}(\rho_m)\rho_m^2 x^2/2k$$
$$+E^{(3)}(\rho_m)\rho_m^3 x^3/6k^{3/2} + \dots]/Q\}R_{n,l}(\rho)$$
$$= W_{n,l}\rho_m^2 R_{n,l}(\rho)/k. \tag{50}$$

The effective potential (48) is of order k^2 and is identified with the leading approximation in the shifted $1/N$ expansion method. This expansion is physically justified by the fact that when $N \to \infty$, the quantum system behaves as a static classical system. The next contribution to the energy is of order k and may be written as

$$k[(n+1/2)\hbar\Omega - (2-a)\hbar^2/4M_\rho]/\rho_m^2, \quad n = 0,1,2,\dots, \tag{51}$$

where the frequency of the normal oscillations about the equilibrium position is determined from Eq. (50) and is given by

$$\Omega = [3\hbar^2/4M_\rho^2 + \rho_m^4 E^{(2)}(\rho_m)/M_\rho Q]^{1/2}$$
$$= (\hbar/2M_\rho)[3 + \rho_m E^{(2)}(\rho_m)/E^{(1)}(\rho_m)]^{1/2}. \tag{52}$$

The shift parameter $a = 2 - 2(2n+1)M_\rho\Omega/\hbar$ is determined from the condition that the solutions of the Schrödinger equation obtained by the $1/N$ expansion technique in the leading approximation and the exact analytic results for a harmonic oscillator should be equal. In order to match Eqs. (45) and (46), we assume $Q = k^{1/2}$. Then, using the definition $Q = 4M_\rho\rho_m^3 E^{(1)}(\rho)/\hbar^2$, we obtain

$$2l + N - 2 + (2n+1)[3 + \rho_m E^{(2)}(\rho_m)/E^{(1)}(\rho_m)]^{1/2}$$
$$= (4M_\rho\rho_m^3 E^{(1)}(\rho_m)/\hbar^2)^{1/2}, \tag{53}$$

from which the position of the local minimum ρ_m is determined for a fixed set of quantum numbers n and l. For the three-dimensional case, we obviously need to assume $N = 3$ in Eq. (53). Equation (52) then gives the following rapidly converging series for the eigenvalues:

$$W_{n,l} = (k/\rho_m)^2[\hbar^2/8M_\rho + \rho_m^2 E_0(\rho_m)/Q$$
$$+ \gamma^{(1)}/k^2 + \gamma^{(2)}/k^3 + O(1/k^4)]$$
$$= W_{n,l}^{(0)} + W_{n,l}^{(1)} + W_{n,l}^{(2)} + \dots. \tag{54}$$

The corrections to the leading contribution to the energy $W_{n,l}^{(0)}$ from quantum fluctuations and anharmonicity effects are cumbersome and are not given here. The analytical form of these corrections was given in Ref. 38. Thus the determination of the spectrum of normal modes associated with the

TABLE I. Vibrational eigenstates of a quasimolecular dimer solvated in ammonia

n	ρ_m, a_0^*	W_n, $10^{-3}e^4 m^*/\varepsilon^{*2}\hbar^2$	Ω, $10^{-3}\hbar/m^*$
0	6.115	-7.220	5.107
1	6.365	-6.105	3.16
2	6.59	-5.329	2.497
3	6.81	-4.666	2.18
4	7.03	-4.09	1.97
5	7.255	-3.57	1.82
6	7.476	-3.086	1.71
7	7.695	-2.600	1.617
8	7.912	-2.11	1.54
9	8.126	-1.606	1.474
10	8.34	-1.086	1.41
11	8.55	-0.551	1.365
12	8.756	-3.12×10^{-3}	1.31
13	8.96	0.552	1.275
14	9.16	1.108	1.235
15	9.36	1.657	1.198
16	9.55	2.19	1.162
17	9.73	2.7	1.13

relative motion of the particles in a two-electron formation reduces to solving the transcendental equation (53).

The results of the calculations are presented in Table I, which gives the vibrational energies W_n, the position of the local minimum ρ_m, and the vibration frequency for a given quantum number n. It was found that the total number of vibrational states is limited, and five are quasi-steady states, since they lie above the asymptotic curve of the pair potential. The position of the local minimum for the ground vibrational state is close to the minimum point of the pair potential (43), which implies that near the minimum the pair potential is close to the harmonic potential. Using the wave functions given in Ref. 38, it is easy to calculate the magnetic susceptibility due to the relative motion of the particles. Since the interval between the vibrational energy levels is of comparable order of magnitude to the thermal energy $k_B T$, we write the magnetic susceptibility as:

$$\chi_\rho^{(d)} = -\frac{\varepsilon^{*2}\mu_B^2 a_0^*}{6e^2} - \left(\frac{m}{m^*}\right)^3$$
$$\times \frac{\sum_n\langle\varphi_n^{(0)}|\rho^2|\varphi_n^{(0)}\rangle\exp(-\Delta W_n/k_B T)}{\sum_n\exp(-\Delta W_n/k_B T)}, \tag{55}$$

where $\Delta W_n = W_n - W_0$.

Summing in Eq. (55) over all the vibrational states (54), we finally obtain the magnetic susceptibility due to the relative motion: $\chi_\rho^{(d)} = -5.11\mu_B^2 a_0/e^2$.

The magnetic susceptibility of the solution will consist of the susceptibilities of individual solvated electrons and of two-electron quasimolecular formations, and may be expressed as:

$$\chi = n_s\chi_1 + n_d\chi_2, \tag{56}$$

where χ_1 is the total susceptibility of a solvated electron [consisting of the Pauli spin paramagnetic susceptibility $\chi_1^{(p)} = \mu_B^2/k_B T$, which gives the main contribution, and the single-electron corrections calculated using formulas (27)

and (43)], χ_2 is the total susceptibility of a bound two-electron formation [with components given by (27), (28), (41), and (55)], and n_s and n_d are the concentrations of solvated electrons and dimers.

The translational effective mass of a solvated electron was taken to be $M_1 = 0.023\alpha_c^4 m^*$ (Refs. 33 and 34). We assume that the total number of electrons n_0 in the solution corresponds to the concentration of dissolved alkali metal. The condition for conservation of the total number of electrons can then be written as $n_s + 2n_d = n_0$. In accordance with the law of mass action,[39] the equilibrium constant K of the reaction $n_d \rightleftarrows 2n_s$ in the thermodynamic equilibrium state at temperature T is given by

$$n_0 K = \frac{n_s^2}{n_d} = 4\left(\frac{M_1 k_B T}{2\pi\hbar^2}\right)^3 \left(\frac{2\pi\hbar^2}{M_R k_B T}\right)^{3/2} \frac{\exp(-D/k_B T)}{Z_{\text{vib}} Z_{\text{rot}}}, \tag{57}$$

where Z_{vib} and Z_{rot} are the vibrational and rotational partition functions of the dimer, and D is its dissociation energy.

The factor of 4 in formula (57) is due to the spin variables of the quasiparticles. We neglect the contributions to the partition functions associated with electronic transitions, since the energy interval between the ground electronic level and the first excited level for single-particle and two-particle formations is 0.9 eV $\gg k_B T$. The first two factors in parentheses in formula (57) are attributed to the translational partition functions of the solvated electrons and dimers, respectively. The vibrational partition function was calculated by direct summation using the results of the shifted $1/N$ expansion method. The rotational energy levels of the dimer were so close together (with an energy interval of the order of $W_{n,l} - W_{n,l-1} \simeq 10^{-4}$ Ry*) that the rotational spectrum can be treated as continuous. In this case, the rotational partition function is replaced by an integral and, in the approximation $T > T_{\text{rot}}$, may be written as $Z_{\text{rot}} = T/T_{\text{rot}}$, with $T_{\text{rot}} = \hbar^2/2k_B J_{\text{rot}}$. The rotational moment of inertia of the dimer relative to the x axis, perpendicular to the dimer coupling axis, may be written as:[23]

$$J_{\text{rot}} = (4/3)\pi e^2 (\omega_0^2 \varepsilon^*)^{-1} \int \chi_0^4(\rho_1, \rho_2)$$
$$\times (\rho_{1y}^2 + \rho_{1z}^2) d\tau_1 d\tau_2 = 0.35\alpha_c^2 \hbar/\omega_0.$$

Figure 1 shows the change in the relative equilibrium concentrations of single-electron and two-electron states, calculated according to formula (57), as a function of the concentration of dissolved alkali metal at the temperatures 198 K and 238 K for which the susceptibility of a metal–ammonia solution has been measured.[1,2] It can be seen that the formation of bound electron pairs is initiated at alkali metal concentrations of $\simeq 5 \times 10^{17}$ cm^{-3}, in the range $n_0 \simeq 3 \times 10^{18}$–$10^{19}$ cm^{-3}, the number of singlet dimers and solvated electrons is equalized, and as n_0 increases further, the number of diamagnetic formations begins to exceed the number of single-particle states. It is important to stress that in this comparatively narrow range of concentrations, the maximum of the optical absorption band undergoes a long-wavelength shift[7] which is identified with the formation of singlet dimers.[8] The concentration changes in the optical
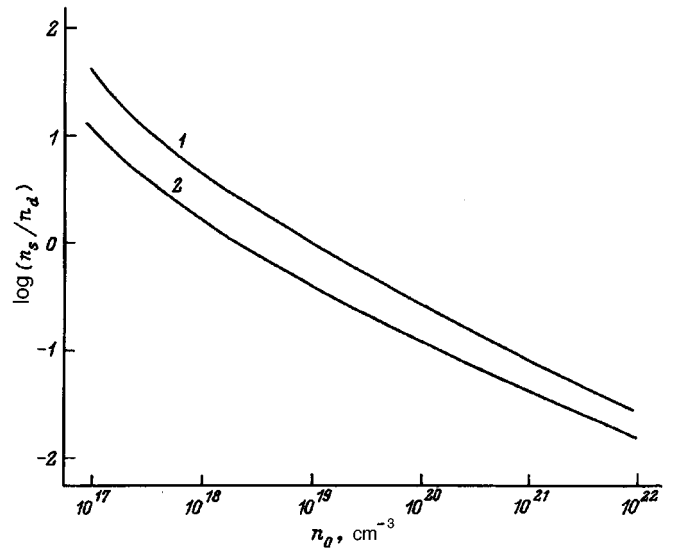

FIG. 1. Relative concentrations of solvated electrons and bound two-electron states as a function of the dissolved metal concentration. T, K: *1* — 238 and *2* — 198.

characteristics are completely correlated[15] with a drop in the spin susceptibility of the electronic subsystem. In this range of concentrations, a drop in the equivalent conductivity of a metal–ammonia solution to approximately 20% of its value in dilute solutions is observed experimentally, which may be explained by the formation of bound two-electron formations[15] whose effective size is greater than that of a solvated electron. The results of the calculations plotted in Fig. 1 are also confirmed by measurements[40] of the rates of proton relaxation due to the hyperfine interaction between the proton spin and the spin system of the electron ensemble. These investigations also show that the formation of singlet spin pairs is initiated at electron concentrations $n_s \simeq (5 \times 10^{17})$–$(3 \times 10^{18})$ cm^{-3}.

To compare the results of the theoretical calculations of the static susceptibility with experiment, we introduce the quantity

$$C_k = N_A \chi T/n_s$$
$$= N_A \mu_B^2 [\chi_1/\chi_1^{(p)} - (\chi_2/\chi_1^{(p)} - 2)n_d/n_s]/k_B, \tag{58}$$

where N_A is Avogadro's number.

In Fig. 2 a comparison is made between the theoretical (58) and experimental results[1,2] for the magnetic susceptibility of metal–ammonia solutions at temperatures of 198 and 238 K as a function of the degree of dilution of the solution $\lg(n/n_0)$, where $n = 2.43 \times 10^{22}$ cm^{-3} is the concentration of solvent atoms[4] and n_0 is alkali metal concentration, which is assumed to be equal to the number of electrons in the solution. The theoretical curves obtained using the model which presupposes the coexistence of two types of particles in a metal–ammonia solution — solvated electrons and two-electron singlet formations — accurately describes the changes in the susceptibility of a metal–ammonia solution as a function of temperature and concentration. The paramagnetism decreases with increasing concentration of the solu-
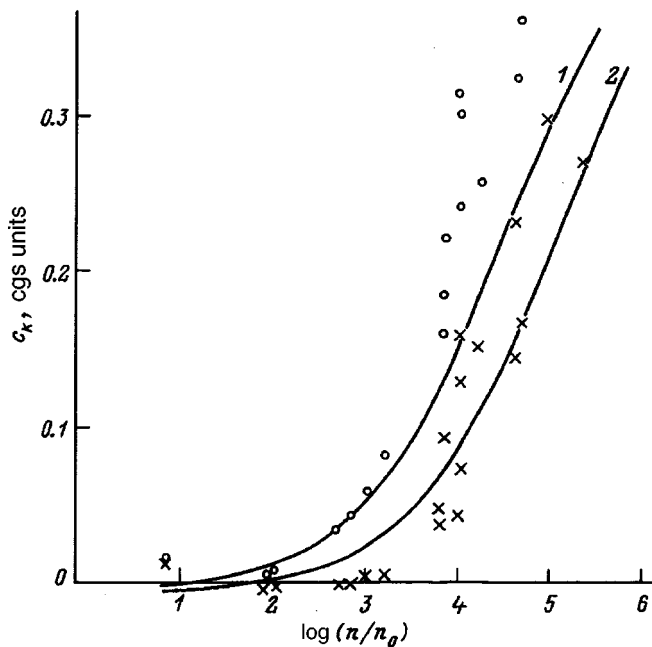
FIG. 2. Static susceptibility of a metal–ammonia solution as a function of the dilution. The solid curves give the calculations using formula (58); $\bigcirc$, $\times$ — experimental values[1,2] for $T = 238$ and 198 K, respectively.

tion and with decreasing temperature, and at concentrations $n_0 > 10^{21}$ cm^{-3} the susceptibility changes sign.

Some discrepancy with experiment in the case of comparatively weak dilutions $\leqslant 4 \times 10^{17}$ cm^{-3} may be attributed to the influence of a repulsive potential barrier in the pair interparticle potential $E_0^{(0)}(\rho)$, the maximum of which, $\simeq 2.9 \times 10^{-3}$Ry* is attained at distances $\rho \simeq 21 a_0^*$ (Refs. 8 and 23). This obviously shifts the reaction $n_d \rightleftarrows 2n_s$ for the temperature range under study toward the formation of individual solvated electrons and thus enhances the paramagnetism of the solution at low electron concentrations. However, as the concentration of the dissolved alkali metal increases to $\simeq 4 \times 10^{18}$ cm^{-3}, the barrier height is reduced by Debye screening, so that the equilibrium constant K of Eq. (57) is no longer sensitive to the presence of the potential barrier.

[1] E. Huster, Ann. Phys. (Leipzig) **33**, 477 (1938).
[2] S. Freed and N. Sugarman, J. Chem. Phys. **11**, 354 (1943).
[3] A. Demortier and G. Lepoutre, C. R. Acad. Sci. (Paris) **268**, 453 (1969).
[4] M. F. Deĭgen, Zh. Éksp. Teor. Fiz. **26**, 293, 300 (1954).
[5] J. C. Thompson, *Electrons in Liquid Ammonia* (Oxford University Press, Oxford) [Russian trans., Mir, Moscow, 1979].
[6] R. Caterall and M. C. R. Symons, J. Chem. Soc. 3763 (1965); J. Chem. Soc. A 13 (1966).

[7] W. H. Koehler and J. J. Lagovski, J. Phys. Chem. **73**, 2329 (1969).
[8] V. K. Mukhomorov, Opt. Spektrosk. **69**, 71 (1990) [Opt. Spectrosc. (USSR) **69**, 43 (1990)].
[9] N. Kestner and D. A. Copeland, in *Proceedings of the International Conference on Metal–Ammonia Solutions*, edited by J. J. Lagovski and M. J. Sienko (London, 1970), p. 27.
[10] J. Zimbric and L. Kevan, J. Am. Chem. Soc. **89**, 2484 (1967).
[11] Schettler and G. Lepoutre, J. Phys. Chem. **79**, 2823 (1975).
[12] T. L. Hill, J. Chem. Phys. **16**, 394 (1948).
[13] R. A. Ogg, J. Am. Chem. Soc. **68**, 155 (1946).
[14] R. H. Land and D. E. O'Reilly, J. Chem. Phys. **46**, 4496 (1967).
[15] N. F. Mott, *Metal–Insulator Transitions* (Taylor and Francis, 1974) [Russian trans., Mir, Moscow, 1979].
[16] E. Arnold and A. Paterson, J. Chem. Phys. **41**, 3089 (1964).
[17] L. Onsager, Rev. Mod. Phys. **40**, 710 (1968).
[18] F. Tehan, M. Lok, and J. Dye, in *Proceedings of the 165th National Meeting of the American Chemical Society, Dallas, 1973*.
[19] D. F. Burow and J. J. Lagowski, Adv. Chem. Ser. **50**, 125 (1965).
[20] C. A. Hutchinson and R. Pastor, Rev. Mod. Phys. **25**, 285 (1953).
[21] V. K. Mukhomorov, Opt. Spektrosk. **74**, 1083 (1993) [Opt. Spectrosc. **74**, 644 (1993)].
[22] V. K. Mukhomorov, Fiz. Tekh. Poluprovodn. **14**, 1587 (1980); **16**, 1095 (1982) [Sov. Phys. Semicond. **14**, 940 (1980); **16**, 700 (1982)].
[23] V. K. Mukhomorov, Khim. Fiz. **2**, 642 (1983); Opt. Spektrosk. **55**, 246 (1983) [Opt. Spectrosc. (USSR) **55**, 145 (1983)].
[24] J. Jortner, J. Chem. Phys. **30**, 839 (1959).
[25] V. N. Shubin and S. A. Kabalka, *Theory and Methods of Radiation Chemistry of Water* [in Russian], Nauka, Moscow (1969).
[26] Yu. T. Mazurenko and V. K. Mukhomorov, Opt. Spektrosk. **41**, 930 (1976); [Opt. Spectrosc. (USSR) **41**, 550 (1976)].
[27] J. Jortner and S. Rice, Adv. Chem. Ser. **50**, 7 (1965).
[28] I. Z. Fisher, *Statistical Theory of Liquids* (University of Chicago Press, Chicago, 1964) [Russian orig., Fizmatgiz, Moscow, 1961].
[29] A. I. Gubanov, *Quantum Electron Theory of Amorphous Conductors* (Consultants Bureau, New York, 1965) [Russian original, USSR Academy of Sciences Press, Moscow (1963, 1969)].
[30] V. K. Mukhomorov, Fiz. Tverd. Tela. (St. Petersburg) **34**, 2766 (1992) [Sov. Phys. Solid State **34**, 1480 (1992)].
[31] V. K. Mukhomorov, Fiz. Tverd. Tela. (Leningrad) **29**, 2902 (1987) [Sov. Phys. Solid State **29**, 1668 (1987)].
[32] W. B. da Costa and F. M. Peeters, J. Phys. Condens. Mater. **7**, 1293 (1995).
[33] N. N. Bogolyubov, Ukr. Mat. Zh. **2**, 3 (1950).
[34] S. V. Tyablikov, Zh. Éksp. Teor. Fiz. **21**, 377 (1951).
[35] A. Anderson and S. A. Walmsley, Mol. Phys. **9**, 1 (1965).
[36] Ya. G. Dorfman, *Diamagnetism and the Chemical Bond* (American Elsevier, New York, 1965) [Russian original, Fizmatgiz, Moscow, 1961].
[37] V. K. Mukhomorov, Opt. Spektrosk. **77**, 21 (1994) [Opt. Spectrosc. **77**, 28 (1994)].
[38] T. Imbo, A. Pagnamenta, and U. Sakhatme, Phys. Rev. D **29**, 1669 (1984).
[39] R. Kubo, *Statistical Mechanics: an Advanced Course with Problems and Solutions* (North-Holland, Amsterdam, 1965) [Russian translation, Mir, Moscow, 1967].
[40] R. A. Newmark, J. C. Stephenson, and J. S. Waugh, J. Chem. Phys. **46**, 3514 (1967).

Translated by R. M. Durham

Localization of a nonlinear switching wave in an active bistable medium with an isolated inhomogeneity

S. V. Petrovskiĭ

P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences, 117851 Moscow, Russia

An analysis is made of the one-dimensional propagation of a switching wave in an active medium when one of the parameters of the medium is a function of the coordinate over a certain region and remains constant outside this region. The possibility of localization (stopping) of the wave at this type of inhomogeneity is investigated. An analytical expression is obtained for the wave shift in the case of a small-amplitude inhomogeneity. Constraints on the amplitude and length of the inhomogeneity at which localization takes place are obtained for an inhomogeneity of finite dimensions. © *1997 American Institute of Physics.*
[S1063-7842(97)00208-0]

INTRODUCTION

Various aspects of the dynamics of nonlinear waves in active dissipative media have been addressed in numerous studies (see, for example, Refs. 1–7). A detailed mathematical analysis of wave regimes as applied to flame propagation problems was reported in Ref. 8. General aspects of the wave dynamics of various physical types of dissipative media were considered in Refs. 9–11.

Most of these studies have been concerned with the propagation of nonlinear waves in spatially homogeneous media. The question arises as to what extent the inhomogeneity that is always present to some degree in real systems may alter the nature of the dynamics. It was shown in Ref. 11 that a point inhomogeneity may lead to localization (stopping) of the wave. Here, an analysis is made of the propagation of a nonlinear switching wave in an active medium in the presence of an isolated inhomogeneity of finite dimensions. It is shown that if the amplitude of the inhomogeneity is small, the presence of an inhomogeneous section merely shifts the position of the wave, and an expression for this shift is derived in the first order of perturbation theory. However, if the amplitude is large (but finite), even an inhomogeneity of short length may lead to stopping of the wave. Analytical expressions describing the relation between the amplitude and length of a critical-scale inhomogeneity are obtained for a bistable medium with an arbitrary nonlinearity. The results show good agreement with the numerical modeling data.

FORMULATION OF THE PROBLEM

In many physically different cases the process of wave propagation in an active medium may be described by the same equation[8–10]

$$\partial_t u = D\partial^2_{xx}u + f(u). \tag{1}$$

Here u is a quantity describing the state of the medium (for instance, the temperature), D is the thermal conductivity or the diffusion coefficient, x is the coordinate, t is the time, and the function $f(u) = -\alpha u + f_1(u)$ describes the kinetics of the active medium (a typical form of $f(u)$ is shown in Fig.

1a), where the term $-\alpha u$ describes the heat losses for thermal waves and α in general depends on the coordinates.

Let us first consider a homogeneous medium $\alpha(x) \equiv \alpha_0 = $ const. In this case, the description of the process using Eq. (1) corresponds to the model of a bistable medium:[10] the form of the function $f(u)$ allows (subject to various constraints on the value of the coefficient α_0) three uniform stationary states to exist in the system: $u = u_2$, $u = u_1$, and $u = u_0$ of which the intermediate state is unstable with respect to small perturbations. Note that the value $u_2 - u_0$ is the natural scale parameter for the function $u(x,t)$: as a result of a linear fractional transformation of the dependent variable $\widetilde{u} = (u - u_0)/(u_2 - u_0)$, keeping the form of Eq. (1) unchanged, we have $\widetilde{u}_0 = 0$, $\widetilde{u}_2 = 1$. Thus, we subsequently assume throughout that $u_0 = 0$, $u_2 = 1$.

We determine the dimensionless variables $x' = x(D/\alpha_0)^{-1/2}$, $t' = t\alpha_0$ by a standard method (the primes are subsequently omitted). Equation (1) has the form

$$\partial_t u = \partial^2_{xx}u + F(u), \tag{2}$$

where $F(u) = -u + (1/\alpha_0)f_1(u)$.

It is well known[1,2,8–10] that subject to the boundary conditions

$$u(x \to -\infty, t) = 1, \quad u(x \to \infty, t) = 0 \tag{3}$$

and transient initial conditions, the solution of Eq. (2) for fairly long times is an aperiodic steady-state "switching wave." The wave profile is described by the equation

$$\Theta''(\xi) + v\Theta'(\xi) + F(\Theta) = 0 \tag{4}$$

(where $\xi = x - vt$, v is the wave velocity) with the corresponding conditions at infinity

$$\Theta(\xi \to -\infty) = 1, \quad \Theta(\xi \to \infty) = 0. \tag{5}$$

No solution of the nonlinear problem (4), (5) exists for any values of the velocity v (Refs. 1,3, and 5). The wave velocity is determined from the condition for the existence of a solution that is an eigenvalue of the boundary-value problem.[1,5,8] The structure of the eigenvalue spectrum is determined by the form of the nonlinearity of F.
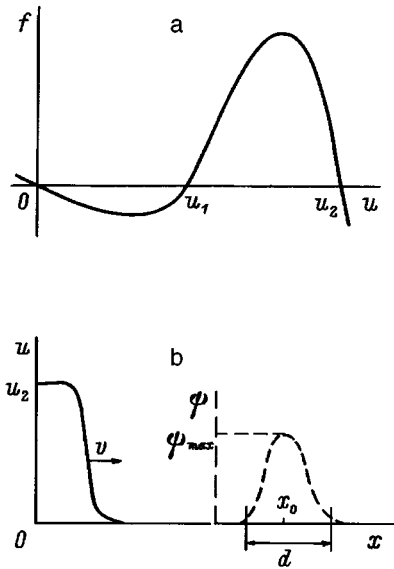
FIG. 1. Typical form of the function $f(u)$ (a) and diagram of wave propagation in a bistable medium with an isolated inhomogeneity (b).

For an arbitrary nonlinearity, an explicit solution of the steady-state wave equation (4) is unknown. The case where the function $F(u)$ is a third-degree polynomial has been studied in fairly great detail. The authors of Refs. 12 and 13 obtained a special two-phase solution of the reaction diffusion equation (2) with cubic nonlinearity which describes the profile of a steady-state wave as a kink in the long-time limit. A broad class of solutions of the steady-state equation (4), expressed in terms of the Weierstrass function, was obtained in Refs. 14 and 15.

For an arbitrary form of the function $F(\Theta)$ it is easy to obtain an asymptotic representation of the solution of Eqs. (4) and (5) for $\xi \to \pm\infty$, i.e., for any finite value of the coordinate x for $t \to \mp\infty$. In fact, let us assume $t \to +\infty$. Then, for a steady-state wave we have $u(x,t) \to u_2 = 1$ for any finite x, and $F(u) \simeq -\beta(u-1)$, where $\beta = -F'(u=1) > 0$. Thus, Eq. (4) becomes linear

$$\Theta''(\xi) + v\Theta'(\xi) - \beta(\Theta-1) \cong 0, \tag{6}$$

and its solution (with allowance for the condition of boundedness for $\xi \to -\infty$) has the form

$$\Theta(\xi) \cong 1 - \exp[\mu(\xi+\Delta)], \tag{7}$$

where $\mu = 0.5((v^2+4\beta)^{1/2}-v)$ and Δ is an arbitrary constant.

Similarly, for $t \to -\infty$, Eq. (4) gives

$$\Theta''(\xi) + v\Theta'(\xi) - \alpha\Theta \cong 0, \tag{8}$$

and its solution is

$$\Theta(\xi) \cong \exp[-\lambda(\xi+\Delta)], \tag{9}$$

where $\lambda = 0.5((v^2+4\alpha)^{1/2}+v)$, $\alpha = -F(u=u_0=0) > 0$, and Δ is an arbitrary constant generally different from that appearing in Eq. (7).

In an inhomogeneous medium the wave propagation is generally not steady-state. Here we confine our analysis to an ''isolated inhomogeneity'': the dependence $\alpha(x)$ has the form

$$\alpha(x) = \alpha_0(1+\Psi(x)), \tag{10}$$

where the function $\Psi(x)$ is finite. Its behavior will be characterized by two quantities: by the inhomogeneity amplitude Ψ_{max} and by the effective length d (see Fig. 1b).

In traveling-wave coordinates, Eq. (2) together with Eq. (10) becomes

$$\partial_t u - v\partial_\xi u = \partial^2_{\xi\xi}u + F(u) - \Psi(x+vt)u, \tag{11}$$

where v is the wave velocity in a corresponding homogeneous medium, i.e., for $\Psi(x) \equiv 0$.

We shall assume that the wave front is formed at a large distance from the inhomogeneous section. The initial condition for Eq. (11) is then

$$u(\xi, t \to -\infty) = \Theta(\xi), \tag{12}$$

where $\Theta(\xi)$ describes the wave in a homogeneous medium, i.e., is a solution of the steady-state equation (4).

The aim of this study is to investigate how the propagation of a wave is influenced by the presence of an isolated inhomogeneity (for various values of the parameters).

SMALL-AMPLITUDE INHOMOGENEITY

When the inhomogeneity amplitude Ψ_{max} is fairly small, the wave negotiates the inhomogeneous section and the wave front recovers with time (with increasing distance from the region of inhomogeneity). The presence of an inhomogeneity can then only lead to a shift in the position of the wave.

Thus, in this case

$$u(\xi, t \to +\infty) = \Theta(\xi+\delta), \tag{13}$$

where δ is the shift, which depends on the parameters of the problem.

Then, if $\Psi_{max} = \varepsilon \ll 1$, an explicit expression can be obtained for δ.

Using the notation $\Psi(\xi+vt) = \varepsilon\phi(\xi+vt)$ and substituting into Eq. (11), we seek a solution in the form of the steady-state wave perturbation:

$$u(\xi,t) = \Theta(\xi) + \varepsilon w(\xi,t) + 0(\varepsilon). \tag{14}$$

In the first order with respect to ε, the following equation is obtained for $w(\xi,t)$:

$$\partial_t w - v\partial_\xi w = \partial^2_{\xi\xi}w + F'[\Theta(\xi)]w - \Theta(\xi)\phi(x+vt) \tag{15}$$

with the boundary conditions

$$w(\xi \to \pm\infty, t) = 0, \quad w(\xi, t \to -\infty) = 0. \tag{16}$$

Using the results of Ref. 6, we expand $w(\xi,t)$ in terms of a complete system of eigenfunctions of the boundary-value problem for the equation describing small wave perturbations in a homogeneous medium,

$$w(\xi,t) = \sum_{i=0}^{\infty} G_i(t)\Phi_i(\xi), \tag{17}$$

where $\Phi_i(\xi)$ is the solution of the following problem:

$$\Phi_i''(\xi)+v\Phi_i'(\xi)+F'[\theta(\xi)]\Phi_i(\xi)=-\lambda_i\Phi_i(\xi), \quad (18)$$

$$\Phi_i(\xi\to\pm\infty)=0 \quad (i=0,1,2,\ldots).$$

It should be noted that the Sturm–Liouville problem (18) is not generally identical to that considered in Ref. 6. The difference is that in this case, $F(\Theta)$ is not positive definite so that in principle, nonmonotonic solutions of Eq. (4) may exist. However, it can be shown, as in Ref. 6, that the nonmonotonic solutions are not stable with respect to small perturbations and thus cannot be an intermediate asymptotic limit for any class of initial conditions. Confining our analysis to physically meaningful waves with a monotonic profile, the conclusions drawn in Ref. 6 concerning the properties of the eigenvalues and eigenfunctions can be applied to the problem (18) without being modified.

Substituting Eq. (17) into Eq. (15) with allowance for Eq. (18), we obtain

$$\sum_{i=0}^{\infty} G_i'(t)\Phi_i(\xi)=-\sum_{i=0}^{\infty}\lambda_i G_i(t)\Phi_i(\xi)-\Theta(\xi)\phi(\xi+vt).$$
$$(19)$$

Performing scalar multiplication of Eq. (19) by $\Phi_k(\xi)$, we obtain

$$G_k'(t)+\lambda_k G_k(t)+h_k(t)=0 \quad (20)$$

with the additional condition $G_k(t\to-\infty)=0$ $(k=0,1,2,\ldots)$, where

$$h_k(t)=||\Phi_k||^{-2}\int_{-\infty}^{\infty}\Theta(\xi)\phi(\xi+vt)\Phi_k(\xi)d\xi \quad (21)$$

and $||\Phi_k||^2=\int_{-\infty}^{\infty}\Phi_k^2(\xi)d\xi$.

The solution of Eq. (20) has the form

$$G_k(t)=-\int_{-\infty}^{\infty}h_k(\tau)\exp[-\lambda_k(t-\tau)]d\tau. \quad (22)$$

It is easy to establish that for any $k\geq1$, $G_k(t\to\infty)=0$. It was shown in Ref. 6 that the lowest eigenvalue of the problem (18) is $\lambda_0=0$ and $\lambda_k>0$ for $k\geq1$. Then, denoting by t_0 the time at which the wave propagates through the inhomogeneity and assuming that the function $\phi(\xi+vt)=\phi(x)$ is localized near a certain $x=x_0$, we can write the integral (22) as

$$G_k(t)=h_k(t_0)\exp[-\lambda_k(t-t_0)]\Delta t, \quad (23)$$

where Δt is the time taken by the wave to propagate through the inhomogeneity. Obviously, Δt is finite, $\Delta t\to(v/\Delta x)$ for $\varepsilon\to0$. This gives $G_k(t\to\infty)=0$ for all $k\geq1$.

Thus, Eqs. (14), (17), and (22) yield

$$u(\xi,t\to\infty)\cong\Theta(\xi)+\varepsilon g_0\Phi_0(\xi), \quad (24)$$

where

$$g_0=G_0(t\to\infty)=-\int_{-\infty}^{\infty}h_0(t)dt, \quad (25)$$

i.e., with allowance for Eq. (13)

$$\Theta(\xi+\delta)\cong\theta(\xi)+\varepsilon g_0\Theta'(\xi) \quad (26)$$

(since $\Phi_0(\xi)=\Theta'(\xi)$ (Ref. 6)); this gives $\delta=\varepsilon g_0$.

Then, substituting Eq. (21) into Eq. (25), changing the order of integration, and using the notation $B\equiv\int_{-\infty}^{\infty}\Psi(x)dx$ (the integration is actually performed over the interval in which the function Ψ has finite values), we have

$$g_0=(B/2\varepsilon v||\Phi_0||^2)[\Theta^2(-\infty)-\Theta^2(\infty)] \quad (27)$$

or with the boundary conditions (5)

$$\delta=(B/2)\left(v\int_{-\infty}^{\infty}[\Theta'(\xi)]^2d\xi\right)^{-1}. \quad (28)$$

Taking into account the form of Eq. (4), we can easily show that

$$v\int_{-\infty}^{\infty}[\Theta'(\xi)]^2d\xi=\int_0^1 F(\Theta)d\Theta\equiv M. \quad (29)$$

The final expression for the shift is

$$\delta=B/2M. \quad (30)$$

Note that formula (30) was derived without any special assumptions about the kinetics of the medium, i.e., as to the form of the function $f_1(u)$, and for an arbitrary inhomogeneity $\Psi(x)$.

WAVE LOCALIZATION: RESULTS OF A NUMERICAL SIMULATION

The results obtained above do not answer the most interesting question as to whether a wave can be localized at an isolated inhomogeneity. For small inhomogeneity amplitudes Ψ_{max}, no localization occurs. Expression (30) does not describe any critical phenomena: the wave shift increases monotonically but remains finite for any finite values of the inhomogeneity parameters.

However, the possibility of localization at isolated inhomogeneities of infinitely large amplitude was demonstrated in Ref. 11. Thus, different limiting values of Ψ_{max} correspond to different types of system dynamics. Consequently, it may be predicted that for each length d there exists a critical inhomogeneity amplitude Ψ_{cr}, such that for $\Psi_{max}>\Psi_{cr}$ the wave stops and for $\Psi_{max}<\Psi_{cr}$ the wave passes through.

These qualitative conclusions are supported by the results of computer simulations. We confine our analysis to the case of an inhomogeneity in the form of a rectangular barrier:

$$\Psi(x)=\Psi_{max}, \quad x_1<x<x_1+d,$$

$$\Psi(x)=0, \quad x<x_1 \text{ or } x>x_1+d \quad (31)$$

(for numerical results for other cases see Ref. 16).

The kinetics of the active medium (following Refs. 9 and 10) will be described by the cubic polynomial

$$F(u)=-u+au^2-bu^3. \quad (32)$$

Equation (2) together with Eqs. (10), (31), and (32) was solved by a finite-difference method over a wide range of parameters Ψ_{max} and d. It should be noted that a numerical
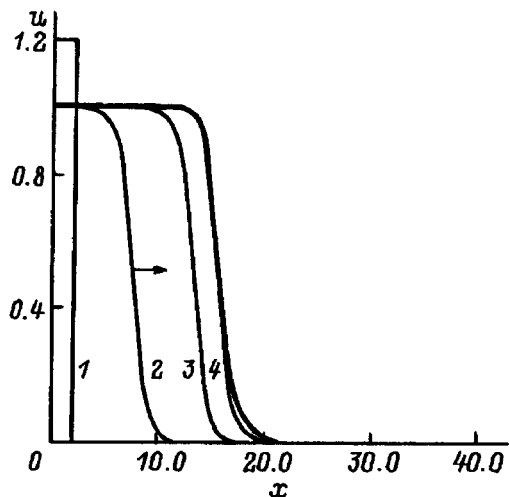
FIG. 2. Wave profiles (in dimensionless variables) calculated for successive equidistant times for supercritical values of the inhomogeneity parameters: *1* — initial condition, *2–4* — wave profiles for successive equidistant times.

FIG. 3. As Fig. 2 for subcritical values of the parameters — the wave propagates: *1* — initial condition, *2–7* — wave profiles for successive equidistant times.

solution of nonlinear partial differential equations may present appreciable difficulties (see Ref. 17). The results of the numerical calculations must be thoroughly checked. A convenient method is to check against an analytical solution of the problem (if one is known), such as the solution describing a steady-state wave.

The existence and uniqueness of the "classical" (i.e., continuous and bounded) solution of the problem (4), (5) for an arbitrary function $F(\Theta)$ describing a bistable medium was demonstrated in Ref. 7. An explicit analytical solution of the problem (4), (5) in the form of a kink for the case of cubic nonlinearity was first obtained in Refs. 12 and 13 (as the asymptotic form of the more general two-phase solution of the reaction diffusion equation (2) for long times). Note that the constraint for boundedness of the solution is important for its uniqueness. It was shown in Refs. 14 and 15 that if this constraint is removed, the nonlinear problem (4), (5) has a rich new class of analytical solutions expressed in terms of the Weierstrass function. These solutions, which generally have second-order discontinuities, describe physical systems possessing fairly exotic properties. Here we confine our analysis to the dynamics of continuous wave fronts having conventional physical applications. An analytical solution describing a two-phase wave[12,13] was used to check the results of the numerical calculations.

The initial condition was a finite Π-shaped distribution of width and amplitude sufficient for the formation of a wave.[18] The size of the calculation region was sufficiently large that the perturbation did not reach the boundary over the times studied. Figures 2 and 3 show wave profiles calculated for successive equidistant times (half the region is shown) for values of the parameters corresponding to different regimes. Figure 2 gives the profiles obtained for supercritical values of the parameters $d=2.0$, $\Psi_{max}=0.7$ (for $a=5$, $b=4$, corresponding to $u_2=1$, $M=0.167$); curve *1* gives the initial condition and the solid line corresponds to the asymptotically established steady-state distribution. Figure 3 gives the wave profiles for the subcritical parameters
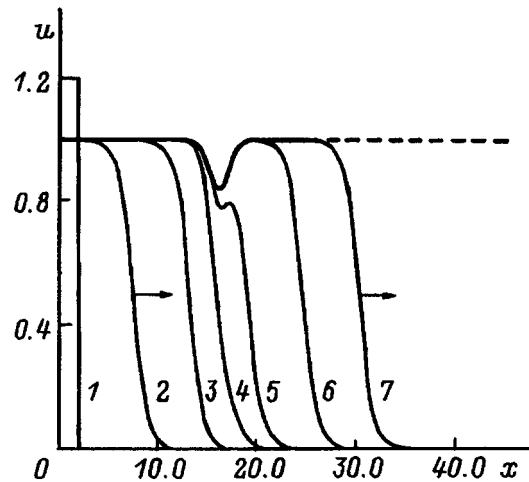
$d=2.0$, $\Psi_{max}=0.5$. The wave shift for small inhomogeneity amplitudes agrees well with formula (30), although it increases considerably faster as Ψ_{max} approaches the critical value.

By performing a fairly large number of computer simulations for various values of Ψ_{max} and d, it is possible to estimate the critical dependence $\Psi_{cr}(d)$. Figure 4 shows the pattern on the (Ψ_{max}, d) plane obtained from the results of the calculations. The curve $\Psi_{cr}(d)$ (solid curve) divides the plane of the inhomogeneity parameters into two regions. For values of the parameters from region *1* (i.e., for fairly small amplitude or width) the wave overcomes the inhomogeneity, whereas for values from region *2* the wave stops. The results of the numerical simulations show that a change in the parameters of the active medium only alters the position of the curve $\Psi_{cr}(d)$ on the plane (Ψ_{max}, d), while the nature of the critical dependence is almost independent of a and b. Note also that by applying known comparison theorems for solutions of nonlinear parabolic equations,[19] the monotonicity of the function $\Psi_{cr}(d)$ can be proven with mathematical rigor.
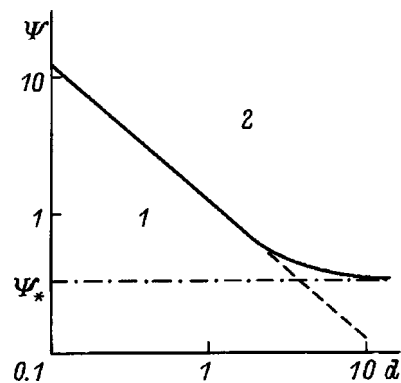

FIG. 4. Diagram on the length–amplitude plane for an isolated inhomogeneity. Solid curve — numerical simulation, dashed curve — asymptotic curve for short lengths (formula (49)), and dot-dash curve — asymptotic curve for long lengths (formula (34)).

In the above model the type of system dynamics (wave propagation or stopping) is predicted by calculating the dependence $\Psi_{cr}(d)$. The exact form of this dependence cannot be derived, although asymptotic formulas for $d \to 0$ and $d \to \infty$ may be obtained for the very general case of a bistable medium with arbitrary nonlinearity.

We begin with the case of an inhomogeneity in the form of a long rectangular barrier. The key to solving the problem can be found in the so-called ''limit theory'' of wave propagation in a homogeneous medium.[4] For a single-component bistable medium described by the single equation (2), this implies that the propagation of a wave switching the system from the lower homogeneous state, $u \equiv 0$, to the upper state, $u \equiv 1$, becomes impossible when $M \equiv \int_0^1 F(\Theta)d\Theta \leq 0$ (Refs. 7 and 10) (for $M < 0$ a ''reverse'' switching wave, switching the system from the upper to the lower state propagates in the system, but this will not be considered separately: this case reduces to the previous one upon the change variables $u \to 1 - u$, $x \to -x$).

Clearly, for a long barrier the conclusions of the limit theory become applicable to the inner region, and no switching wave is found in the region $x_1 < x < x_1 + d$ if

$$M_{in} = \int_0^1 [-(1 + \Psi_{max})\Theta + (1/\alpha_0)f_1(\Theta)]d\Theta \leq 0, \quad (33)$$

i.e., for

$$\Psi_{max} > \Psi_* = 2M. \quad (34)$$

The asymptotic curve $\Psi_{cr}(d \to \infty) = \Psi_*$ is shown by the dot-dash line in Fig. 4. Expression (34) thus shows excellent agreement with the results of the numerical analysis for lengths $d \gtrsim 10$.

Let us now consider the case $d \to 0$. For arbitrary finite Ψ_{max} the inhomogeneity clearly has a weaker influence on the wave propagation for smaller lengths d. In particular, distortions of the wave front as it propagates through the inhomogeneity may be neglected for fairly small d. Thus to a first approximation we shall assume that the wave profile is unchanged and the influence of the inhomogeneity then reduces to a gradual shift of the wave position with time relative to the position it would have occupied in a corresponding homogeneous medium. This implies that the solution of the time-dependent equation (11) may be sought in the form

$$u(\xi, t) = \Theta(\xi; \Delta(t)), \quad (35)$$

where $\Theta(\xi)$ is the solution of Eq. (4) for a steady-state wave, $\Delta(t)$ is the shift, and $\Delta(t \to -\infty) = 0$.

The time-independent equation (4) is invariant with respect to the shift, so that $\Theta(\xi; \Delta(t)) = \Theta(\xi + \Delta(t))$. Hence, substituting Eq. (35) into Eq. (11), we obtain

$$[\Delta'(t) - v]\partial_\xi \Theta = \partial_{\xi\xi}^2 \Theta + F(\Theta) - \Psi(\xi + vt)\Theta. \quad (36)$$

Multiplying Eq. (36) by $\partial_\xi \Theta$ and integrating over ξ between $-\infty$ and ∞, we obtain an equation for the function $\Delta(t)$:

$$\Delta'(t) = -(v/M) \int_{-\infty}^{\infty} \Psi(\xi + vt)\Theta(\xi + \Delta(t))$$

$$\partial_\xi \Theta(\xi + \Delta(t))d\xi \quad (37)$$

(the convergence of the integral is ensured by the finiteness of the function $\Psi(x)$).

The idea of the subsequent analysis is as follows. Solving Eq. (37) in the long-time limit gives the total shift $\Delta(\infty) = \delta$, which, as has been shown above, increases with increasing inhomogeneity parameters Ψ_{max} and d. If the total shift becomes infinite for a certain ratio between the parameters, this implies (within the limits of this approach) that the wave is localized.

The solution of the steady-state equation (4) for an arbitrary type of nonlinearity $F(\Theta)$ is unknown, so that it is also impossible to solve Eq. (37). However, the behavior of the solution for $t \to +\infty$ is of particular interest. In this case (i.e., for any finite x for $\xi \to -\infty$), an asymptotic representation (7) is found for the steady-state wave. Substituting Eq. (7) into Eq. (37) yields the following equation for $\Delta(t)$:

$$\Delta'(t) \simeq (v\mu/M)(L \exp[\mu(\Delta - vt)]$$

$$- N \exp[2\mu(\Delta - vt)]), \quad (38)$$

where $L = \int_0^d \Psi(x)\exp(\mu x)dx$, $N = \int_0^d \Psi(x)\exp(2\mu x)dx$ (assuming, without loss of generality, that $x_1 = 0$).

Since $\exp(\mu\xi)$ is small for $\xi \to -\infty$, we confine ourselves to linear terms with respect to $\exp(\mu\xi)$ in Eq. (38), which then gives

$$\Delta'(t) \simeq (v\mu L/M)\exp[\mu(\Delta - vt)]. \quad (39)$$

Equation (39) can easily be solved

$$\Delta(t) \simeq -(1/\mu)\ln(C + (\mu L/M)\exp(-\mu vt)). \quad (40)$$

The integration constant C in Eq. (40) cannot be determined directly, since the ''initial'' condition has the form $\Delta(-\infty) = 0$ whereas the asymptotic representation (39) can only be considered to be approximately valid from a certain t_0. The value of C can be determined from the matching conditions of the asymptotic representations of the solution for $t \to \infty$ and $t \to -\infty$.

The asymptotic behavior of the solution of Eq. (37) for $t \to -\infty$ is determined by the behavior of the solution of the steady-state equation (4) for $\xi \to \infty$. Substituting Eq. (9) into Eq. (37) gives

$$\Delta'_{(-)}(t) \simeq (v\lambda K/M)\exp[-2\lambda(\Delta - vt)], \quad (41)$$

where $K = \int_0^d \Psi(x)\exp(-2\lambda x)dx$, which gives

$$\Delta_{(-)}(t) \simeq (1/2\lambda)\ln(1 + (\lambda K/M)\exp(2\lambda vt)). \quad (42)$$

However, the simplest condition $\Delta_{(-)}(t_0) = \Delta(t_0)$ is not satisfactory since the times for which the interaction between the wave and the inhomogeneity is strongest are then neglected. The contribution of these times can be described by introducing the effective shift parameter Δ_0. The matching condition has the form

$$\Delta_{(-)}(0) + \Delta_0 = \Delta(0) \quad (43)$$

(assuming $t_0 = 0$). Using Eqs. (40), (42), and (43), we obtain the integration constant

$$C = \exp(-\mu\Delta_0)[1 + (\lambda K/M)]^{-\mu/2\lambda} - (\mu L/M). \quad (44)$$

It is easy to see from Eq. (40) that the total shift becomes infinite when $C = 0$. Thus the inhomogeneity becomes impassable for a wave with the following relation between the parameters:

$$(\mu L/M)[1 + (\lambda K/M)]^{\mu/2\lambda} = A, \qquad (45)$$

where $A = \exp(-\mu\Delta_0)$, and the values of K, L, and N are the integral characteristics of the function $\Psi(x)$.

We stress that expression (45) was obtained for a bistable medium with an arbitrary nonlinearity and an arbitrary isolated inhomogeneity. For each specific form, $\Psi(x)$ belongs to a certain two-parameter family $\Psi = \Psi(x; \Psi_{max}, d)$. In this case, the values of K, L, and M are functions of the inhomogeneity amplitude Ψ_{max} and the length d, and expression (45) describes the required relation between the critical inhomogeneity parameters.

Relation (45) may be simplified by taking into account the smallness of d proposed above. Let us assume that $d \ll \min\{1/\mu, 1/\lambda\}$, in which case we obtain, to first order in d,

$$L \cong K \cong \int_0^d \Psi(x)dx \equiv B = B(\Psi_{max}, d), \qquad (46)$$

and Eq. (45) becomes

$$(\mu B/M)[1 + (\lambda B/M)]^{\mu/2\lambda} = A. \qquad (47)$$

Relation (47) has the simplest form for a bistable medium with cubic nonlinearity. For this case, using the results of Ref. 20, it is easily shown that $\lambda = \mu$ and Eq. (47) gives

$$(\mu B/M)[1 + (\mu B/M)]^{1/2} = A, \qquad (48)$$

where $\mu = (b/2)^{1/2}$.

Expression (48) is a nonlinear algebraic equation for the variable $z = \mu B/M$. Denoting its solutions as z_0, we obtain a simple relation between the parameters of a critical-scale inhomogeneity

$$B(\Psi_{max}, d) = z_0 M/\mu. \qquad (49)$$

The normalization coefficient z_0 is not determined in this approach. It may be determined by means of a comparison with the results of a computer simulation.

For a rectangular barrier inhomogeneity we have $B = \Psi_{max}d$. The critical curve $\Psi_{cr}(d)$ calculated for this case in accordance with Eq. (49) (dashed curve in Fig. 4) demonstrates almost ideal matching with the results of the numerical simulation up to lengths $d \cong 2$.

Thus, the asymptotic expressions (34) and (49) satisfactorily describe the dependence $\Psi_{cr}(d)$ for any lengths d. The value Ψ_* is an exact lower bound on the critical values of the amplitude (with allowance for the monotonic behavior of the function $\Psi_{cr}(d)$).

CONCLUSIONS

Information on the wave dynamics of an active medium with an isolated inhomogeneity, obtained only for one case of a specific type of inhomogeneity (for example, using relations (34) or (49)) may be used to predict the propagation of a wave in a medium with an arbitrary type of inhomogeneity when the inhomogeneity $\Psi_0(x)$ is supercritical. Then, using the comparison theorems for solutions of nonlinear parabolic equations,[19] it is easily shown that any inhomogeneity $\Psi_1(x)$ such that $\Psi_1(x) \geqslant \Psi_0(x)$ for any x is supercritical. Conversely if the inhomogeneity $\Psi_0(x)$ is subcritical, any inhomogeneity $\Psi_2(x)$ such that $\Psi_2(x) \leqslant \Psi_0(x)$ is also subcritical. In particular, this implies that in a given active medium the coefficient z_0 in relation (49) (or in a similar relation obtained from the more general expression (47)) need be determined only once. The dependence (49) will then serve as a lower bound for supercritical inhomogeneities and as an upper bound for subcritical inhomogeneities.

To conclude, we note that the assumption made at the beginning of this study that the function $\Psi(x)$ is finite is not a fundamental constraint. It is easy to see that a necessary condition is that $\Psi(x)$ should tend fairly rapidly to zero for $x \to \pm\infty$, which ensures convergence of the appropriate integrals.

[1] A. N. Kolmogorov, I. G. Petrovskiǐ, and N. S. Piskunov, Byull. Mosk. Gos. Univ. Ser. Mat. Mekh. **1**(6), 1 (1937).
[2] R. Fisher, Ann. Eugenics **7**, 355 (1937).
[3] Ya. B. Zel'dovich and D. A. Frank-Kamenetskiǐ, Zh. Fiz. Khim. **12**, 100 (1938).
[4] Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **11**, 159 (1941).
[5] Ya. B. Zel'dovich, Zh. Fiz. Khim. **22**, 27 (1948).
[6] F. I. Barenblatt and Ya. B. Zel'dovich, Prikl. Mat. Mekh. **21**, 856 (1957).
[7] Y. I. Kanel', Dokl. Akad. Nauk SSSR **132**, 268 (1960) [sic].
[8] Ya. B. Zel'dovich, F. I. Barenblatt, V. B. Libovich, and G. M. Makhviladze, *Mathematical Theory of Combustion and Explosion* [in Russian], Nauka, Moscow (1980).
[9] V. P. Maslov, V. G. Danilov, and K. A. Volosov, *Mathematical Modeling of Heat and Mass Transfer Processes* [in Russian], Nauka, Moscow (1987).
[10] A. Yu. Loskutov and A. S. Mikhaǐlov, *Introduction to Synergetics* [in Russian], Nauka, Moscow (1990).
[11] A. V. Gurevich and R. G. Mints, Usp. Fiz. Nauk. **142**, 61 (1984) [Sov. Phys. Usp. **27**, 19 (1984)].
[12] T. Kawahara and M. Tanaka, Phys. Lett. A **97**, 311 (1983).
[13] V. G. Danilov and P. Yu. Subochev, Teor. Mat. Fiz. **89**, 25 (1991).
[14] A. M. Samsonov, *Future Directions of Nonlinear Dynamics in Physical and Biological Systems*, edited by P. L. Christiansen *et al.* (Plenum Press, New York, 1993).
[15] A. M. Samsonov, Applicable Analysis **57**, 85 (1995).
[16] M. E. Vinogradov, G. I. Barenblatt, and S. V. Petrovskiǐ, Okeanologiya **35**(2), 98 (1995).
[17] G. I. Barenblatt, M. Bertsch, R. Dal Passo *et al.*, J. Fluid. Mech. **253**, 341 (1993).
[18] S. V. Petrovskiǐ, Inzh.–Fiz. Zh. **66**, 398 (1994).
[19] A. I. Vol'pert and S. I. Khudyaev, *Analysis in Classes of Discontinuous Functions and the Equations of Mathematical Physics* [in Russian], Nauka, Moscow (1975).
[20] M. V. Ognev, S. V. Petrovskiǐ, and V. M. Prostokishin, Zh. Tekh. Fiz. **65**(6), 1 (1995) [Tech. Phys. **40**, 521 (1995)].

Translated by R. M. Durham

Jet engine electrification as a phenomenon reflecting the evolution of charge instability in an outflowing combustion product mixture

V. A. Pinchuk

Baltic State Technical University, 198005 St. Petersburg, Russia

It is hypothesized that charge instabilities developing in a mixture of outflowing combustion products make some contribution to the mechanism of engine electrification in jet engines. Investigations are carried out which support this hypothesis. © *1997 American Institute of Physics.* [S1063-7842(97)00308-5]

INTRODUCTION

The electrification of jet engines is a phenomenon that is still in need of a physical interpretation. The most frequently advanced concepts of engine electrification as a process essentially caused by the removal of electric charge from the chambers in the form of electrically charged volumes of gas from regions adjacent to the walls in the outflow of weakly ionized combustion products into the surrounding space[1,2] have difficulty accounting for (and may even be inconsistent with) many of the general characteristics observed in experiment.

In particular, this model is not supported by investigations of the downstream distribution of the charge removed from the chamber of a jet engine: the observed distributions of the exhaust currents do not reveal any preferential concentration of charge in the peripheral layers and correspond to the cross-sectional distributions of electrical charge in jets with a unipolar ionic charge.[3]

The general features of the identified dependences of the polarity of the charge extracted from the chambers on the engine operating conditions are difficult to explain in terms of these concepts. For instance, experimental investigations of engine electrification have convincingly demonstrated that high-temperature operation generally leads to the removal of positive charge from the chambers. At the same time, in low-temperature operation (generator modes) it is negative charge that is most often found to be expelled from the chambers.[2–4]

Another approach has been developed in which the phenomenon of engine electrification is attributed to the charged particles being ''frozen'' in the combustion product mixture, i.e., as being due to their low mobility, on the assumption that the charge carriers in the chamber are predominantly small carbon-containing particles of unburned fuel.[3] However, engine electrification as a phenomenon is characteristic of a whole range of jet engines, including some not using carbon-containing substances as fuel (for example, those using hydrogen fuel). Moreover, the experimental observations of engine electrification do not reveal any fundamental changes in its behavior related to the presence or absence of carbon-containing components in the fuel.[4]

In other words, a comprehensive explanation of the observed characteristics of engine electrification has not yet been formulated. The combined results of attempts to obtain a picture of engine electrification indicate that the conventionally used models of these characteristics are fundamentally incapable of discriminating between alternative choices as to the physical causes of engine electrification.

The overall situation clearly suggests that analysis is limited to the processes taking place in the volumes of combustion products adjacent to the chamber walls and to the removal of electricity from the chambers by barely mobile carbon-containing particles of unburned fuel, it is difficult (if not impossible) to put forward scientifically substantiated concepts of engine electrification.

Without dismissing the contribution of the factors conventionally taken into account in the mechanism of engine electrification, it is suggested in the present paper that models of engine electrification may be constructed to include phenomena reflecting the evolution of charge instabilities in a weakly ionized mixture of combustion products flowing out of the combustion chambers.[5]

MODEL MECHANISM

We start from the fact that the known experimental data on engine electrification suggest that there is some mechanism which not only leads to a contribution to engine electrification from the entire stream of combustion products flowing out of the chambers (and not only from the peripheral layers adjacent to the chamber walls) but also causes the electrification process to depend on the operating characteristics in a manner that cannot be explained by the conventional reasoning.

In this context, it is assumed that, under the conditions prevailing in the chambers of jet engines, the combustion products form predominantly a weakly ionized gaseous medium which is subjected to an intense acoustic perturbation as it flows out into the surrounding space. In view of Ref. 5, it is natural to expect that the outflow of combustion products from the chambers into the surrounding space will be accompanied by the evolution of charge instabilities in the medium. The characteristics of the expected manifestations of these charge instabilities in the form of erosion currents arising in the stream could be of particular interest in this regard.

Let us construct a system to describe the outflow of combustion products. The description is based on the concepts of the three-fluid model.[5] Note that among other advantages,

this model eliminates the need to make a detailed estimate of the mixture composition: both the individual characteristics of a composition and the possibility that they may differ substantially can be taken into account by specifically determining or, alternatively, by varying over a wide range, one of the important characteristics of ionized media — the effective ionization potential.

Adopting the notation of Ref. 5, assuming that the problem is steady-state, and using a one-dimensional representation, we determine the composition of the system using the equations of motion for the electron component

$$m_e n_e U_e \frac{\partial U_e}{\partial x} + \frac{e n_e}{\mu_e}(U_e - U_a) + \frac{\partial p_e}{\partial x} - e n_e E = 0, \quad (1)$$

and for the ion component

$$m_i n_i U_i \frac{\partial U_i}{\partial x} + \frac{e n_i}{\mu_i}(U_i - U_a) + \frac{\partial p_i}{\partial x} - e n_i E = 0 \quad (2)$$

the Saha ionization equilibrium condition for a three-fluid medium of arbitrary composition, expressed in the form

$$\frac{\alpha \eta}{(1-\alpha)(1+\eta)} - 6.666798_{10^{-2}} \frac{T^{5/2}}{p} \exp\left(-\frac{eV}{kT}\right) = 0 \quad (3)$$

expressions for the mass-averaged density

$$\rho_\Sigma - \frac{m_e \eta + m_i \alpha + m_a(1-\alpha)}{1+\eta} \frac{p}{kT} = 0, \quad (4)$$

and velocity

$$U_\Sigma \left[\eta + \frac{m_i}{m_e}\alpha + \frac{m_a}{m_e}(1-\alpha) \right] - U_e \eta - \frac{m_i}{m_e} U_i \alpha$$

$$- \frac{m_a}{m_e} U_a (1-\alpha) = 0 \quad (5)$$

of a three-fluid medium, the equation of flux continuity

$$\rho_\Sigma \bar{F} \frac{\partial U_\Sigma}{\partial x} + \rho_\Sigma U_\Sigma \frac{\partial \bar{F}}{\partial x} + U_\Sigma \bar{F} \frac{\partial \rho_\Sigma}{\partial x} = 0 \quad (6)$$

the electric charge conservation law

$$U_\Sigma \bar{F} \frac{\partial q}{\partial x} + j_k \frac{\partial \bar{F}}{\partial x} + \bar{F} \frac{\partial j_k}{\partial x} = 0, \quad (7)$$

and the Poisson equation for the electric component of the field

$$\frac{\partial E}{\partial x} - \frac{e}{\varepsilon_0} \frac{\alpha - \eta}{1+\eta} \frac{p}{kT} = 0. \quad (8)$$

Using the relations between the parameters, and also the expressions for the mobilities μ_e and μ_i of the components, and the densities of the excess charge q and the convective current j_k (Ref. 5), the system is reduced to eight linear, independent equations for eleven unknowns: U_e, U_i, U_a, ρ_Σ, U_Σ, α, η, E, p, T, and $\bar{F}$.

Equations (1)–(8) are then supplemented by an equation linking the pressure and temperature of the medium for isentropic flow

$$T = T_0 \left(\frac{p}{P_0} \right)^{\frac{n-1}{n}}, \quad (9)$$

and also by a relation for the expansion of the nozzle as a function of the relative pressure in the stream and the characteristic of the process

$$\bar{F} = \frac{\left(\frac{2}{n+1} \right)^{\frac{1}{n-1}} \left(\frac{n-1}{n+1} \right)^{0.5}}{\left[\left(\frac{p}{P_0} \right)^{2/n} - \left(\frac{p}{P_0} \right)^{\frac{n+1}{n}} \right]^{0.5}}. \quad (10)$$

To close the system and to ensure the possibility of using the formalism developed in Ref. 5 to obtain quantitative estimates for the stated problem, we adopt the following form for the characteristic pressure distribution over the length of the supersonic part of the nozzle of a jet engine:

$$p = p* \left(\frac{p_a}{p*} \right)^{\left(\frac{x}{L_a} \right)}. \quad (11)$$

Here $x \in [0, L_a]$ is the linear coordinate,

$$p* = P_0 \left(\frac{2}{n+1} \right)^{\frac{n}{n-1}}$$

is the pressure at the nozzle throat. The system (1)–(11) was thus used to obtain an electrogasdynamic description of the outflow of combustion products from the chamber of a jet engine. The electrical coupling between the stream and nozzle walls is neglected.

NATURE OF THE INVESTIGATION AND RESULTS

The system (1)–(11) was used to study the characteristics of the supersonic part of the stream during the outflow of combustion products from the chambers. It was assumed that the electrophysical parameters of the combustion product stream in the throat cross section of the nozzle are unperturbed, i.e., they are consistent with the general conditions of quasineutrality and are characterized by the absence of any internal electric fields.

With allowance for these physical concepts, a difference scheme was formulated using the initial integration conditions

$$x = 0: \quad U_e = U_i = U_a = U_\Sigma = [nRT_0/(n+1)]^{0.5} = W* = a*,$$

$$\alpha = \eta = f(T*, p*, V), \quad E = 0, \quad \rho_\Sigma = \rho*, \quad p = p*,$$

$$T = T*, \quad \bar{F} = 1.$$

The problem was thus reduced to numerical integration of the system (1)–(11) in the range $x \in [0, L_a]$, where L_a is the length of the supersonic part of the nozzle. At each step during the integration process, a check was made to ensure that the condition mod $E < E_{bd}$ was satisfied, where E_{bd} is the breakdown strength of the electric field (arbitrarily taken to be 10^7 V/m). If this condition was not satisfied, the integration was terminated.

As in Ref. 5, an estimate was made of the complex characteristics such as the exhaust current density j_k, the local electric charge density q caused by the charge instability, and the degree of degradation of the quasineutrality of the medium ξ. The specific conditions for the calculation variant were determined by defining the pressure P_0 in the chamber, the pressure p_a of the combustion products in the nozzle edge cross section, the temperature T_0 of the combustion products in the chamber, the outflow characteristic n, the ionization potential V for the conditions of a three-fluid medium, and the nozzle length L_a. The input parameters for the calculations were varied widely: $P_0 \in (0.3, 5)$ MPa, $p_a \in (0.05, 0.5)$ MPa, $T_0 \in (1500, 3500)$ K, $n \in (1.1, 1.4)$, $L_a \in (0.5, 1.5)$ m, and $V \in (6, 16)$ V. However, it was ensured that when selecting P_0 and p_a, the condition $P_0/p_a \leq 10$ was satisfied. At present there is no commonly held, definitive view as to how far the composition of the combustion product mixture flowing out of the chambers of real jet engines corresponds to the ionization equilibrium conditions as given by Eq. (3). Thus, the calculations using this technique were duplicated with calculations in which the degree of ionization over the nozzle length was considered to be ''frozen'' (other conditions being equal, the Saha equation was only used in these cases when formulating the initial conditions for the numerical analysis; this condition was subsequently dropped when integrating the system).

The general results of the numerical analysis justify the validity of considering engine electrification as a phenomenon which reflects the evolution of instabilities in the bulk distribution of electric charge in a weakly ionized mixture of combustion products flowing out of the chambers.

In fact, the solutions reveal that over the entire range of input parameters, nonzero exhaust currents are necessarily formed in the streams of combustion products flowing out of the nozzle, and these give rise to macroscopic formations with excess electric charge (but quasineutral) and internal electric fields.

In general, the results indicate that either positive or negative electric charge may be carried out of the chambers as a result of the evolution of charge instabilities in outflowing combustion product streams. Other conditions being equal, the probability of the extraction of positive charge increases with decreasing ionization potential or increasing mixture temperature, while the probability of the extraction of negative charge increases with increasing ionization potential or decreasing temperature. The general nature of the results stays the same regardless of the calculation variant (equilibrium or frozen outflow in terms of degree of ionization).

Figure 1 shows specific distributions of the exhaust current density along the nozzle length, obtained from these solutions. These curves also show how the polarity of the electric charge carried out of the chamber with the combustion product stream depends on the composition of the outflowing medium, as was noted above. Figure 2 gives the extracted current density as a function of the temperature of the combustion products, keeping other conditions equal.

Although these curves are plotted for ionization equilibrium outflow, they are nevertheless typical, and as a whole
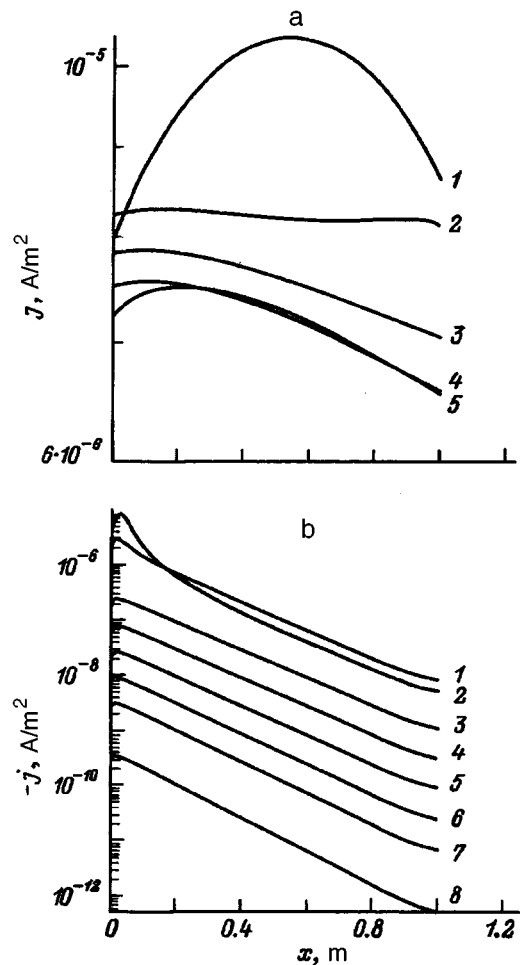

FIG. 1. Current density distribution over nozzle length as a function of mixture composition. $P_0 = 0.4$ MPa, $T_0 = 3000$ K, $p_a = 0.1$ MPa, $L_a = 1.0$ m, $n = 1.3$. V in volts for a: 1 — 8.5, 2 — 10.5, 3 — 10.0, 4 — 9.5, and 5 — 9; for b: 1 — 13.00, 2 — 12.75, 3 — 14.00, 4 — 14.50, 5 — 15.00, 6 — 15.50, 7 — 16.00, and 8 — 17.00.

reflect the general features of the results. In the range of T_0 and V levels covered by these investigations, three characteristic regimes of combustion product outflow from the chambers of jet engines can be identified in particular. The first regime is characterized by stable extraction of positive electric charge from the chambers along the nozzle length. This regime is established when high-temperature (or easily ionized) combustion product mixtures flow out of the chamber and may be arbitrarily assigned to the ''chamber'' group (Figs. 1a and 2a).

The second characteristic group of regimes exhibits stable extraction of negative charge along the nozzle length and is established when low-temperature (barely ionizable) media flow out into the surrounding space. These regimes may be arbitrarily assigned to the ''generator'' group (Figs. 1b and 2b).

Finally, the third, transition, group of regimes exhibits some uncertainty as to the sign of the charge carried out of the chamber with the combustion product stream. The distributions of the flow parameters over the nozzle length, which determine the evolution of charge instabilities in the stream and the local electrophysical properties of the combustion
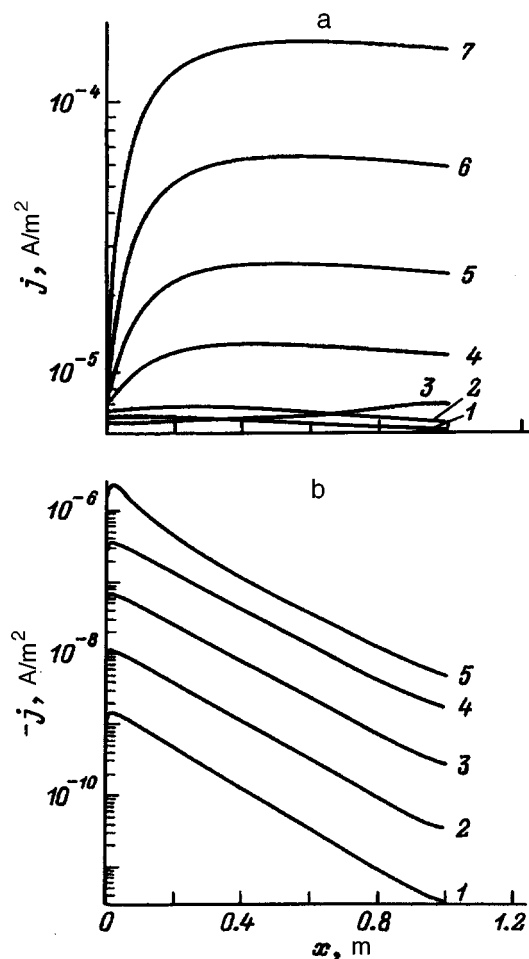
FIG. 2. Current density distribution over nozzle length as a function of temperature in chamber. $P_0 = 0.4$ MPa, $V = 9$ V, $p_a = 0.1$ MPa, $L_a = 1.0$ m, $n = 1.3$, T, K for a: 1 — 2800, 2 — 3000, 3 — 2500, 4 — 3200, 5 — 3300, 6 — 3400, 7 — 3500; for b: 1 — 1700, 2 — 1800, 3 — 1900, 4 — 2000, and 5 — 2100.

products as a whole, have a nonunique character for the various realizations of this group of regimes, which fully reflects the complex mechanism (via the evolution of well-defined charge instability in the stream) of reversal of the charge extracted from the chamber during the outflow process (Fig. 3).

In other words, as in Ref. 5, it is observed that in the region where the sign of the exhaust currents reverses (see, for example, $j = f(V)|T_0 = \text{const}$ in Fig. 4, the dependence $j = f(T_0)|V = \text{const}$ is similar), the electrophysical characteristics of the charge instabilities estimated by the solutions not only increase abruptly in absolute value, but their directions become highly indeterminate (under these conditions, the charge instability may described as "catastrophic"). Thus, the observed pattern reflects the general behavior of instabilities in bulk distribution of electric charge,[5] which develops in weakly ionized media as a result of parametric transience (which, in this particular case, is caused by the outflow).

These results support the initial assumption that charge may be carried out of the chambers as a result of the evolution of charge instabilities in the outflowing combustion product mixture, and therefore confirm their contribution to

the overall mechanism of engine electrification (at least among those conventionally taken into account by known models).

Confirmation that charge instabilities formed in the outflowing combustion product mixture contribute to the mechanism of engine (i.e., caused directly by the operation of the engine) electrification is also provided by comparing the general characteristics of the experimental observations of engine electrification with predictions from the present results.

Although there are generally very few known experimental studies of engine electrification, it should nevertheless be acknowledged that specialists at the Kazan School of Science (V. E. Alemasov and A. F. Dregalin) have produced extremely reliable results in this field taking into account an extensive, complex range of factors, some of which are reported in Refs. 2, 4, and 6. We note the general features of these results.

Unique experimental results of investigations of the characteristics of engine electrification obtained in tests on liquid-fueled jet engines utilizing different fuel mixtures, carried out using a specially equipped test rig, were reported in Ref. 4. In particular, it was established in Ref. 4 that there are three characteristic regimes of engine electrification. The first is achieved when the engine becomes negatively charged as a result of positive charge being carried out of the chamber with the outgoing stream. This regime is mainly typical of high-temperature operation (determined by fuel component ratios in the chamber being approximately stoichiometric). The engine electrification characteristics under these conditions are stable. During operation the engine acquires and stably retains a negative potential with respect to ground.

Conversely, low-temperature (generator) operation, according to Ref. 4, is characterized by the removal of negative charge from the chamber. The engine then becomes positively charged. It acquires and stably retains a positive potential with respect to ground.

Finally, the existence of a regime of unstable potential formation is identified, where the sign of the charge (potential) extracted with the stream changes. It is observed that the absolute levels of the electrophysical characteristics of engine electrification in this region are many times greater than those of the inherent stable regimes and also vary in sign.

These general characteristics of engine electrification therefore show complete qualitative agreement with the present results (Figs. 1–4). The more specific results also agree with the known observations.

It was observed in Ref. 6 that the region of unstable regimes becomes larger as the expansion ratio of the outflowing medium increases. This is also confirmed by the results of the present numerical calculations. Agreement may be noted between the results of the observations[2,3] and the extracted charge distribution over the stream cross section taken into account initially. In other words, comparisons reveal a qualitative analogy between the nature of the predicted engine electrification assuming that charge instabilities make a decisive contribution to the mechanism, and the generalized results of the experimental observations.
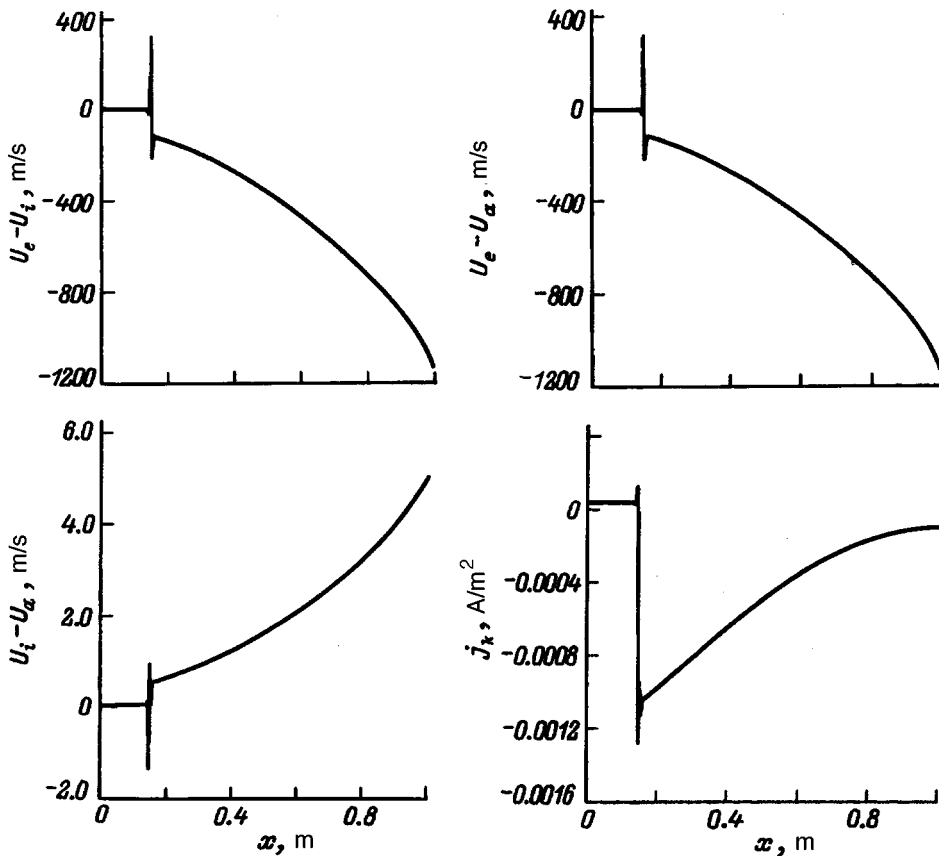
FIG. 3. Distributions $U_e - U_i = f(x)$, $U_e - U_a = f(x)$, $U_i - U_a = f(x)$, and $j_k = f(x)$ over nozzle length (respectively). $P_0 = 0.4$ MPa, $p_a = 0.1$ MPa, $T_0 = 3000$ K, $V = 12.1$ V, $L_a = 1.0$ m, and $n = 1.3$.

Thus, the results as a whole support the assumption that the phenomenon of motive electrification reflects the evolution of instabilities in the bulk electric charge distribution of the combustion products flowing out of the chambers of jet engines.

CONCLUSIONS

Theoretical analyses using the three-fluid model have confirmed that engine electrification in jet engines can be accurately treated as an effect reflecting the evolution of instabilities in the bulk electric charge distribution of the combustion products flowing out of the chambers. Further theoretical and, particularly, experimental investigations are needed to determine in greater detail the importance of the relative contribution of charge instabilities to the mechanism of engine electrification and to obtain a generally deeper understanding of the concepts of engine electrification within the proposed model. In particular, the mechanism for limitation of the charge accumulated at the casing of jet engines as a result of engine electrification requires detailed study under flight conditions.

The author would like to thank G. E. Skvortsov for helpful discussions and pertinent observations which were of particular assistance to the author in the final preparation of this manuscript for publication.
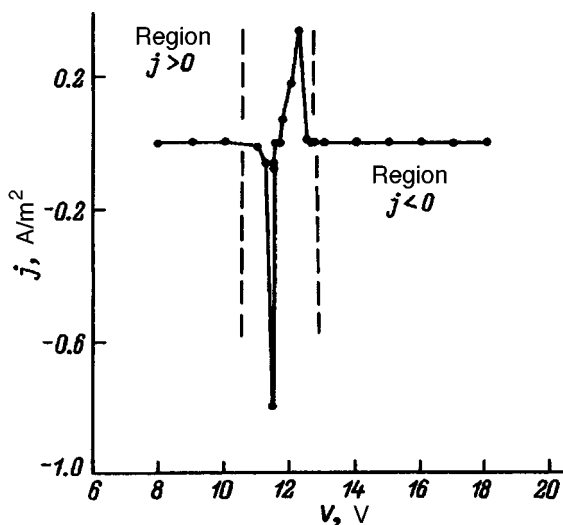

FIG. 4. Distribution $j_k = f(V)$ in nozzle edge cross section. $P_0 = 0.4$ MPa, $p_a = 0.1$ MPa, $T_0 = 3000$ K, $L_a = 1.0$ m, and $n = 1.3$.

[1] L. Aronowitz, IEEE Trans. Electromagn. Compat. **EMC-10**, 341 (1968).
[2] G. P. Potapov, Izv. Vyssh. Uchebn. Zaved. Aviats. Tekh. No. 4, 112 (1978).
[3] A. B. Vatazhin, V. I. Grabovskiǐ, V. A. Likhter, and V. I. Shul'gin, *Electrogasdynamic Flows* [in Russian], Nauka, Moscow (1983).
[4] V. A. Shkalikov, *Thermal Processes and Properties of the Working Parts of Aircraft Engines*, edited by V. E. Alsamasov [in Russian], Kai, Kazan (1980), pp. 41–47.
[5] V. A. Pinchuk, Inzh.-Fiz. Zh. **67**, 112 (1994).
[6] V. A. Shkalikov, A. A. Lebedev, and Zh. K. Kul'zhanov, *Thermal Processes and Properties of the Working Parts of Aircraft Engines*, edited by V. E. Alsamasov [in Russian], Kai, Kazan (1980), pp. 79–83.

Translated by R. M. Durham

Capillary oscillations and Tonks–Frenkel instability of a liquid layer of finite thickness

A. I. Grigor'ev, S. O. Shiryaeva, V. A. Koromyslov, and D. F. Belonozhko

Yaroslavl State University, 150000 Yaroslavl, Russia

A dispersion relation is derived and analyzed for the spectrum of capillary motion at a charged flat surface of viscous liquid covering a solid substrate with a layer of finite thickness. It is shown that for waves whose wavelengths are comparable with the layer thickness, viscous damping at the solid bottom begins to play an important role. The spectrum of capillary liquid motion established in this system has high and low wave number limits. The damping rates of the capillary liquid motion with wave lengths comparable with the layer thickness are increased considerably and the Tonks–Frenkel instability growth rates are reduced compared with those for a liquid of infinite depth. © *1997 American Institute of Physics.*
[S1063-7842(97)00408-X]

INTRODUCTION

The instability of a charged liquid surface (Tonks–Frenkel instability) of finite depth must be considered in various scientific, technical, and technological applications. Such applications include the stability of a water layer on the surface of an ice core (melting hailstone) in a thundercloud or in a vacuum type of liquid mass spectrometer, and the stability of a layer of liquid metal in liquid-metal ion sources, where electrospraying takes place from the lateral surface of the emitter needle along which liquid metal is supplied (see Ref. 1). It is therefore of interest to calculate the capillary oscillations and electrostatic instability of a liquid layer of finite thickness. A similar problem was considered in Refs. 2–5 but some problems were not resolved: in Ref. 2 there was an error — one term was lost in the dispersion relation. In Refs. 3–5, the authors only considered asymptotic cases. Situations with high-viscosity and zero-viscosity liquid layers of extremely large and very small thickness were analyzed in Ref. 3, but the asymptotic expression of the dispersion relation for thin layers of liquid was written incorrectly. Intermediate situations of the greatest interest for practical applications were not considered. In Ref. 4 the approximation of an ideal liquid was used to study the influence of the disjoining pressure on the stability of a charged thin liquid layer. In Ref. 5 the problem was solved for thin layers of low-viscosity liquids.

1. We propose to calculate the spectrum of capillary waves on a flat charged surface of an ideally conducting liquid of finite depth d, bounded by vacuum, having the density ρ, viscosity ν, and surface tension γ, exposed to the gravitational field g and an electrostatic field. The upper medium has the permittivity ε. The strength of the electric field $\mathbf{E}$ at the liquid surface is determined by the potential difference between the electrodes: a lower electrode positioned at $z=-d$, covered by a layer of liquid maintained at the potential $\Phi_1=0$, and a parallel counterelectrode, separated from the liquid surface by the distance b and having the potential $\Phi_2=V$.

We position a Cartesian coordinate system such that the z axis is directed vertically upward $\mathbf{n}_z \| -\mathbf{g}$ ($\mathbf{n}_z$ is the unit vector of the z Cartesian coordinate) and the x axis lies in the direction of motion of a plane capillary wave $\sim \exp(st+ikx)$. We also assume that the $z=0$ plane coincides with the free, unperturbed surface of the liquid (s is the complex frequency, k is the wave number, t is the time, and i is an imaginary unit). Let us suppose that the function $\xi(x,t)=\xi_0\exp(st+ikx)$ describes a small perturbation of an equilibrium flat liquid surface induced by thermal capillary wave motion of extremely small ($\xi_0\sim(kT/\gamma)^{1/2}$) amplitude, where k is Boltzmann's constant, T is the absolute temperature, and $\mathbf{U}(\mathbf{r},t)$ is the velocity field of the liquid motion induced by the perturbation $\xi(x,t)$, having the same order of smallness.

To simplify the following reasoning, we cast all the quantities in dimensionless form to their characteristic values, taking $g=\rho=\gamma=1$,

$$d_*\left(\frac{\rho g}{\gamma}\right)^{-1/2}, \qquad \nu_*=\left(\frac{\gamma^3}{\rho^3 g}\right)^{1/4}, \qquad k_*=\left(\frac{\rho g}{\gamma}\right)^{1/2},$$

$$s_*=\left(\frac{\rho g^3}{\gamma}\right)^{1/2}, \qquad W_*=\sqrt{\rho g \gamma}.$$

The linearized system of hydrodynamic equations for a viscous liquid (we retain the previous notation for all the dimensionless quantities), describing the liquid motion in this particular system, has the form

$$\frac{\partial \mathbf{U}}{\partial t}=-\frac{1}{\rho}\nabla P(\mathbf{U})+\nu\nabla^2\mathbf{U}+\mathbf{g}, \tag{1}$$

$$\mathrm{div}\,\mathbf{U}=0, \tag{2}$$

$$z=-d: \quad \mathbf{U}=0, \tag{3}$$

$$z=0: -\frac{\partial \xi(x,t)}{\partial t}+U_z=0, \tag{4}$$

$$z=0: \quad \mathbf{n}\cdot(\boldsymbol{\tau}\cdot\nabla)\mathbf{U}+\boldsymbol{\tau}\cdot(\mathbf{n}\cdot\nabla)\mathbf{U}=0, \tag{5}$$

$$z=0: \quad -P(\mathbf{U})+\rho g\xi+2\rho\nu\mathbf{n}\cdot(\mathbf{n}\cdot\nabla)\mathbf{U}$$
$$-P_E(\xi)+P_\sigma(\xi)=0, \tag{6}$$

$$\text{div } \mathbf{E}_i = 0, \quad \mathbf{E}_i = -\nabla \Phi_i \quad (i = 1, 2), \tag{7}$$

$$z = -d: \quad \Phi_1 = 0, \tag{8}$$

$$z = b: \quad \Phi_2 = V, \tag{9}$$

$$z = \xi: \quad \Phi_1 = \Phi_2, \tag{10}$$

where $P(\mathbf{U})$, $P_E(\xi)$, and $P_\sigma(\xi)$ are the corrections to the pressure of the electrical forces and the pressure of the surface tension forces caused by the surface perturbation ξ, having the first order of smallness in ξ (Refs. 6 and 7), $\mathbf{n}$ and τ are the unit vectors of the normal and the tangent to the liquid surface, and the subscripts "1" and "2" refer to the liquid and the external medium, respectively.

2. The two-dimensional nature of the problem (the perturbation of the surface profile $\xi(x,t)$ and the field $\mathbf{U}(\mathbf{r},t)$ are assumed to be independent of the coordinate y) allows the problem to be scalarized using the Helmholtz theorem, introducing the velocity field potential $\varphi(\mathbf{r},t)$ and the stream function $\psi(\mathbf{r},t)$ (Ref. 6),

$$\mathbf{U} = \hat{\mathbf{N}}_1 \varphi + \hat{\mathbf{N}}_2 \psi.$$

$$\hat{\mathbf{N}}_1 \equiv \nabla; \quad \hat{\mathbf{N}}_2 \equiv [\nabla \times \mathbf{n}_y], \tag{11}$$

where $\mathbf{n}_y$ is the unit vector of the Cartesian coordinate y, $\hat{\mathbf{N}}_1$ and $\hat{\mathbf{N}}_2$ are vector differential operators satisfying the orthogonality relations and the commutation conditions with the Laplacian operator. The Hermitian operator $\hat{\mathbf{N}}_1$ isolates the potential component of the motion, while the anti-Hermitian operator $\hat{\mathbf{N}}_2$ isolates the solenoidal (vortex) component.

Substituting the expansion (11) into the vector equations (1) and (2) and taking the eigenvalues of the operators $\hat{\mathbf{N}}_1 \cdot \hat{\mathbf{N}}_2$ and $\hat{\mathbf{N}}_2 \cdot \hat{\mathbf{N}}_1$ to be nonzero, we obtain the system of scalar equations

$$\frac{\partial \psi}{\partial t} - \nu \nabla^2 \psi = 0, \tag{12}$$

$$\nabla^2 \varphi = 0, \tag{13}$$

$$P(\mathbf{U}) = -\rho \frac{\partial \varphi}{\partial t} - \rho g z. \tag{14}$$

Having substituted the expansion (11) into Eqs. (3)–(6), we transform the boundary conditions for the vector velocity field $\mathbf{U}(\mathbf{r},t)$ to give the boundary conditions for the scalar functions $\varphi(\mathbf{r},t)$ and $\psi(\mathbf{r},t)$. Condition (3) on the lower electrode is transformed to give

$$z = -d: \quad \frac{\partial \varphi}{\partial x} - \frac{\partial \psi}{\partial z} = 0, \tag{15}$$

$$\frac{\partial \varphi}{\partial z} + \frac{\partial \psi}{\partial x} = 0. \tag{16}$$

Conditions (4)–(6) on the free liquid surface have the form

$$z = 0: \quad \frac{\partial \xi}{\partial t} = \frac{\partial \varphi}{\partial z} + \frac{\partial \psi}{\partial x}, \tag{17}$$

$$2 \frac{\partial^2 \varphi}{\partial x \partial z} + \frac{\partial^2 \psi}{\partial x^2} - \frac{\partial^2 \psi}{\partial z^2} = 0, \tag{18}$$

$$\rho \frac{\partial \varphi}{\partial t} + \rho g \xi + 2 \rho \nu \left\{ \frac{\partial^2 \varphi}{\partial z^2} - \frac{\partial^2 \psi}{\partial x \partial z} \right\}$$
$$- P_E(\xi) + P_\sigma(\xi) = 0. \tag{19}$$

The system (12)–(14) with the boundary conditions (15)–(19) is the hydrodynamic part of the problem in scalarized form.

In order to satisfy the dynamic boundary condition for the normal components of the stress tensor (19), we need to use expression (A19) for $P_E(\xi)$ obtained in the Appendix and the known expression for the correction to the pressure of the surface tension forces P_σ, which has the following form in the linear approximation with respect to ξ (Ref. 6)

$$P_\sigma(\xi) \approx -\sigma \frac{\partial^2 \xi}{\partial x^2}.$$

3. Bounded solutions of the system (12) and (13), periodic in x, will be sought in the Cartesian coordinate system in the form

$$\varphi = (C_1 \sinh(mz) + C_2 \cosh(mz)) \exp(st - ikx), \tag{20}$$

$$\psi = (C_3 \sinh(qz) + C_4 \cosh(qz)) \exp(st - ikx), \tag{21}$$

where C_1, C_2, C_3, C_4, and s are complex quantities.

Substituting Eqs. (23) and (24) into the boundary conditions (15)–(19) yields a homogeneous system of five linear algebraic equations for the unknown constants C_1, C_2, C_3, C_4, and ξ_0

$$GC_1 + 2\rho \nu q C_3 + F \xi_0 = 0,$$

$$2 \rho \nu k C_2 + G C_4 = 0,$$

$$-C_1 \sinh(kd) + C_2 \cosh(kd) - C_3 \sinh(qd)$$
$$+ C_4 \cosh(qd) = 0,$$

$$kC_1 \cosh(kd) - kC_2 \sinh(kd) + qC_3 \cosh(qd)$$
$$- qC_4 \sinh(qd) = 0,$$

$$C_2 + C_4 - s \xi_0 = 0,$$

where

$$G = \frac{\rho \nu}{k} (k^2 + q^2), \quad q^2 = k^2 + s/\nu,$$

$$F = \rho g + \gamma k^2 - \frac{\varepsilon E_0^2}{4\pi} k \coth(kb).$$

A necessary and sufficient condition for the existence of a nontrivial solution of this system of equations is that its determinant, consisting of the coefficients of the required quantities C_i and ξ_0 should be zero,

$$\begin{vmatrix} G & 0 & 2\rho\nu q & 0 & F \\ 0 & 2\rho\nu q & 0 & G & 0 \\ -\sinh(kd) & \cosh(kd) & -\sinh(qd) & \cosh(qd) & 0 \\ k\cosh(kd) & -k\sinh(kd) & q\cosh(qd) & -q\sinh(qd) & 0 \\ 0 & 1 & 0 & 1 & -s \end{vmatrix} = 0.$$

This condition gives the dispersion relation for the spectrum of capillary liquid motion in the system being analyzed

$$4qk^2(k^2+q^2)+(k^2+q^2)^2$$

$$\times(k\,\sinh(kd)\sinh(qd)-q\,\cosh(kd)\cosh(qd))$$

$$+4k^3q(q\,\sinh(kd)\sinh(qd)$$

$$-k\,\cosh(kd)\cosh(qd))-\frac{Z(k)}{\nu^2}(q\,\cosh(qd)\sinh(kd)$$

$$-k\,\cosh(kd)\sinh(qd))=0,$$

$$Z(k)\equiv k+k^3-Wk^2\,\coth(kb),$$

$$W=\frac{\varepsilon E_0^2}{4\pi},\quad E_0=-V/b. \tag{22}$$

In the limiting case of an infinitely deep ideal liquid, Eq. (22) is written quite simply

$$s^2=\frac{k}{\rho}(\varepsilon E_0^2 k\,\coth(kb)-\rho g-\gamma k^2)$$

and determines the critical conditions for the establishment of Tonks–Frenkel instability. This relation is similar to the expression derived in Ref. 7 for an ideal, ideally conducting liquid, and only differs by the factor $\coth(kb)$, which allows for the finite distance from the upper electrode. This, at first glance, minor circumstance leads to some interesting physical consequences and specifically indicates that the critical conditions for the establishment of instability of a charged liquid surface depend on the distance from the upper counterelectrode b, and this dependence becomes appreciable for k comparable to b.

4. We briefly analyze the dispersion relation (22) and consider some of its asymptotic forms. It is easy to see that if (k,q) is a solution of Eq. (22), then $(k,-q)$ and $(k,\bar{q})$ are also solutions of Eq. (22) because of the properties of hyperbolic functions[8] ($\bar{q}$ is the complex conjugate of q). A trivial solution of Eq. (22) is $q=0$. In addition, $q=k$ is always a solution of Eq. (22). In fact, substituting $q=k$ into Eq. (22) gives the equality

$$8q^5+4q^4(q\,\sinh^2(qd)-q\,\cosh^2(qd))$$

$$+4q^4(q\,\sinh^2(qd)-q\,\cosh^2(qd))$$

$$-\frac{Z(k)}{\nu^2}(q\,\cosh(qd)\sinh(qd)$$

$$-q\,\cosh(qd)\sinh(qd))=0,$$

which is an identity because of the known relation between the hyperbolic sine and cosine. Thus, the simplest obvious solutions of Eq. (22) are $s=0$ ($q=k$) and $s=-\nu k^2$ ($q=0$). The number of nontrivial roots of the dispersion relation for the case where the viscosity and depth of the liquid layer are nonzero bounded quantities, is an infinite denumerable set. The first two members (in order of increasing real components) of this set determine potential–vortex motion similar to that existing in an infinitely deep liquid, while the others determine aperiodically damped vortex motion formed as a result of interaction between the flows and the bottom.

We now consider the transition from this situation of a liquid layer of finite depth with a solid flat bottom to an infinitely deep liquid. For $d\to\infty$: $\coth(kd)\to1$, $\coth(qd)\to1$ and the dispersion relation (22) may be written as

$$(\sqrt{k^2+s/\nu}-k)\left\{4k^3\sqrt{k^2+s/\nu}-(2k^2+s/\nu)^2-\frac{Z(k)}{\nu}\right\}=0. \tag{23}$$

The nontrivial solutions are determined by the factor in braces. It is easy to see that on going to the limit of an infinitely deep layer, the trivial solution $s=0$ is retained but the infinite denumerable set of nontrivial solutions is reduced to four, of which only two lie on the upper sheet of the two-sheet Riemann surface on which the entire expression (23) is determined, and thus are observable.

In the approximation of a liquid layer of small depth $d\ll1$, we assume that the product qd is small and in Eq. (22) we take $\cosh(qd)\approx1$ and $\sinh(qd)\approx qd$, so that Eq. (22) can we rewritten as

$$\sqrt{k^2+s/\nu}\left\{4k^2\left(2k^2+\frac{s}{\nu}\right)+\left(2k^2+\frac{s}{\nu}\right)^2\right.$$

$$\times(kd\,\sinh(kd)-\cosh(kd))$$

$$+4k^3\left(\left(k^2+\frac{s}{\nu}\right)d\,\sinh(kd)-k\,\cosh(kd)\right)$$

$$\left.-\frac{Z(k)}{\nu^2}(\sinh(kd)-kd\,\cosh(kd))\right\}=0. \tag{24}$$

It is easy to see that the simplest solutions of Eq. (22) are retained.

If we assume $kd\approx qd$ (i.e., if the analysis is performed in the vicinity of the solution $s=0$), the terms $kd\,\sinh(kd)\approx k^2d^2$ should be neglected in Eq. (24) and $\cosh(kd)=1$ should be assumed. In this approximation, we obtain a dispersion relation which only has the simplest solutions

$$q(k^2 - q^2)^2 = 0.$$

The nontrivial solutions of Eq. (24), determined by the factor in braces, are the roots of a quadratic equations and can easily be written as

$$s_{1,2} = -2k^2 \nu (1 + AB)$$
$$\pm \sqrt{4k^4 \nu^2 (AB - 2A - 1) - Z(k)AC},$$

$$A = 1/(dk \sinh(kd) - \cosh(kd)), \quad B = 1 + dk \sinh(kd),$$

$$C = dk \cosh(kd) - \sinh(kd).$$

Thus, in the thin layer approximation both of the simplest solutions and only two nontrivial solutions are retained, which may be real or complex depending on the values of the physical parameters, i.e., they may determine aperiodic and wave potential–vortex motion, similar to the motion in an infinitely deep liquid. Aperiodically damped vortex motion associated with the presence of a bottom does not occur in this approximation.

Since an attempt to go to the limit of an infinitely thin layer, where $kd \approx qd$ (for $\sinh(kd) \approx kd$, $\cosh(kd) \approx 1$ $(d \to 0)$), yields a dispersion relation having only trivial solutions, we analyze this case, retaining small quantities up to the third order of smallness

$$\sinh(kd) \approx kd + \frac{k^3 d^3}{6}, \quad \cosh(kd) \approx 1 + \frac{k^2 d^2}{2}.$$

We substitute these expansions into Eq. (22) and write the resulting equation in a form which explicitly contains s

$$\frac{s}{\nu} \sqrt{k^2 + s/\nu} \left\{ \frac{s}{2} \frac{d^2}{\nu} \left(\frac{s}{\nu} + 4k^2 \right) + \frac{s}{\nu} + \frac{Z(k)kd^3}{3\nu^2} \right\} = 0.$$

It can be seen that this equation contains the two simplest solutions and two nontrivial solutions determined by a quadratic equation with coefficients which depend on different powers of the small parameter

$$s^2 \frac{d^2}{2\nu^2} + \frac{s}{\nu}(1 + 2d^2 k^2) + \frac{Z(k)}{3\nu^2} kd^3 = 0. \tag{25}$$

Both roots of Eq. (25)

$$s_{1,2} = \nu \left(\frac{1}{d^2} + 2k^2 \right)$$
$$\pm \sqrt{\nu^2 \left(\frac{1}{d^2} + 2k^2 \right)^2 - \frac{2}{3} Z(k)kd}, \tag{26}$$

like the solutions of Eq. (24), which reduce to expressions (26), describe the potential–vortex motion of the liquid.

A striking feature of this analysis is that all the solutions of the dispersion relation for a liquid of any finite depth are determined on the upper sheet of the Riemann surface and theoretically may be observable. For a liquid of infinite depth there are also nonobservable solutions determined on the lower sheet of the two-sheet Riemann surface.

A numerical analysis of the dispersion relation (22) reveals that this has an infinite single-parameter family of solutions (there is an infinite number of branches with numbers
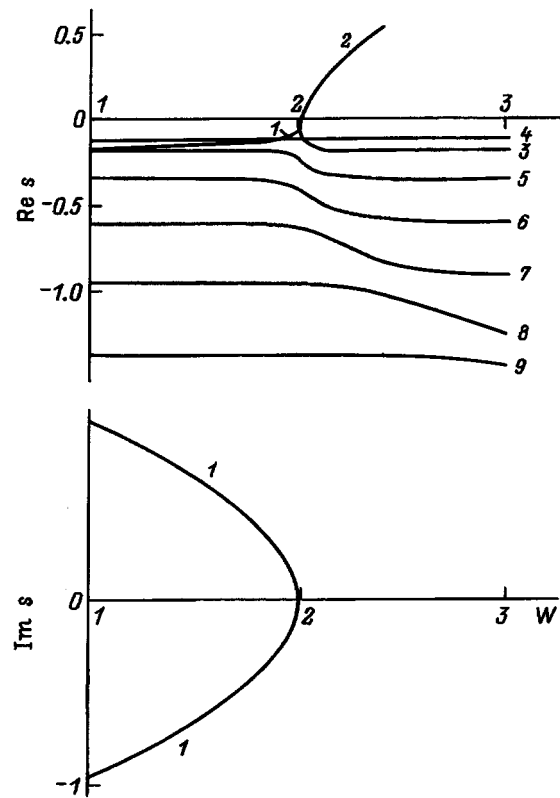

FIG. 1. Real and imaginary components of the dimensionless frequency s of the capillary wave motion as a function of the dimensionless parameter W characterizing the electric field pressure on the free liquid surface: $b = 10$, $k = 1$, $\nu = 0.1$, and $d = 5$.

greater than 3 and with ordinates whose absolute value increases with increasing branch number) (Fig. 1), unlike the dispersion relation for a liquid of infinite depth[9] whose spectrum of capillary motion is described by curves 1–3 in Fig. 1. From the mathematical point of view, this circumstance arises because hyperbolic functions of the complex frequency appear in the dispersion relation in this situation. From the physical point of view, the appearance of an infinite family of aperiodic, strongly damped motion is caused by reflection of the moving liquid from the bottom. Curve 2 in Fig. 1 determines the growth rate of instability of a charged liquid surface in the region Re $s > 0$ (the growth rate of Tonks–Frenkel instability). Curve 1 describes the damped wave motion. Curve 3 together with part of curve 2 determines the damped aperiodic motion of the liquid in the region Re $s < 0$. All new branches of the dispersion relation compared with the case of an infinitely deep liquid determine strongly aperiodically damped vortex motion.

The influence of the finite layer thickness on the parameters of the liquid motion becomes substantial when the thickness is comparable to the wavelength, as can be seen from Figs. 2 and 3, which give the complex frequency as a function of the wave number for subcritical and supercritical values of the electric field at the liquid surface. Figure 2b gives the left-hand side of Fig. 2a on a larger scale to show how the transition takes place from a state of stable capillary liquid motion to Tonks–Frenkel instability, which is achieved for $W > 2$. The part of branch 2 tangent to the ab-
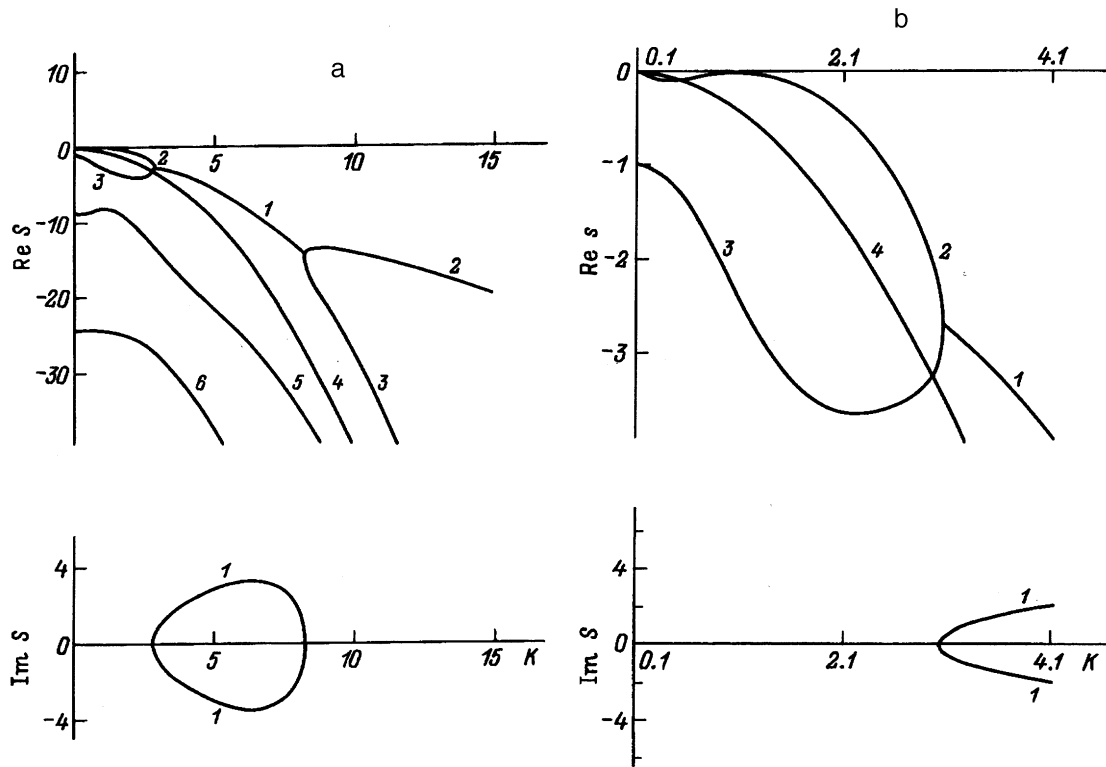
FIG. 2. Real and imaginary components of the frequency s as a function of the wave number k for $b = 10$, $\nu = 0.4$, $d = 1$, and $W = 2$ (a) and left-hand side of Fig. 2a at higher resolution (b).

scissa at $k = 1$ rises above it and gives the growth rate of the Tonks–Frenkel instability, as can be seen from Fig. 3, where Re $s = $ Re $s(k)$ is plotted for $W = 4$.

Figures 4 and 5 give the complex frequency as a function of the liquid layer depth d for $k = 1$ for subcritical ($W = 1.9$) and supercritical ($W = 2.02$) values of the electric field at the liquid surface. It can be seen that the growth rate of the Tonks–Frenkel instability, the capillary wave frequencies, and the damping rates of the aperiodic vortex liquid motion depend on d. This dependence is strongest for $d < 2$. Calculations of the complex frequency as a function of viscosity in the subcritical and supercritical regimes show that as the viscosity increases, the damping rates increase and the growth rate of the Tonks–Frenkel instability decreases.

The fact that the condition for the existence of capillary waves is influenced by the viscosity and the layer thickness is also confirmed by the dependence plotted in Fig. 6 which for $k = 1$ (curve *1*) and $k = 10$ (curve *2*) and a fixed layer thickness, determines the characteristic viscosity for which the capillary wave motion completely disappears. The points lying above the curve correspond to the situation where there is no capillary wave motion with the given wave number while for those lying below the curve, such motion does exist. It can be seen that the layer thickness has an appreciable influence when $d < \lambda$ and a weak influence when $d > \lambda$.

CONCLUSIONS

This analysis indicates that when the thickness of a charged liquid layer decreases to a level comparable with the

wavelength, this increases the damping rates of all the damped liquid motion and reduces the growth rate of the Tonks–Frenkel instability. A decrease in the thickness of the liquid layer also narrows the spectrum of established wave motion by imposing high and low wave number limits. The spectrum of wave numbers characterizing the motion involved in the buildup of the Tonks–Frenkel instability also becomes narrower. The wave number of the most unstable mode increases with increasing electric field strength at the liquid surface and with decreasing layer thickness.

APPENDIX

We shall calculate the electrical pressure on the surface of the liquid layer in the geometry described above. The
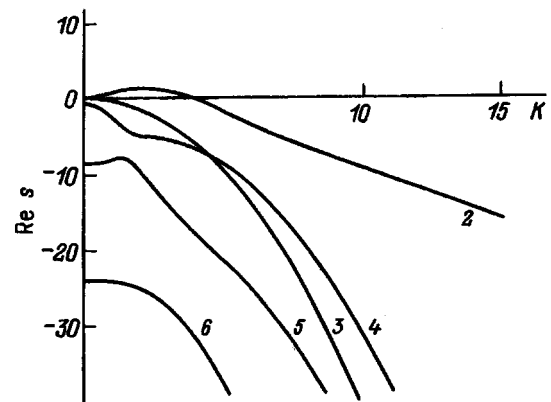

FIG. 3. Real component of s versus wave number k for $b = 10$, $\nu = 0.4$, $d = 1$, and $W = 4$.
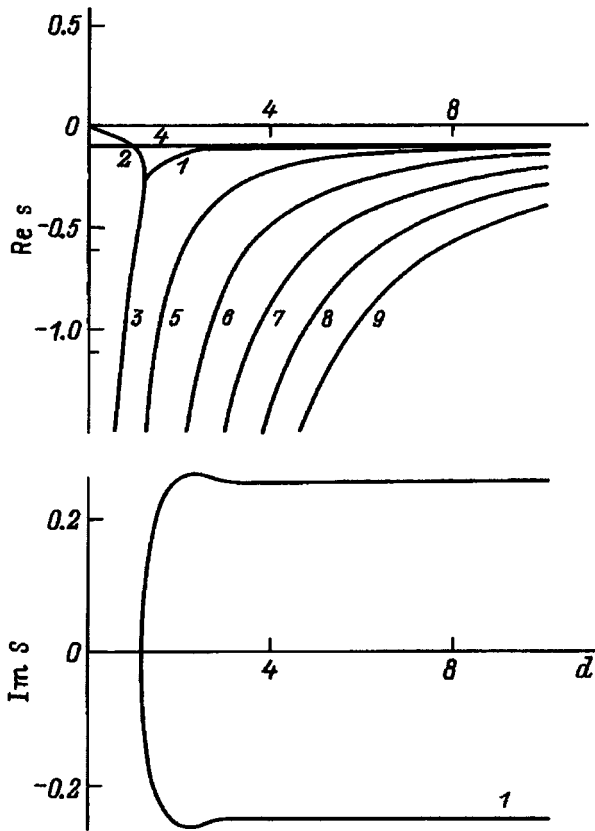
FIG. 4. Real and imaginary components of s versus liquid layer depth d for $k=1$, $b=10$, $\nu=0.1$, and $W=1.9$.

mathematical formulation of the problem of calculating the electric field between the electrodes has the form (7)–(10). Since it is known that there is no electric field in a conductor, it can immediately be assumed that the electric field potential in the liquid layer is $\Phi_1 = \mathrm{const} = 0$. Thus, the problem is reduced to the simpler one of determining the electric potential Φ_2 in the region between the perturbed liquid surface and the flat counterelectrode

$$\nabla^2\Phi_2=0, \tag{A1}$$
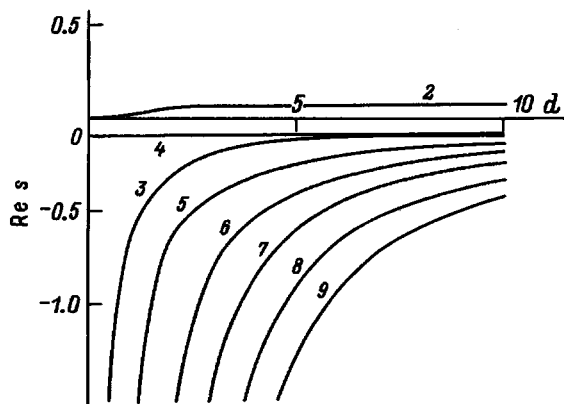
$$z=\xi: \quad \Phi_2\equiv 0, \tag{A2}$$

FIG. 5. Real component of s versus liquid layer depth d for $k=1$, $b=10$, $\nu=0.1$, and $W=2.02$.
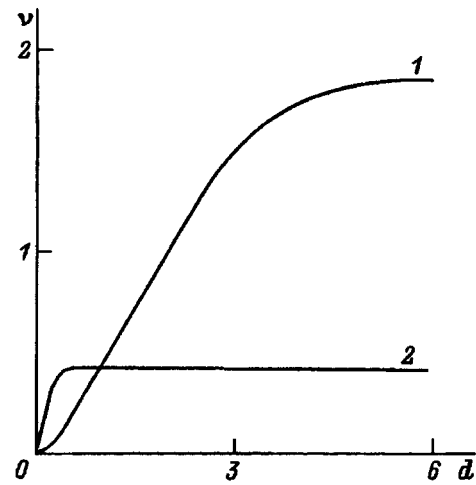

FIG. 6. Viscosity versus liquid layer thickness for which capillary wave motion ceases, for $b=10$, $W=0$, $k=1$ (*1*), and $k=10$ (*2*).

$$z=b: \quad \Phi_2\equiv V. \tag{A3}$$

We shall seek Φ_2 in the form

$$\Phi_2=\Phi^{(0)}+\Phi^{(1)}, \tag{A4}$$

where $\Phi^{(0)}$ is the steady-state solution obtained for $\xi\equiv 0$, $\Phi^{(1)}$ is a small correction to $\Phi^{(0)}$ caused by the deformation of the surface $\xi(x,t)$ and having the same order of smallness.

Decomposing the boundary condition (A2) near the unperturbed liquid surface $z=0$ in the linear approximation in ξ and $\Phi^{(1)}$

$$z=0: \quad \Phi^{(0)}+\Phi^{(1)}+\xi\frac{\partial\Phi^{(0)}}{\partial z}\equiv 0 \tag{A5}$$

and substituting Eq. (A4) into Eqs. (A1), (A3) and (A5), it is easy to obtain the problem to determine the potentials in the zeroth-order $\Phi^{(0}$ and first-order $\Phi^{(1)}$ approximations

$$\nabla^2\Phi^{(0)}=0, \tag{A6}$$

$$z=0: \quad \Phi^{(0)}\equiv 0, \tag{A7}$$

$$z=b: \quad \Phi^{(0)}\equiv V, \tag{A8}$$

$$\nabla^2\Phi^{(1)}=0, \tag{A9}$$

$$z=0: \quad \Phi^{(1)}\equiv -\xi\frac{\partial\Phi^{(0)}}{\partial z}, \tag{A10}$$

$$z=b: \quad \Phi^{(1)}\equiv 0. \tag{A11}$$

It is also clear that $\Phi^{(1)}$ should be sought as a class of functions, bounded and periodic in x.

The solution of the problem in the zeroth-order approximation has the form

$$\Phi^{(0)}=-E_0z, \quad E_0=-V/b. \tag{A12}$$

We seek the solution in the first-order approximation in the form

$$\Phi^{(1)}(x,z,t)=\Phi^{(1)}_*(x,z)\exp(mt),$$

which, after substituting into Eqs. (A9)–(A11), gives

$$\Phi^{(1)} = -\frac{E_0\xi}{\sinh(kb)} \sinh[k(z-b)]. \qquad (A13)$$

Finally the solution of Eqs. (A1)–(A3) apart from small, first-order terms has the form

$$\Phi = -E_0 z + \frac{E_0\xi}{\sinh(kb)} \sinh[k(z-b)]. \qquad (A14)$$

For the projections of the electric field strength on the axis required to write the expression for the electric field pressure on the liquid surface, we find

$$E_x = -\frac{ikE_0\xi}{\sinh(kb)} \sinh[k(z-b)], \qquad (A15)$$

$$E_z = E_0 + \frac{kE_0\xi}{\sinh(kb)} \cosh[k(z-b)]. \qquad (A16)$$

Near the unperturbed liquid surface $z=0$, the components of the electric field strength vector on the perturbed surface needed to use the dynamic boundary condition are written as

$$E_x = ikE_0\xi, \qquad (A17)$$

$$E_z = E_0(1 + k\xi\coth(kb)). \qquad (A18)$$

These quantities determine the components of the Maxwellian stress tensor

$$\Pi_{ik} = \varepsilon E_i E_k - \frac{\varepsilon}{2} E^2 \delta_{ik}$$

on the perturbed liquid surface

$$-\Pi_{zz} = \Pi_{xx} = -\frac{\varepsilon}{4\pi}\left(\frac{E_0^2}{2} + E_0^2 k\xi \coth(kb)\right),$$

$$\Pi_{xy} = \Pi_{yz} = \Pi_{yy} = 0, \quad \Pi_{xz} = \frac{\varepsilon}{4\pi} E_0^2 ik\xi.$$

Assuming that the unit vector $\mathbf{n}$ of the normal to the flat free liquid surface perturbed by the wave motion in the linear approximation with respect to the small parameter is written as

$$\mathbf{n} = ik\xi\mathbf{n}_x + \mathbf{n}_z,$$

and retaining terms of the zeroth and first orders of smallness, it is easy to derive an expression for the momentum flux density entering the liquid (i.e., for the force acting per unit surface area),

$$\Pi_{ik} n_k = \frac{\varepsilon E_0^2}{8\pi}(ik\xi\mathbf{n}_x + (1 + 2k\xi\coth(kb))\mathbf{n}_z).$$

After scalar multiplication of this vector by the normal vector $\mathbf{n}$ in the linear approximation with respect to ξ, it is easy to derive an expression for the electric field pressure on the free liquid surface in this electrode system, induced by the perturbation $\xi(x,t)$

$$P_E(\xi) = \frac{\varepsilon E_0^2}{8\pi}(1 + 2k\xi\coth(kb)). \qquad (A19)$$

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

[2] J. He, N. M. Miskovsky, P. H. Citler, and M. Chung, J. Appl. Phys. **68**, 1475 (1990).

[3] N. G. Surgy, J. -P. Chabrerie, O. Denoux, and J. E. Wesfreid, J. Phys. II (Paris) **3**, 1201 (1990).

[4] A. I. Grigor'ev. M. I. Munichev, and S. O. Shiryaeva, J. Coll. Int. Sci. No. 166, 267 (1994).

[5] A. É. Lazaryants and A. I. Grigor'ev, Zh. Tekh. Fiz. **60**, 29 (1990) [Sov. Phys. Tech. Phys. **35**, 651 (1990)].

[6] V. G. Levich, *Physicochemical Hydrodynamics* [in Russian], Fizmatgiz, Moscow (1959).

[7] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, (Pergamon Press, Oxford, 1960) [Russian original, Nauka, Moscow, 1992].

[8] I. N. Bronshteĭn and K. A. Semendyaev, *Handbook of Mathematics for Mathematicians and University Students* [in Russian], Nauka, Moscow (1980).

[9] A. I. Grigor'ev, O. A. Grigor'ev, and S. O. Shiryaeva, Zh. Tekh. Fiz. **62**, 12 (1992) [Sov. Phys. Tech. Phys. **37**, 904 (1992)].

Translated by R. M. Durham

Capillary oscillations of a flat charged surface of liquid with finite conductivity

S. O. Shiryaeva, A. I. Grigor'ev, and V. A. Koromyslov

Yaroslavl State University, 150000 Yaroslavl, Russia

A dispersion relation is proposed and analyzed for the spectrum of capillary motion at a charged flat liquid surface with allowance made for the finite rate of charge redistribution accompanying equalization of the potential as a result of the wave deformation of the free surface. It is shown that when the conductivity of the liquid is low, a highly charged surface becomes unstable as a result of an increase in the amplitude of the aperiodic charge-relaxation motion of the liquid and not of the wave motion, as is observed for highly conducting media. The finite rate of charge redistribution strongly influences the structure of the capillary motion spectrum of the liquid and the conditions for the establishment of instability of its charged surface when the characteristic charge relaxation time is comparable with the characteristic time for equalization of the wave deformations of the free surface of the liquid. © *1997 American Institute of Physics.* [S1063-7842(97)00508-4]

INTRODUCTION

In view of the wide range of different applications in physics, geophysics, engineering, and technology, many studies have been devoted to the laws governing the establishment of capillary wave motion at a charged liquid surface.[1] However, some particular aspects have not yet been sufficiently well studied. This particularly applies to the influence of the finite rate of potential equalization in a real liquid on the characteristics of the establishment of capillary motion and its stability. The few theoretical studies that have taken into account the finite conductivity of the liquid (see, for instance, Refs. 2–5) have not fully clarified the specific characteristics of the phenomenon. A preliminary analysis of this problem[5] has indicated that allowance for the finite rate of equalization of the electric potential in a real liquid first increases the order of the dispersion relation and second, increases the damping rates of all the branches. In addition, when the physical parameters of the problem are varied (the surface charge density, permittivity, and electrical conductivity), the various branches of the dispersion relation become deformed and reclose onto one another, which indicates some change in the physical meaning of the liquid motion established under the new conditions.

1. We shall calculate the spectrum of capillary waves at a flat surface of infinitely deep liquid, bounded by vacuum, having the density ρ, conductivity γ, viscosity ν, surface tension σ, and permittivity ε, exposed to the influence of a gravitational field $\mathbf{g}$ and an external electrostatic field whose strength vector $\mathbf{E}_0$ is perpendicular to the flat liquid surface ($\mathbf{E}_0 \| -\mathbf{g}$). We position a Cartesian coordinate system such that the z axis is directed vertically upward, perpendicular to the surface, and the x axis is parallel to the direction of motion of a plane capillary wave. Let us assume that the function $\xi(x,t)$ describes a small perturbation of the equilibrium flat liquid surface and $\mathbf{U}(\mathbf{r},t)$ is the velocity field of the liquid motion induced by the perturbation ξ, having the same order of smallness.

The system of hydrodynamic equations determining the formulated problem, linearized with respect to the small quantities $\mathbf{U}$ and ξ, has the form[6]

$$\frac{\partial \mathbf{U}}{\partial t} = -\frac{1}{\rho}\nabla P(\mathbf{U}) + \nu\nabla^2\mathbf{U}; \tag{1}$$

$$\nabla \cdot \mathbf{U} = 0; \tag{2}$$

$$z \to -\infty: \quad \mathbf{U} \to 0; \tag{3}$$

$$z = 0: \quad -\frac{\partial \xi(x,t)}{\partial t} + U_z = 0; \tag{4}$$

$$z = 0: \quad (\Pi_\tau^{(2)} - \Pi_\tau^{(1)}) - \rho\nu[\mathbf{n}\cdot(\boldsymbol{\tau}\cdot\nabla)\mathbf{U}$$
$$+ \boldsymbol{\tau}\cdot(\mathbf{n}\cdot\nabla)\mathbf{U})] = 0; \tag{5}$$

$$z = 0: \quad -P(\mathbf{U}) + \rho g \xi + 2\rho\nu\mathbf{n}\cdot(\mathbf{n}\cdot\nabla)\mathbf{U}$$
$$- P_E(\xi) + P_\sigma(\xi) = 0, \tag{6}$$

where $P(\mathbf{U})$, $P_E(\xi)$, and $P_\sigma(\xi)$ are the corrections to the pressure inside the liquid in the presence of an electric field, the pressure of the electrical forces, and the pressure of the surface tension forces caused by the surface perturbation ξ and having the first order of smallness in $\mathbf{U}$ and ξ, $\mathbf{n}$ and $\boldsymbol{\tau}$ are the unit vectors of the normal and the tangents to the liquid surface, $\Pi_\tau = (\varepsilon/4\pi)E_n E_\tau$, E_n and E_τ are the normal and tangential components of the electric field strength, and the subscripts "1" and "2" refer to the liquid and the external medium, respectively.

We supplement the system (1)–(6) with the equations taking into account the finite rate of electric charge redistribution accompanying equalization of the potential at the oscillating liquid surface,

$$\text{div}\mathbf{E}_j = 0; \quad \mathbf{E}_j = -\nabla\Phi_j \quad (j=1,2); \tag{7}$$

$$z \to \infty: \quad \mathbf{E}_2 \to \mathbf{E}_0; \quad z \to -\infty: \quad \mathbf{E}_1 \to 0; \tag{8}$$

$$z = \xi: \quad E_{2n} - \varepsilon E_{1n} = 4\pi\varkappa; \quad \Phi_1 = \Phi_2; \tag{9}$$

$$z = \xi: \quad -\gamma(\mathbf{n}\cdot\mathbf{E}_1) + (\mathbf{U}\cdot\nabla_\Sigma)\varkappa + \frac{\partial\varkappa}{\partial t} = 0;$$

$$\nabla_\Sigma \equiv \frac{\partial}{\partial x}\mathbf{n}_x + \frac{\partial}{\partial y}\mathbf{n}_y, \tag{10}$$

where Φ_i is the electric potential and $\varkappa(x,t)$ is the surface charge density (in the equilibrium state $\varkappa(x,t) \equiv \varkappa_0 = $ const: $4\pi\varkappa_0 = E_0$.)

Note that in the linear approximation with respect to small quantities, the second (convective) term in Eq. (10) is dropped, since it has the second order of smallness because it is proportional to the product of $\mathbf{U}(\mathbf{r},t)$ and the correction $\varkappa(x,t)$ to the surface charge density, appearing as a result of deformation of the liquid surface and having the value $\sim\xi$.

2. Using a scalarization method described in detail in Ref. 7, we express the velocity field $\mathbf{U}$ as the sum of three orthogonal fields

$$\mathbf{U}(\mathbf{r},t) = \sum_{j=1}^{3} \hat{\mathbf{N}}_j \Psi_j(\mathbf{r},t), \tag{11}$$

where $\Psi_j(\mathbf{r},t)$ are scalar functions determined by the form of the field $\mathbf{U}(\mathbf{r},t)$, $\hat{\mathbf{N}}_j$ are vector operators satisfying the orthogonality relations and the commutation conditions with the Laplacian operator,

$$\hat{\mathbf{N}}_j \cdot \hat{\mathbf{N}}_k = 0 \,(\text{for } j \neq k); \quad \nabla^2\hat{\mathbf{N}}_j = \hat{\mathbf{N}}_j\nabla^2 \quad (j=1,2,3) \tag{12}$$

and have the form

$$\hat{\mathbf{N}}_1 = \nabla; \quad \hat{\mathbf{N}}_2 = \nabla \times \mathbf{e}_z; \quad \hat{\mathbf{N}}_3 = \nabla \times (\nabla \times \mathbf{e}_z). \tag{13}$$

The operator $\hat{\mathbf{N}}_1$ isolates the potential component of the liquid motion, and $\hat{\mathbf{N}}_2$ and $\hat{\mathbf{N}}_3$ isolate its solenoidal (vortex) components.

Substituting the expansion (11) into the system of vector equations (1) and (2), and assuming that the eigenvalues of the operators $\hat{\mathbf{N}}_j^+ \cdot \hat{\mathbf{N}}_j$ are nonzero, we obtain a system of scalar equations equivalent to (1) and (2)

$$P(\mathbf{U}) = -\rho\frac{\partial\Psi_1}{\partial t}; \tag{14}$$

$$\nabla^2\Psi_j - (1 - \delta_{1j})\frac{1}{\nu}\frac{\partial\Psi_j}{\partial t} = 0 \quad (j=1,2,3). \tag{15}$$

Having substituted the expansion (11) into Eqs. (3)–(6), we transform the boundary conditions for the vector function $\mathbf{U}$ to give the conditions for the scalar functions Ψ_j and after simple transformations, we obtain

$$z \to -\infty: \quad \Psi_1 + \frac{\partial\Psi_3}{\partial z} = 0;$$

$$\frac{\partial\Psi_1}{\partial z} - \frac{\partial^2\Psi_3}{\partial x^2} = 0; \quad \Psi_2 = 0. \tag{16}$$

The kinematic boundary condition at the free liquid surface has the form

$$z = 0: \quad -\frac{\partial\xi}{\partial t} + \frac{\partial\Psi_1}{\partial z} - \frac{\partial^2\Psi_3}{\partial x^2} = 0. \tag{17}$$

The boundary condition (5) for the tangential components of the stress tensor is split into two because the unit vectors of the coordinate axes $\mathbf{e}_x$ and $\mathbf{e}_y$ may be selected as the tangent vector τ

$$z = 0: \quad (\Pi_{\tau(x)}^{(2)} - \Pi_{\tau(x)}^{(1)}) - \rho\nu\frac{\partial}{\partial x}$$
$$\times\left[2\frac{\partial\Psi_1}{\partial z} + \left(\frac{\partial^2}{\partial z^2} - \frac{\partial^2}{\partial x^2}\right)\Psi_3\right] = 0;$$

$$z = 0: \quad (\Pi_{\tau(y)}^{(2)} - \Pi_{\tau(y)}^{(1)}) + \rho\nu\frac{\partial^2\Psi_2}{\partial x\partial z} = 0. \tag{18}$$

The boundary condition (6) for the normal component of the stress tensor with allowance for Eq. (14) has the form

$$z = 0: \quad \rho\frac{\partial\Psi_1}{\partial t} + \rho g\xi + 2\rho\nu\left\{\frac{\partial^2\Psi_1}{\partial z^2} - \frac{\partial^2}{\partial x^2}\frac{\partial\Psi_3}{\partial z}\right\}$$
$$-p_E(\xi) + P_\sigma(\xi) = 0. \tag{19}$$

The system (14) and (15) with the boundary conditions (16)–(19) is the hydrodynamic part of the problem in scalarized form.

3. Suppose that the velocity field $\mathbf{U}$ and the free surface perturbation ξ depend exponentially on the time t: $U \sim \exp(\omega t)$, $\xi \sim \exp(\omega t)$, where ω is the complex frequency. Then the solutions of system (15) satisfying the conditions (16) may be written in Cartesian coordinates as

$$\Psi_1(\mathbf{r},t) = \int_0^\infty dk B_1 \exp(kz)\exp(ikx)\exp(\omega t);$$

$$\Psi_j(\mathbf{r},t) = \int_0^\infty dk B_j \exp\left(\sqrt{k^2 + \frac{\omega}{\nu}}z\right)$$
$$\times\exp(ikx)\exp(\omega t)(j=2,3), \tag{20}$$

where i is an imaginary unit.

The function $\xi(x,t)$ describing the perturbation of the liquid surface can also be expressed as an expansion in terms of plane waves

$$\xi(x,t) = \int_0^\infty dk C \exp(ikx)\exp(\omega t). \tag{21}$$

In Eqs. (20) and (21) B_j and C are constants, expansion coefficients whose relation is determined by the boundary conditions (17)–(19).

In order to satisfy the dynamic boundary conditions (18) and (19), we use the expressions derived in the Appendix for the electric parts of the tangential components of the stress tensor (A15) and (A16), and for the correction to the pressure of the electric forces $P_E(\xi)$ (A14), as well as the known expression for the correction to the surface tension forces P_σ in the linear approximation with respect to ξ (Ref. 6)

$$P_\sigma(\xi) \approx -\sigma\frac{\partial^2\xi}{\partial x^2}. \tag{22}$$

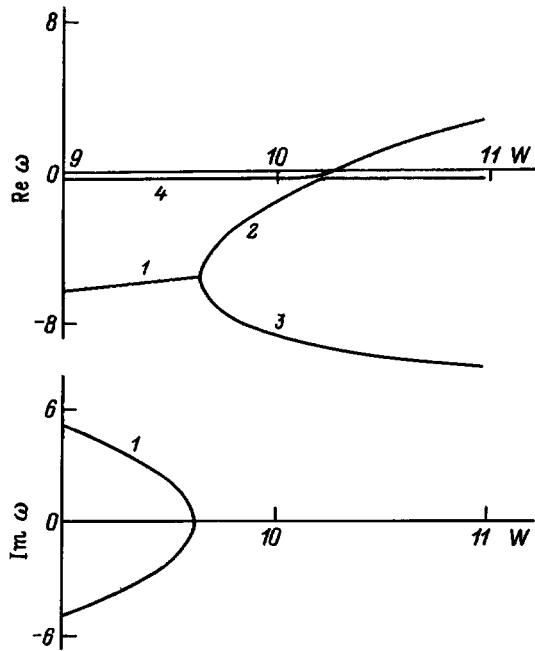Satisfying the boundary conditions (17)–(19) with allowance for these relations, we can obtain the dispersion

FIG. 1. Real and imaginary components of the dimensionless frequency ω of capillary motion of a liquid as a function of the dimensionless parameter W characterizing the electric field pressure (the surface charge density) on the free surface of the liquid, calculated for $\nu = 0.1$ and $\gamma = 0.1$.

FIG. 2. As Fig. 1 for $\nu = 0.01$ and $\gamma = 0.1$.

relations characterizing the capillary motion of the liquid. Note that we are dealing with two dispersion relations, since the problem of determining the vortex motion described by the function $\Psi_2(\mathbf{r},t)$ is completely autonomous, does not depend on the functions $\Psi_1(\mathbf{r},t)$, $\Psi_3(\mathbf{r},t)$, and $\xi(x,t)$ (i.e., has no influence on the formation of the liquid surface relief), and leads to a separate dispersion relation. However, we first consider the boundary-value problem for $\Psi_1(\mathbf{r},t)$, $\Psi_3(\mathbf{r},t)$, and $\xi(x,t)$.

Substituting the solutions (20) and (21) into the boundary conditions (17)–(19), with allowance for expressions (38), (39), and (22), we obtain a homogeneous system of three linear equations for the coefficients B_1, B_3, and C, which has a nontrivial solution when its determinant vanishes — this condition also gives one of the dispersion relations

$$\frac{k}{\rho}(\sigma k^2 + g\rho) + [(\omega + 2k^2\nu)^2 - 4(k^2\nu)^{3/2}\sqrt{\omega + \nu k^2}]$$

$$- \frac{k^2}{\rho}\frac{4\pi\varkappa_0^2}{4\pi\gamma + \omega(\varepsilon + 1)}\{(4\pi\gamma + \omega\varepsilon)$$

$$+ (\omega + 2k^2\nu) - 2(k^2\nu)^{1/2}\sqrt{\omega + \nu k^2}\} = 0. \quad (23)$$

We now change to dimensionless variables where $\rho = g = \sigma = 1$. Then, all the quantities (for which we retain the previous notation) will be expressed in fractions of the characteristic values

$$k_* = \left(\frac{\rho g}{\sigma}\right)^{1/2}, \quad \omega_* = \left(\frac{\rho g^3}{\sigma}\right)^{1/4}, \quad \gamma_* = \left(\frac{\rho g^3}{\sigma}\right)^{1/4},$$
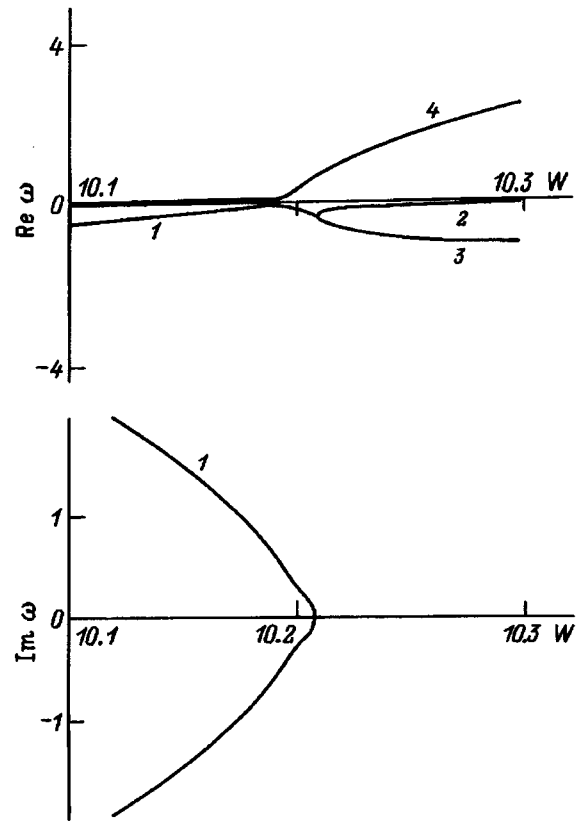
$$\varkappa_* = (\rho g \sigma)^{1/4}, \quad \nu_* = \left(\frac{\sigma^3}{\rho^3 g}\right)^{1/4}.$$

The dispersion relation (23) then has the form

$$k(K^2 + 1) + [(\omega + 2k^2\nu)^2 - 4(k^2\nu)^{3/2}\sqrt{\omega + \nu k^2}]$$

$$- \frac{Wk^2}{4\pi\gamma + \omega(\varepsilon + 1)}\{(4\pi\gamma + \omega\varepsilon) + (\omega + 2k^2\nu)$$

$$- 2(k^2\nu)^{1/2}\sqrt{\omega + \nu k^2}\} = 0,$$

$$W = 4\pi\varkappa_0^2. \quad (24)$$

It is easy to see that Eq. (24) has a higher order than the dispersion relation neglecting the charge relaxation effect[7-11] (derived from Eq. (24) for $\gamma \to \infty$). In addition to the capillary wave branches usual for a flat charged surface of ideally conducting liquid, branches of damped liquid motion also appear as a result of redistribution of the charge over the liquid surface during its deformation. These branches are logically called branches of charge-relaxation motion.

The results of the numerical calculations using Eq. (24) for $k = 10$, $\varepsilon = 100$, and various values of the viscosity ν and conductivity γ are illustrated in Figs. 1–4, which give dependences of the real and imaginary components of the dimensionless frequency on the dimensionless electric field pressure on the unperturbed liquid surface. It is easy to see that in addition to the branches of capillary liquid motion usual for a flat charged liquid surface (branches 1–3), other branches 4–7 of liquid motion also appear as a result of
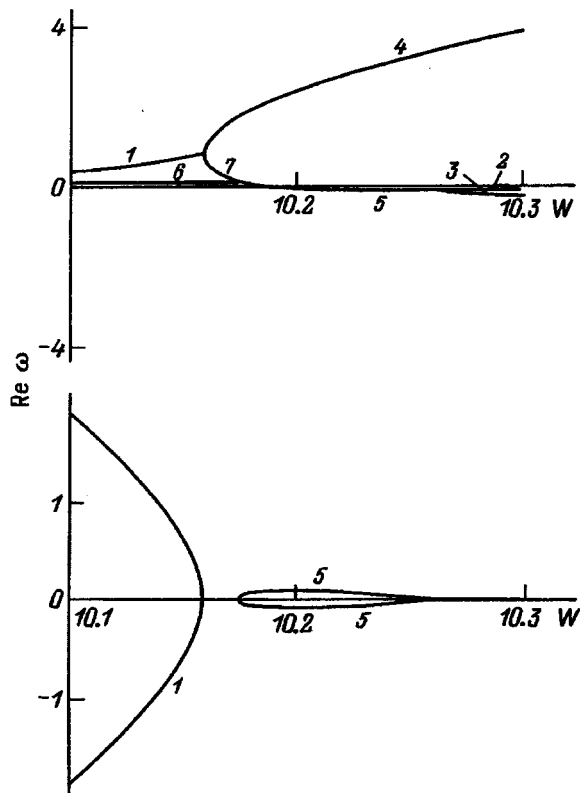
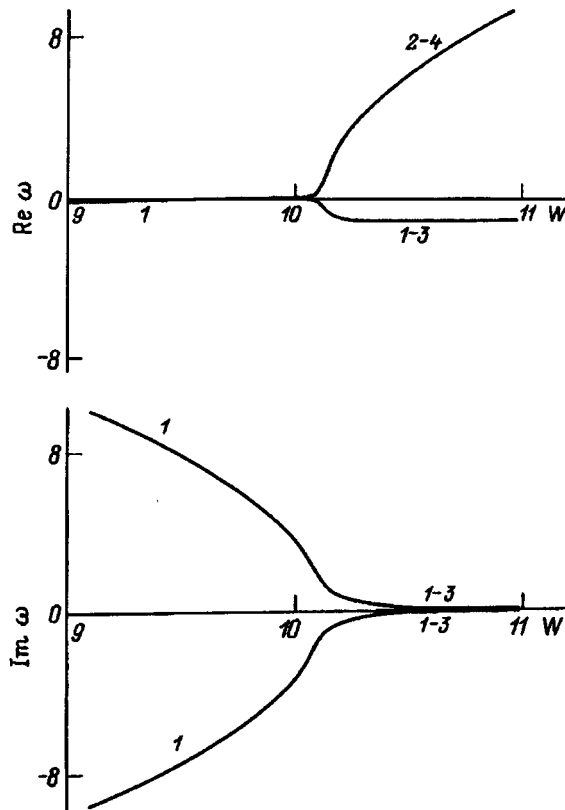FIG. 3. As Fig. 1 for $\nu=0.001$ and $\gamma=0.1$.

FIG. 4. As Fig. 1 for $\nu=0.001$ and $\gamma=10$.

redistribution of the charge over the liquid surface as it undergoes deformation. The branches of the dispersion relation lying on the lower sheet of the two-sheet Riemann surface on which the dispersion relation (24) is determined and thus not realizable in practice, are merely given in order of elimination in the figures, where this helps to understand the meaning of the new branches. In Figs. 1–4 the curves $\mathrm{Re}\,\omega(W)$ are plotted for the range of W values containing the point where the capillary branch of the $k=10$ mode is transferred to an unstable state.

The calculations show (and it is obvious from the physical meaning) that the finite rate of equalization of the potential along the liquid surface has a substantial effect at low conductivities γ. It can also be seen from Figs. 1–3 that the viscosity of the liquid plays an important role in the structuring of its charge-relaxation motion and as the viscosity increases, the structure of these branches of motion becomes appreciably more complex.

Note that for an ideal, but real, conducting liquid, it is impossible to allow exactly for the finite rate of charge redistribution during capillary motion of the liquid because the charge relaxation motion is generated by tangential stresses at the liquid surface, which are neglected in an ideal liquid (at the free surface of an ideal liquid there are no boundary conditions for the tangential component of the stress tensor).

A striking feature of Fig. 1 is that as W increases, the charge-relaxation motion 4 becomes unstable and not the capillary motion branch 2, as is observed for an ideally conducting liquid. It can be seen from Figs. 1 and 2 that, other conditions being equal, as the viscosity of the liquid increases, the capillary-motion branches 1–3 ascend toward the abscissa. It is interesting to note that in the range $10.1<W\leqslant10.2$, directly after branch 4 has passed from the region of negative to positive $\mathrm{Re}\,\omega$ at $W=10.1$, the instability growth rate of the charge relaxation waves changes negligibly with increasing W up to $W\approx10.2$, remaining extremely low. At $W\approx10.2$ the growth rate then begins to increase rapidly with increasing W.

A further increase in viscosity (Fig. 3) leads to a substantially more complex structure of charge-relaxation and capillary motion of the liquid as a result of the displacement and reclosure of various branches of the dispersion relation and the appearance of new branches. For instance, branch 1 ascends partially into the region $\mathrm{Re}\,\omega>0$ and closes onto branch 4. Then, part of branch 4 in the region $\mathrm{Re}\,\omega>0$ for $10.1\leqslant W\leqslant10.2$ lies below branch 1 and is renumbered as branch 6. Curve 7, linking the point of intersection of curves 1 and 4 with the point of intersection of curves 6 and 4 describes the growth rate of aperiodic instability which decreases with increasing surface charge density (with increasing W). To the right of the point of intersection of branches 4 and 6 is the start of branch 5, which passes from the region $\mathrm{Re}\,\omega>0$ to $\mathrm{Re}\,\omega<0$ and has an imaginary component, i.e., describing periodic wave motion, which increases with time in the left part of branch 5 in the range $10.175\leqslant W\leqslant10.195$ and decays in the right part in the range $10.195\leqslant W\leqslant10.257$. The end of the real part of branch 5 spawns two branches of aperiodically damped motions 2 and 3.

Part of branch 1 lying in the region $\mathrm{Re}\,\omega>0$ passes into the region $\mathrm{Re}\,\omega<0$ at $W=9.91$. In the range
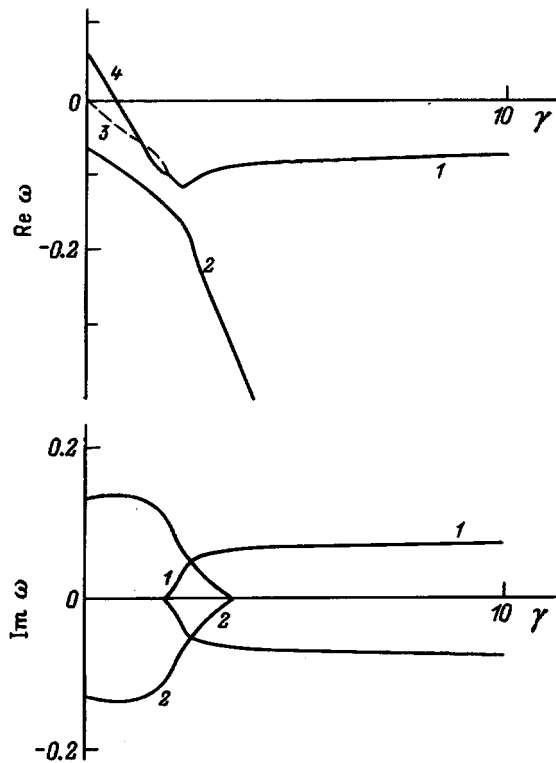
FIG. 5. Real and imaginary components of the dimensionless frequency ω of the capillary motion of a liquid as a function of the dimensionless parameter γ characterizing the liquid conductivity for $\varepsilon = 100$, $W = 1.98$, and $\nu = 0.1$.

FIG. 6. As Fig. 5 for $W = 2.06$ and $\nu = 0.1$.

$9.91 \leqslant W \leqslant 10.15$, branch *1* describes oscillatory instability of the capillary waves. For $W < 9.91$ branch *1* describes damped capillary waves.

In Fig. 4 these dependences are plotted for a liquid of substantially higher conductivity $\gamma = 10$. The curves $\mathrm{Re}\,\omega \approx \mathrm{Re}\,\omega(W)$ and $\mathrm{Im}\,\omega \approx \mathrm{Im}\,\omega(W)$ in this case are close to those for an ideally conducting liquid: branch *1* describes damped capillary waves. Branch *2–4* gives the growth rate of the Tonks–Frenkel instability. Branch *1–3* is a direct continuation of branch *1* and describes damped capillary waves. However, in practice the motion determined by branch *1–3* is aperiodic, since the frequencies of this motion are substantially lower than their damping rates. In one oscillation period the amplitude of this wave decreases by more than an order of magnitude. As a result of a further increase in the conductivity γ, the imaginary part of branch *1–3* disappears for W corresponding to the point of intersection of curves *1* and *2* and its real part numbered *3* describes purely aperiodically damped motion. Thus, the region of physical characteristics of liquids for which the finite rate of charge redistribution has a substantial influence on the spectrum of capillary motion is limited by media having low conductivities.

Figures 5–7 give the real and imaginary components of the dimensionless frequency as a function of the liquid conductivity γ for fixed W, ε, and ν. Figure 5 shows the calculated capillary motion associated with the $k = 1$ mode for $\nu = 0.1$ and subcritical charge density (for $W = 1.98$). Branch *1* describes damped capillary wave motion. Branch *2* describes damped charge-relaxation wave motion. At the right

end of the real and imaginary components of branch *2*, this branch is transferred to the lower sheet of the Riemann surface. Branch *3* lies on the lower sheet of the Riemann surface and is nonobservable. Branch *4* exhibits charge-relaxation behavior, where the part of this branch in the region $\mathrm{Re}\,\omega > 0$ describes aperiodic instability established in a narrow range of conductivities γ and characterized by low growth rates (as can be seen from Fig. 5).

Figure 6 gives results of calculations of the capillary motion of the $k = 1$, $\nu = 0.1$ mode for a supercritical surface charge density (for $W = 2.05$). The instability growth rate is

FIG. 7. As Fig. 5 for $W = 2.05$ and $\nu = 1$.

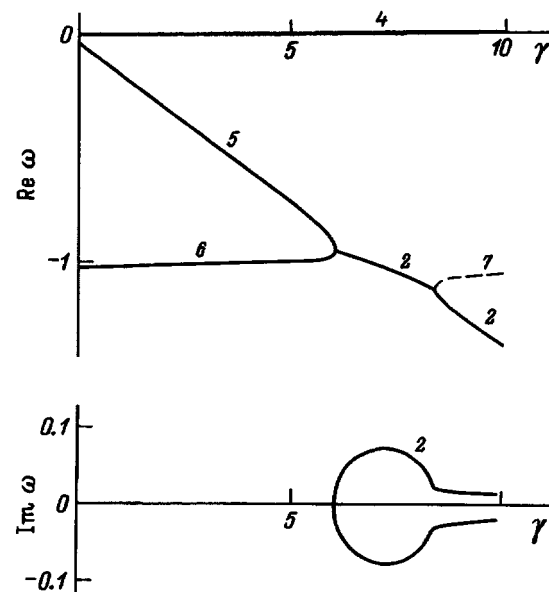determined by branch *4*. Branch *2* determines damped charge-relaxation waves while branches *5* and *6* determine damped aperiodic motion.

Figure 7 gives the same curves as in Fig. 6, but for a substantially higher viscosity $\nu = 1$. Branch *7* lies on the lower sheet of the Riemann surface. It is easy to see that the instability growth rate of branch *4* decreased substantially with increasing viscosity and the range of existence of the charge-relaxation waves was shifted toward higher conductivity.

Note that this charge-relaxation motion is generated by a phase lag between the capillary motion of the liquid deforming the surface and the associated currents of redistributing charge, which dissipate energy as Joule heat.

It is clear from the physical meaning of this effect that the change in the structure of the capillary wave motion described only occurs for those wavelengths in poorly conducting liquids for which the characteristic relaxation time of the electric charge $\tau_\varepsilon = \varepsilon / \gamma$ is greater than or comparable with the period T_k of the capillary wave motion

$$T_k = 2\pi(\rho/(\sigma k^2 + g\rho - 4\pi\varkappa_0^2 k)k)^{1/2}.$$

The constraint $\tau > T$ is not too stringent and may be satisfied for capillary motion of high wave numbers even for a liquid such as water, which is commonly used in practical electrospraying, and is not considered to be a poor conductor.

It is interesting to note that the frequencies of the capillary and relaxation waves have qualitatively different dependences on viscosity: for capillary waves, the frequency decreases with increasing viscosity whereas for relaxation waves, it increases. This behavior evidently arises because the capillary wave motion is generated by a pressure change at the free surface (from the mathematical point of view, it is generated by the dynamic boundary condition for the normal components of the stress tensor), whereas the relaxation waves are associated with the establishment of tangential stresses at the liquid surface (with a change in the form of the dynamic boundary condition for the tangential components of the stress tensor). The relation between the frequency and the viscosity in the dynamic boundary conditions for the normal and tangential components of the stress tensor is the opposite.

4. Let us now consider the boundary-value problem for the function $\Psi_2(\mathbf{r}, t)$. From the second boundary condition (18) for Ψ_2 with allowance for the solution (20) and expression (A16), dispensing with the integration over k by Fourier transformation, we obtain

$$B_2\left(k\sqrt{k^2 + \frac{\omega}{\nu}}\right) = 0. \tag{25}$$

The condition may be satisfied in two ways: 1) by imposing the constraint that all the coefficients B_2 vanish, which is equivalent to equating the entire function to zero $\Psi_2(\mathbf{r}, t) = 0$; 2) by assuming that $\Psi_2(\mathbf{r}, t) \neq 0$, in which case the dispersion relation for the vortex motion of the liquid, determined by the function $\Psi_2(\mathbf{r}, t)$ is obtained by equating to zero the expression in braces in Eq. (25). This

dispersion relation has the solution $\omega = -\nu k^2$, i.e., the corresponding liquid motion is aperiodically damped with the damping rate $\sim k^2$.

It is easy to see that this type of motion does not depend on the finite rate of charge redistribution, as should be expected, strictly speaking, from the general physical meaning.

CONCLUSIONS

The effect of electric charge relaxation makes the structure of the capillary motion spectrum of a liquid considerably more complex, leading to new capillary wave and aperiodic motion. For poorly conducting liquids, charge relaxation leads to the onset of periodic instability which is established when the electric field strength at the liquid surface is below the critical value for the Tonks–Frenkel instability, which is essentially an aperiodic instability. The conditions under which the viscosity relaxation effect has a strong influence reduce to the requirement that the characteristic charge relaxation time τ_ε should be of the order of or greater than the characteristic time taken for equalization of the deformations of the free liquid surface or the characteristic viscous relaxation time.[12,13] This last constraint implies that the effect is important for liquids which are viscous in the sense of electrospraying from the meniscus at the end of a capillary, along which liquid is supplied to a discharge chamber.[12,13] Typical examples of such liquids are glycerin and cyclohexanol.

APPENDIX

We shall solve the problem (7)–(10) which will allow us to obtain expressions for the electric part of the tangential components of the stress tensor $(\Pi_\tau^{(2)} - \Pi_\tau^{(1)})$ and the correction to the pressure of the electric forces $P_E(\xi)$ with allowance for the electric charge relaxation effect, contained in the dynamic boundary conditions (18) and (19).

The system (7)–(10) allows us to formulate the boundary-value problem for the electric potentials Φ_j

$$\nabla^2\Phi_j = 0 \quad (j = 1,2), \tag{A1}$$

$$z \to -\infty: \quad \Phi_1 \text{ is bounded}; \quad \Phi_2 \to \Phi_0 = -4\pi\varkappa_0 z;$$

$$z = \xi: \quad \varkappa = \frac{1}{4\pi}\left(\varepsilon\frac{\partial\Phi_1}{\partial n} - \frac{\partial\Phi_2}{\partial n}\right); \quad \Phi_1 = \Phi_2; \tag{A2}$$

$$z = \xi: \quad \gamma\frac{\partial\Phi_1}{\partial n} + (\mathbf{U}\cdot\nabla_\Sigma)\varkappa + \frac{\partial\varkappa}{\partial t} = 0. \tag{A3}$$

We express the potentials Φ_j in the form of expansions

$$\Phi_j = \Phi_j^{(0)} + \delta\Phi_j \quad (j = 1,2), \tag{A4}$$

where $\Phi_j^{(0)}$ are the potentials for the equilibrium state of the system and $\delta\Phi_j$ are the corrections caused by the perturbation of the liquid surface ξ, which are of the same order of smallness.

Substituting Eq. (A4) into the system (A1)–(A3), we write the boundary-value problems for $\Phi_j^{(0)}$ and $\delta\Phi_j$ separately

$$\nabla^2(\Phi_j^{(0)}) = 0 \quad (j = 1,2); \tag{A5}$$

$z \to -\infty: \quad \Phi_1^{(0)} = \text{const};$

$z \to +\infty: \quad \Phi_2^{(0)} \to -4\pi\varkappa_0 z;$

$z = 0: \quad \Phi_1^{(0)} = \Phi_2^{(0)};$

$$\nabla^2(\delta\Phi_j) = 0 \quad (j = 1,2); \tag{A6}$$

$z \to -\infty: \quad \delta\Phi_1 \to 0; \quad z \to +\infty: \quad \delta\Phi_2 \to 0; \tag{A7}$

$$z = 0: \quad \delta\Phi_1 + \frac{\partial\Phi_1^{(0)}}{\partial z}\xi = \delta\Phi_2 + \frac{\partial\Phi_2^{(0)}}{\partial z}\xi; \tag{A8}$$

$$z = 0: \quad \sigma\frac{\partial(\delta\Phi_1)}{\partial z} + \frac{\omega}{4\pi}\left[\varepsilon\frac{\partial(\delta\Phi_1)}{\partial z} - \frac{\partial(\delta\Phi_2)}{\partial z}\right] = 0. \tag{A9}$$

The boundary-value problem (A5) for the potentials $\Phi_j^{(0)}$ does not present any difficulties and its solution is easily written

$$\Phi_1^{(0)} = 0; \quad \Phi_2^{(0)} = -4\pi\varkappa_0 z. \tag{A10}$$

The solutions of Eqs. (A6) satisfying the conditions (A7) are logically sought in the form of expansions in terms of plane waves

$$\delta\Phi_1(\mathbf{r},t) = \int_0^\infty dk N_1 \exp(ikx)\exp(kz)\exp(\omega t);$$

$$\delta\Phi_2(\mathbf{r},t) = \int_0^\infty dk N_2 \exp(ikx)\exp(-kz)\exp(\omega t), \tag{A11}$$

where N_1 and N_2 are constant expansion coefficients.

Satisfying the conditions (A8) and (A9) using the solutions (A10) and (A11), we obtain, in the first order of smallness, equations linking the constants N_1 and N_2 with the coefficient C in the expansion (21) of the perturbation $\xi(x,t)$ in terms of plane waves

$$N_1 = -4\pi\varkappa_0 \frac{\omega}{4\pi\gamma + \omega(\varepsilon+1)} C;$$

$$N_2 = 4\pi\varkappa_0 \frac{4\pi\gamma + \omega\varepsilon}{4\pi\gamma + \omega(\varepsilon+1)} C. \tag{A12}$$

We can now write the normal and tangential components of the electric field strength vector, apart from quantities of the first order of smallness

$$E_n^{(2)} \approx 4\pi\varkappa_0 - \frac{\partial(\delta\Phi_2)}{\partial z}; \quad E_n^{(1)} \approx -\frac{\partial(\delta\Phi_1)}{\partial z};$$

$$E_{\tau(x)}^{(2)} \approx -\frac{\partial\Phi_2^{(0)}}{\partial z}\frac{\partial\xi}{\partial x} - \frac{\partial(\delta\Phi_2)}{\partial x}; \quad E_{\tau(x)}^{(1)} \approx -\frac{\partial(\delta\Phi_1)}{\partial x};$$

$$E_{\tau(y)}^{(2)} \approx 0; \quad E_{\tau(y)}^{(1)} \approx 0. \tag{A13}$$

The general expression for the electric field pressure on the surface of a liquid bounded by vacuum has the form[11]

$$P_E = \frac{1}{8\pi}[(E_n^{(2)})^2 - \varepsilon(E_n^{(1)})^2] + (\varepsilon-1)\frac{E_\tau^{(2)^2}}{8\pi}.$$

Then, with allowance for Eq. (A13) in the linear approximation we find

$$P_E \approx P_E^{(0)} + P_E(\xi), \quad P_E^{(0)} \approx 2\pi\varkappa_0^2 = \frac{E_0^2}{8\pi};$$

$$P_E(\xi) = 4\pi\varkappa_0^2 \frac{4\pi\gamma + \omega\varepsilon}{4\pi\gamma + \omega(\varepsilon+1)}$$

$$\times \int_0^\infty dk Ck \exp(ikx)\exp(\omega t). \tag{A14}$$

The electric parts of the tangential components of the stress tensor in the first order of smallness using expressions (A11)–(A13) and (21), and the definition $\Pi_r = (\varepsilon/4\pi)E_n E_\tau$ may be given as

$$z = 0: \quad \Pi_{\tau(x)}^{(2)} - \Pi_{\tau(x)}^{(1)} \approx -\varkappa_0\frac{\partial(\delta\Phi_1)}{\partial x}$$

$$= 4\pi\varkappa_0^2\frac{\omega}{4\pi\gamma + \omega(\varepsilon+1)}\frac{\partial\xi}{\partial x}, \tag{A15}$$

$$z = 0: \quad \Pi_{\tau(y)}^{(2)} - \Pi_{\tau(y)}^{(1)} = 0. \tag{A16}$$

[1] A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk Ser. Mekh. Zhidk. Gaza No. 4, 3 (1994).

[2] J. R. Melcher and W. J. Schwarz, Phys. Fluids, **11**, 2604 (1968).

[3] J. R. Melcher and C. V. Smith, Phys. Fluids **12**, 778 (1969).

[4] M. K. Novoselets and V. S. Nikolaev, Ukr. Fiz. Zh. **32**, 713 (1987).

[5] S. O. Shiryava, A. I. Grigor'ev, and V. A. Koromyslov, Pis'ma Zh. Tekh. Fiz. **21**(19), 77 (1995) [*sic*].

[6] V. G. Levich, *Physicochemical Hydrodynamics* [in Russian], Fizmatgiz, Moscow (1959).

[7] S. O. Shiryaeva, A. É. Lazaryants *et al.*, Preprint No. 27 [in Russian], Institute of Mechanics, Russian Academy of Sciences, Yaroslavl' (1994).

[8] A. V. Aliev and A. V. Filippov, Magn. Gidrodynam. No. 4, 94 (1989).

[9] A. I. Grigor'ev, O. A. Grigor'ev, and S. O. Shiryaeva, Zh. Tekh. Fiz. **62**(9), 12 (1992) [Sov. Phys. Tech. Phys. **37**, 904 (1992)].

[10] J. Zeleny, Phys. Rev. **10**, 1 (1917).

[11] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, (Pergamon Press, Oxford, 1960), Nauka, Moscow, 1992.

[12] S. O. Shiryaeva, A. I. Grigor'ev, and A. A. Svyatchenko, Preprint No. 25 [in Russian], Institute of Mechanics, Russian Academy of Sciences, Yaroslavl' (1993).

[13] S. O. Shiryaeva and A. I. Grigor'ev, J. Electrostatics **34**, 51 (1995).

Translated by R. M. Durham

Limiting nonideal metastable supercooled plasma

A. N. Tkachev and S. I. Yakovlenko

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia

Direct modeling of the dynamics of a system of many Coulomb particles is applied to analyze the formation stage of a metastable plasma state from an initial, highly nonideal state, and also to consider some properties of this metastable supercooled state. It is shown that relaxation of the average particle kinetic energy may be characterized by a universal dimensionless function and in particular, there is a limiting degree of plasma nonideality which may be achieved in the metastable state, in the absence of any external influence. The calculated pair correlation functions agree with the results of the Debye model, even outside its limits of validity. The time dependence of the total dipole moment of the particle system is investigated. It is shown that oscillations of the total dipole moment are observed. These collective oscillations take place at a frequency slightly below the Langmuir frequency and the oscillations of free and bound electrons are in antiphase. The hypothesis is put forward that recombination relaxation is frozen as a result of interaction between quasibound electrons and Langmuir oscillations of free electrons. © *1997 American Institute of Physics.*
[S1063-7842(97)00608-9]

INTRODUCTION

The results of first-principles modeling of the dynamics of many Coulomb particles have yielded some fairly radical conclusions (see Refs. 1 and 2 and the literature cited therein). It has been shown that in a classical Coulomb plasma, recombination processes are frozen if the charged particle motion is not subjected to some stochastic external influence. Here we are talking of an external influence in relation to Newton's dynamic equations which control the particle motion. The stochastic nature of the influence is observed as a loss of dynamic memory, i.e., a loss of reversibility of the particle motion. A stochastic influence has been specially incorporated in codes used to model particle motion and has stimulated recombination processes. In addition, errors in the numerical simulation have a stochastic influence on particle motion (see for instance, Refs. 3–5 and also Refs. 1 and 2).

With the same degree of accuracy as that achieved by numerical simulation, it may be stated that a classical Coulomb plasma may be in a metastable state which is supercooled in terms of degree of ionization. In this state, the total electron energy has a quasi-steady-state distribution function, which differs radically from the Boltzmann distribution in the negative energy range and from the distribution obtained for recombination relaxation. However, the electron velocity distribution remains Maxwellian so that the concept of electron temperature can be used for this nonequilibrium metastable state.

The electron temperature obtained after relaxation to the metastable state is higher than the initial value used to select the initial particle velocities in the calculations (the initial electron and ion coordinates are usually defined assuming with a uniform distribution). If the initial temperature is sufficiently high for the plasma to be ideal, negligible heating occurs. If the initial velocities are close to zero (zero initial temperature), the formation stage of the metastable state is accompanied by heating, which impedes the development of a metastable plasma with an arbitrarily high degree of nonideality.

Here we analyze the formation of a metastable plasma state from an initial, highly ideal state, and we consider some properties of the limiting nonideal state of a metastable plasma isolated from external stochastic influences. The correlation functions between the simultaneous positions of particle pairs are calculated. Collective oscillations of the total dipole moment of the particle system are studied. The results of this new series of calculations confirm the general conclusions reached in Refs. 1–8, that a metastable state exists and that the concepts of Debye length and Langmuir frequency do not become meaningless on transition to a nonideal plasma. The hypothesis is also put forward that recombination relaxation is frozen as a result of interaction between quasibound electrons and Langmuir oscillations of free electrons.

LIMITING NONIDEALITY OF A METASTABLE CLASSICAL COULOMB PLASMA

a) System of equations. An analysis was made of the time evolution of a system of n electrons and n ions contained within a cube with absolutely rigid walls limiting the particle motion (for further details see Refs. 1 and 2). The length of the cube edge a was selected to give the defined charged particle density $N_e = N_i = n/a^3$.

We shall measure quantities having dimensions of length in units of the interparticle spacing $N_e^{-1/3}$, and quantities having dimensions of time in units of the reciprocal Langmuir frequency

$$\omega_L^{-1} = \sqrt{\frac{m_e}{4\pi e^2 N_e}},$$

where m_e is the electron mass and e is the electron charge.

The system of Newton equations for the electrons and ions, cast in dimensionless form is then

$$\frac{d^2\mathbf{x}_k}{d\tau^2} = \frac{1}{4\pi} \sum_{l\neq k} \frac{(-1)^m(\mathbf{x}_k-\mathbf{x}_l)}{|\mathbf{x}_k-\mathbf{x}_l|^3} G\left(\frac{|\mathbf{x}_k-\mathbf{x}_l|}{x_0}\right), \quad (1e)$$

$$\frac{d^2\mathbf{x}_k}{d\tau^2} = \frac{\mu}{4\pi} \sum_{l\neq k} \frac{(-1)^m(\mathbf{x}_k-\mathbf{x}_l)}{|\mathbf{x}_k-\mathbf{x}_l|^3} G\left(\frac{|\mathbf{x}_k-\mathbf{x}_l|}{x_0}\right). \quad (1i)$$

Here, $\mathbf{x}_k = \mathbf{r}_k N_e^{1/3}$ is the dimensionless position vector of the kth particle, $\tau = \omega_L t$ is the dimensionless time, $\mu = m_e/m_i$ is the ratio of the electron to the ion mass (two values of the ion mass were used in the calculations: the proton mass $\mu = 1/1836$ and the electron mass $\mu = 1$), where $m = 1$ for particles of unlike charge and $m = 2$ for particles of like charge, $G(x) = 1$ for $x > x_0$ and $G(x) = 8 - 9x + 2x^2$ for $x \leq x_0$ is a function limiting the Coulomb potential at close distances (the particles are assumed to be charged spheres of small radius $r_0/2$, $x_0 = r_0 N_e^{1/3}$), summation is performed over all $2n$ particles, and the index of the kth particle has values between 1 and n in the equations for the electrons (1e) and the ions (1i).

The initial conditions were defined using a pseudorandom number generator assuming a uniform electron and ion spatial distribution and a Maxwellian velocity distribution with the initial temperature T_0; the boundary conditions corresponded to specularly reflecting walls (for further details see Refs. 1 and 2).

The degree of ideality (or nonideality) of the plasma is characterized by

$$\gamma = \delta^{1/3} = \frac{e^2(2N_e)^{1/3}}{T_e},$$

$$\delta = \frac{2e^6 N_e}{T_e^3}, \quad n_D = \frac{1}{\sqrt{36\pi\delta}}, \quad (2)$$

where $n_D = 4\pi r_D^3 N_e/3$ is the number of particles of like charge in the Debye sphere, $r_D = \sqrt{T_e/(8\pi e^2 N_e)}$ is the Debye length, and $e^2 N_e^{1/3}$ serves as a natural unit of energy measurement.

Here the electron temperature was calculated as two thirds of the kinetic energy

$$T_e = \frac{2}{3} \sum_{k=1}^{k=n} \frac{m_e v_k^2}{2}, \quad (3)$$

where $v_k = dr_k/dt$ is the velocity of the kth electron.

Experience of previous calculations shows that this value is close to that calculated using the velocity distribution obtained.

Calculations were also made of the physical quantities for various ranges of the total electron energy ε. The total energy of an individual particle was given by:

$$\varepsilon_j = \frac{m_j v_j^2}{2} + q_j \sum_{k=1}^{2n} \frac{q_k}{|\mathbf{r}_j-\mathbf{r}_k|} G(|\mathbf{r}_j-\mathbf{r}_k|/r_0). \quad (4)$$
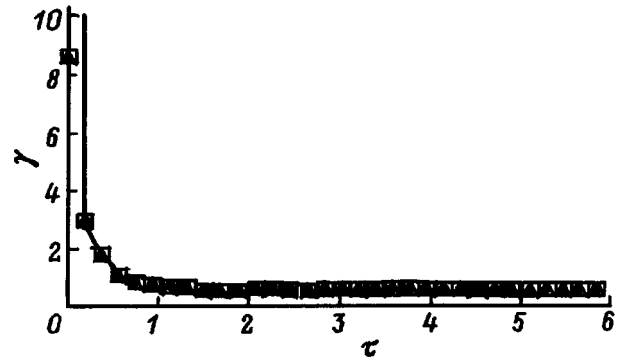

FIG. 1. Relaxation of the degree of plasma ideality γ to a quasi-steady-state value: $\square$ — $N_i = N_e = 10^{14}$ cm^{-3}, $T_0 = 0.001$ eV, $\triangle$ — $N_i = N_e = 10^{17}$ cm^{-3}, $T_0 = 0.01$ eV, $\bullet$ — $N_i = N_e = 10^{20}$ cm^{-3}, $T_0 = 0.1$ eV, solid curve— $\gamma = 0.5/\tau + 0.4$, total number of particles $2n = 1000$, and $\mu = m_e/m_i = 3 \times 10^{-7}$.

Here $G_1(x) = 1$ for $x \geq 1$, $G_1(x) = 2.4x - 4x^3 + 3x^4 - 0.4x^5$ for $x < 1$ is a function which takes into account the distortion of the Coulomb interaction potential when the distance between the charged particles is smaller than the sum of their radii r_0.

b) Universal relaxation curves. If a sufficiently large number of particles and a sufficiently low initial temperature are considered, the degree of ideality of the plasma will only be a function of the dimensionless time. This is supported by numerical calculations (Fig. 1) made for different densities N_e and different initial temperatures T_0, selected so that the initial ideality parameter for this series of calculations was the same. An analysis was also made of the situation where the ions may be considered to be fixed ($\mu \to 0$) and the initial average electron energy is low ($T_0 \to 0$). The barely noticeable difference between the results of the calculations for different densities is clearly determined by the calculation accuracy. The results of the numerical calculations of the relaxation of the average kinetic energy are well approximated by the simple expression

$$\gamma = \frac{a}{\tau} + b, \quad (5)$$

where a and b are the fitting parameters (for an H plasma $a \approx 0.5$ and $b \approx 0.35$).

Calculations were also made for different ion configurations (these were determined by defining different initial values in the pseudorandom number generator). It was show that for different ion configurations, the time dependence of the degree of ideality is accurately described by the same simple expression.

It follows from these calculations that by defining the initial conditions, it is impossible to obtain a plasma having an arbitrarily high degree of nonideality. Even if the initial state of a classical Coulomb plasma is strongly nonideal ($\gamma \gg 1$), within a comparatively short time interval of the order of the reciprocal Langmuir frequency, electron heating occurs and the plasma becomes closer to ideal. The limiting values of the parameters γ, δ, and n_D characterizing the degree of ideality of the steady-state metastable state are

$\gamma = 0.4 \pm 0.1$, $\delta = 0.076 \pm 0.05$, $n_D = 0.34$. For example, for the electron density $N_e = 10^{17}$ cm^{-3} the lower limit of the electron temperature is $T_e \approx 0.2$ eV.

The fact that the relaxation time is not determined by the Langmuir frequency does not contradict the results of Refs. 9 and 10 (see also Refs. 7 and 8), where the relaxation time is determined not by the Langmuir frequency, but by the electron transit time across the average interparticle spacing $\tau_{ei} = N_i^{-1/3}/\nu_e$. The fact is that, because of the strong heating in this particular case of low initial temperature, the electron velocity (and thus the interparticle transit time) is not determined by the initial low average velocity but by the energy of the Coulomb interaction and the Langmuir frequency

$$\nu_e \cong \sqrt{\frac{2e^2 N_e^{1/3}}{m_e}};$$

$$\tau_{ei} = N_i^{-1/3}/\nu_e \cong \sqrt{\frac{m_e}{2e^2 N_e}} = \sqrt{2\pi}\,\omega_L^{-1}. \tag{6}$$

When the initial temperature is high (the plasma is ideal), the time autocorrelation functions of the potential energy of the various electrons and the potential created by the moving particles at the center of the cube decay within the time τ_{ei} (for further details see Refs. 7–10).

The constraint on the degree of nonideality discussed here clearly applies to an isolated plasma not exposed to any external influence (for further details see Refs. 1 and 2). We could attempt to create a more nonideal plasma by cooling it by means of some external influence such as placing it in a thermostat with cold walls or simulating cooling, inelastic collisions. However, it was shown earlier[6–8] (see also Refs. 1 and 2), that an external stochastic influence with energy exchange impairs the metastable state and stimulates relaxation processes, particularly recombination. Further research is required to determine how a highly nonideal metastable plasma could be formed.

c) Theoretical calculation of the limiting nonideality parameter. We consider the case of equal masses of positively and negatively charged particles $m_i = m_e$ because the theoretical analysis put forward below can only be constructed for a completely steady-state metastable state. For substantially differing masses, it is impossible to investigate relaxation as far as the stage where the electron and ion temperatures become equalized.

Reasoning whereby particle diffusion and drift along the energy axis in the negative energy range are determined by microfields was used previously (see Refs. 1 and 2) to obtain the following expression for the particle distribution function in terms of the total energy $y = \varepsilon/T_e$:

$$f(y) = \frac{2C}{\sqrt{\pi}} \begin{cases} \sqrt{y}\,\exp(-y), & y > \alpha\gamma, \\ C_3 \exp(C_1 y + C_2 y^2/2), & |y| \leqslant \alpha\gamma, \\ C_4 \exp(\beta y/\gamma), & y < -\alpha\gamma. \end{cases}$$

Here

$$C_1 = [-1 + 1/(2\alpha\gamma) + \beta/\gamma + \beta/\gamma]/2,$$

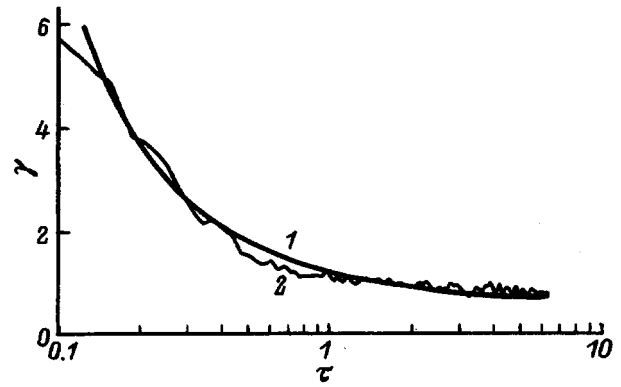$$C_2 = [-1 + 1/(2\alpha\gamma) + \beta/\gamma - \beta/\gamma]/(2\alpha\gamma),$$

FIG. 2. Relaxation of the degree of plasma ideality γ to a quasi-steady-state value in an ion–ion plasma, $\mu = 1$; the other values are the same as in Fig. 1: *1* — $\gamma = 0.55/\tau + 0.64$, *2* — calculations.

$$C_3 = \alpha^{1/2}\sqrt{\gamma}\,\exp[-\alpha\gamma(1 + C_1 + C_2\alpha\gamma/2)],$$

$$C_4 = \alpha^{1/2}\sqrt{\gamma}\,\exp[\alpha\beta - \alpha\gamma(1 + 2C_1)],$$

$$C^{-1} = 1 - (2/\sqrt{\pi}) \cdot F(3/2, \alpha\gamma) + (2C_3/\sqrt{\pi})$$

$$\times \int_{-\alpha\gamma}^{\alpha\gamma} \exp(C_1 y + C_2 y^2/2)dy$$

$$+ (2C_4\gamma/\sqrt{\pi} \cdot \beta)\exp(-\alpha\beta),$$

where $\alpha = 1.5$, $\beta = 0.4$, and $F(a,x) = \int_0^x e^{-t}t^{a-1}dt = \gamma(a,x)$ is an incomplete gamma function.

The function (1) can be used to find the total energy of the system per particle $\langle y \rangle = 3/2 + 2\langle u \rangle$. Here $\langle x \rangle = \int_{-\infty}^{\infty} dy\, x(y)f(y)$ is the average value of x and u is the potential energy per particle, measured in units of T_e.

Assuming that at zero time, both the kinetic energy and the potential energy are close to zero (this corresponds to a spatially uniform initial distribution with a low energy) and bearing in mind that energy is conserved as a result of specular reflection from the walls, in this case we obtain $\langle u \rangle = 3/4$ for the metastable state. The dependence of $\langle u \rangle$ on δ has already been tabulated. According to the calculations, this energy corresponds to $\delta = 0.26$, and therefore $\gamma = 0.64$. The results of a numerical simulation of an ion–ion plasma (Fig. 2) are accurately described by the function (5) with $a = 0.55$ and $b = 0.64$.

It has already been shown[1,2] that the function used above accurately describes the distributions obtained by first-principles modeling and yields expressions for the thermodynamic quantities which gives the results of the Debye theory in the weak Coulomb interaction limit. The agreement between the calculated limiting value of the nonideality parameter and the value obtained by direct modeling of many-particle dynamics indicates that the theoretical reasoning put forward previously accurately describes the new data.

PARTICLE PAIR CORRELATION FUNCTION

a) Calculation of the correlation functions. It is known that the pair correlation functions $w_{ab}(\mathbf{r}_a, \mathbf{r}_b)$ give the probability of particles a and b being found simultaneously near points $\mathbf{r}_a$ and $\mathbf{r}_b$. In a spatially uniform isotropic plasma,

these functions only depend on the interparticle spacing $r=|\mathbf{r}_a-\mathbf{r}_b|$. Thus, the correlation functions were calculated as follows.

An interval equal to half the cube edge length $(0<r<a/2)$ was divided into numerous (two hundred in the calculations) small intervals. In each time interval, we selected all the spacings r_{ij} between those particles for which the correlation function is determined (for example, by electrons and/or ions) and we calculated the number of particle incidences in each of these small intervals. Periodic boundary conditions were used to calculate r_{ij}. For instance, if it was found that the difference between the projections of the particle position vectors on the x axis was greater than half the cube edge length $x_i-x_j>a/2$, the value $x_i-x_j-a/2$ was taken instead. For $x_i-x_j</2$, the value $x_i-x_j+a/2$ was taken. We stress that these standard periodic boundary conditions were not used to solve the dynamic equations but merely to calculate the correlation functions.[1] It was therefore assumed that the entire space was filled with close-packed cubes with specularly reflecting walls. This avoids a nonphysical decay of the correlation functions over distances comparable with half the cube edge length.

As a result of this procedure, we calculated a quantity proportional to r^2w_{ab}. The normalization of the correlation function was determined by

$$4\pi\int_0^{a/2} r^2 dr w_{ab}(r) = \frac{4}{3}\pi\left(\frac{a}{2}\right)^3.$$

It was assumed that the number of incidences in a given small interval defines the value of the function r^2w_{ab} at the center of the interval.

Calculations were made of the electron–ion and electron–electron correlation functions. In addition, the electron–ion correlations were characterized by functions calculated for different energy ranges: a) free electrons with the energy $\varepsilon>1.5\gamma T_\varepsilon$, b) "quasicontinuum" electrons with $-1.5\gamma T_\varepsilon<\varepsilon<1.5\gamma T_\varepsilon$, and c) electrons which may be arbitrarily assumed to be bound with $-1.5\gamma T_\varepsilon>\varepsilon$.

b) Results of the calculations. The calculated electron–ion and electron–electron correlation functions were compared with the theoretical expressions obtained from the Bogolyubov theory (see, for example, Ref. 11),

$$w_{ab}=1+(-1)^m\frac{e^2}{T}\frac{\exp(-r/r_D)}{r}$$

$$=1+(-1)^m\gamma\frac{\exp(-2^{4/3}\sqrt{\pi}\gamma x)}{2^{1/3}x}. \tag{7a}$$

Here, as in Eq. (2), $x=|\mathbf{r}_a-\mathbf{r}_b|N_e^{1/3}$ is the dimensionless interparticle spacing, $m=1$ for particles of unlike charge and $m=2$ for particles of like charge, and the value of γ is determined by formula (3)

Generally speaking, expression (7a) is more suitable for the case of equal charged particle masses. Times for which screening by ions has not been established were considered for an electron–ion plasma. In this case, it is better to use the electron Debye length to describe the screening by the electrons:
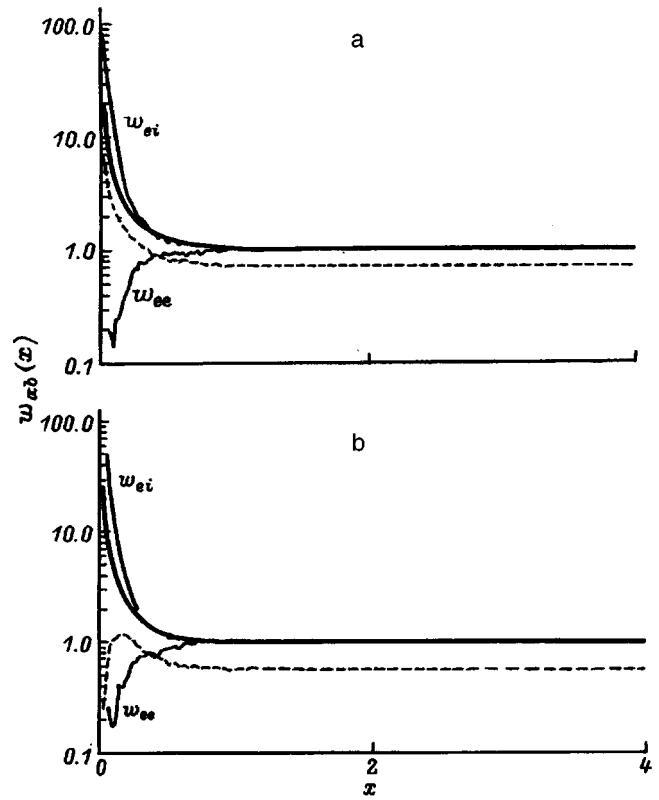

FIG. 3. Electron–ion (e–i) and electron–electron (e–e) pair correlation functions. The initial conditions are taken as the final time in the calculations of the establishment of the metastable state ($t_0\sim T_L$, where $T_L=2\pi/\omega_L=\sqrt{2\pi m_e/e^2 N_e}$), $x=rN_e^{1/3}$ $2n=1024$, $N_i=N_e=10^{17}$ cm^{-3}, the time of observation of the system evolution is $t=T_L$; a — ($m_i=1836m_e$), time-averaged electron temperature $T_e=0.17$ eV, heavy curve — $w_{ei}=1+(0.394/x)e^{-2.22x}$, dashed curve — distribution of electrons having the energy $\varepsilon>-1.5\gamma T_\varepsilon$; b — $m_i=m_e$, time-averaged temperature $T_e=0.12$ eV, heavy curve — $w_{ei}=1+(0.558/x)e^{-3.74x}$, dashed curve — the same as in Fig. 3a.

$$r_{D_e}=\sqrt{\frac{T_e}{4\pi e^2 N_e}}=\sqrt{2}\,r_D.$$

Thus, for the correlation function, we have:

$$w_{ab}=1+(-1)^m\frac{e^2}{T}\frac{\exp(-r/r_{D_e})}{r}$$

$$=1+(-1)^m\gamma\frac{\exp(-2^{5/6}\sqrt{\pi}\gamma x)}{2^{1/3}x}. \tag{7b}$$

A comparison between the calculated correlation functions and theoretical ones reveals reasonable agreement (Fig. 3). Some differences are observed for quite small distances ($x<0.5$), where bound particles make a substantial contribution. For an H plasma the distribution for particles of energy $\varepsilon>-1.5\gamma$ (free electrons and quasicontinuum electrons) almost agrees with the theoretical curve, apart from the shift caused by normalization. For particles of equal mass the distribution of free and quasicontinuum particles decreases for small distances because, for equal masses, there are more bound particles as a result of wall recombination (for further details see Ref. 12).
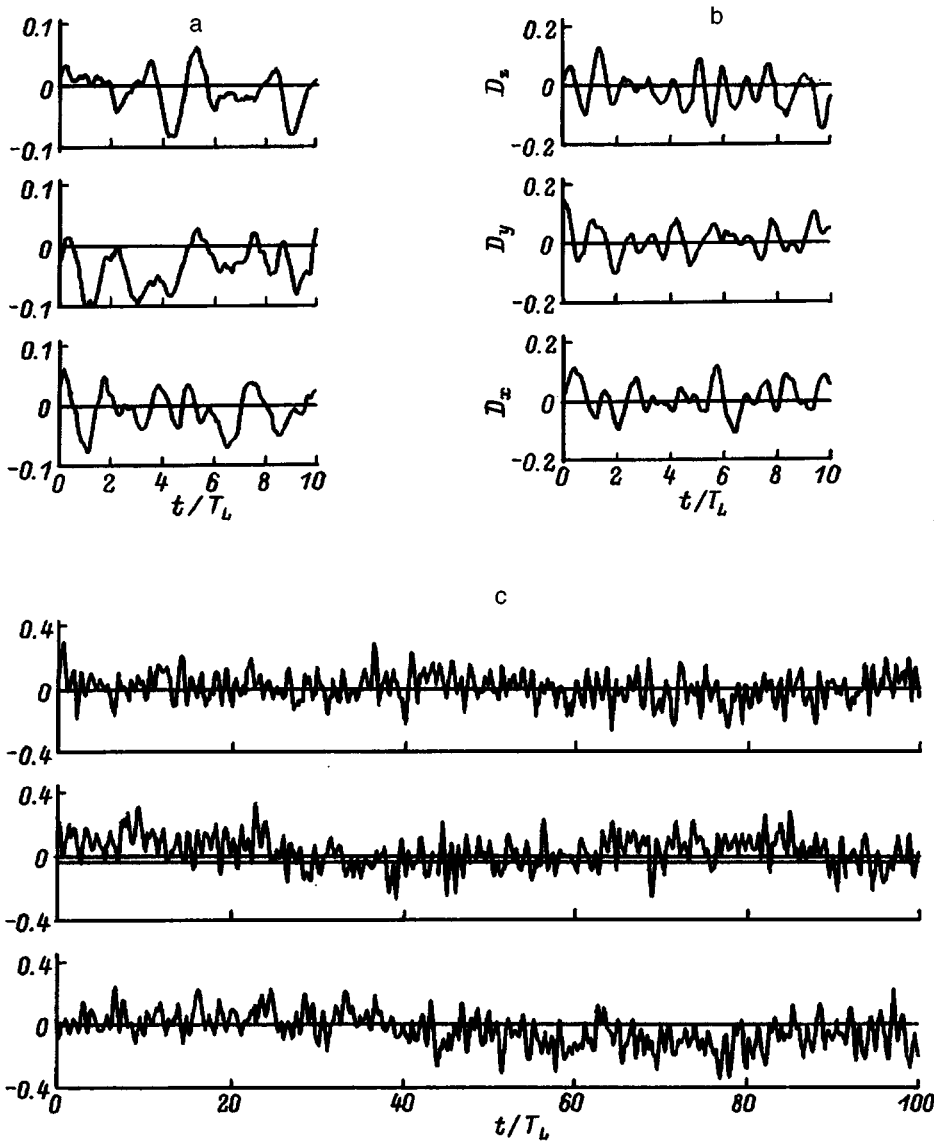
FIG. 4. Time dependence of the projections of the total dipole moment of a system of Coulomb particles on the x, y, and z axes (the moment is measured in units of $neN_e^{-1/3}$): a — short calculations, $m_i = 1836$, initial temperatures $T_{i0} = 0.22$ eV, $T_{e0} = 0.2$ eV, $2n = 1024$, $N_i = N_e = 10^{17}$ cm^{-3}; time of observation of the system evolution $0 < t < 10T_L$; during the calculations the electron temperature increased to $T_e = 0.25$ eV; b — short calculations, $m_i = m_e$, initial temperature $T_0 = 0.43$ eV, $2n = 1024$, $N_i = N_e = 10^{17}$ cm^{-3}, $0 < t < 10T_L$, $T_e = 0.45$ eV; c — long calculations, H-plasma, $T_{i0} = 0.22$ eV, $T_{e0} = 0.2$ eV, $2n = 128$, $N_i = N_e = 10^{17}$ cm^{-3}, $0 < t < 100T_L$; during the calculations the electron temperature increased to $T_e = 0.3$ eV.

Surprisingly, agreement between the results of the modeling and the theory is found outside the range of validity of the Debye model. First, the plasma is nonideal and second, agreement is found up to distances shorter than the Debye length and the average interparticle spacing. This observation agrees with the results of previous studies[9,10] (see also Refs. 7 and 8), where substantial "tightening" of the validity of the Debye concepts was observed: a) the Debye thermodynamic functions agreed with the calculated values up to $\delta \sim 1$ and b) the average potential over a large time interval, around a fixed sphere situated at the center of a cube, agreed with the Debye formula for screening even for short distances smaller than the average interparticle spacing.

DIPOLE MOMENT OF A SYSTEM OF COULOMB PARTICLES

1. Results of the calculations

a) Time dependence. Collective motion of the system electrons should show up particularly in the time dependence of the total dipole moment of the entire system of n positively charged and n negatively charged particles

$$\mathbf{D}(t) = \sum_{k=1}^{2n} q_k \mathbf{r}_k(t) = e\left(\sum_{j=1}^{n} \mathbf{r}_j^{(+)} - \sum_{k=1}^{n} \mathbf{r}_k^{(-)} \right). \qquad (8)$$

Here, $\mathbf{r}_j^{(+)}$ and $\mathbf{r}_k^{(-)}$ are the trajectories of the positively and negatively charged particles, respectively. The results of the calculations (Fig. 4) show that the plasma electrons undergo some ordered motion which may be treated as collective oscillations. More accurately, the center of gravity of the electrons (and the center of gravity of the ions in cases of equal mass) undergoes periodic oscillations, which become appreciable after summation over all the electron position vectors. Summation cancels out displacements as a result of chaotic motion and against the background of this chaotic motion, the regular component becomes appreciable.

The amplitude of the collective oscillations is small and corresponds to the displacement of each electron over a distance of the order of 1/20 of the average interparticle spacing. The characteristic time scale of the collective oscillations is of the same order of magnitude as the Langmuir frequency. However, these values are not exactly the same: in the several series of calculations carried out by us, the
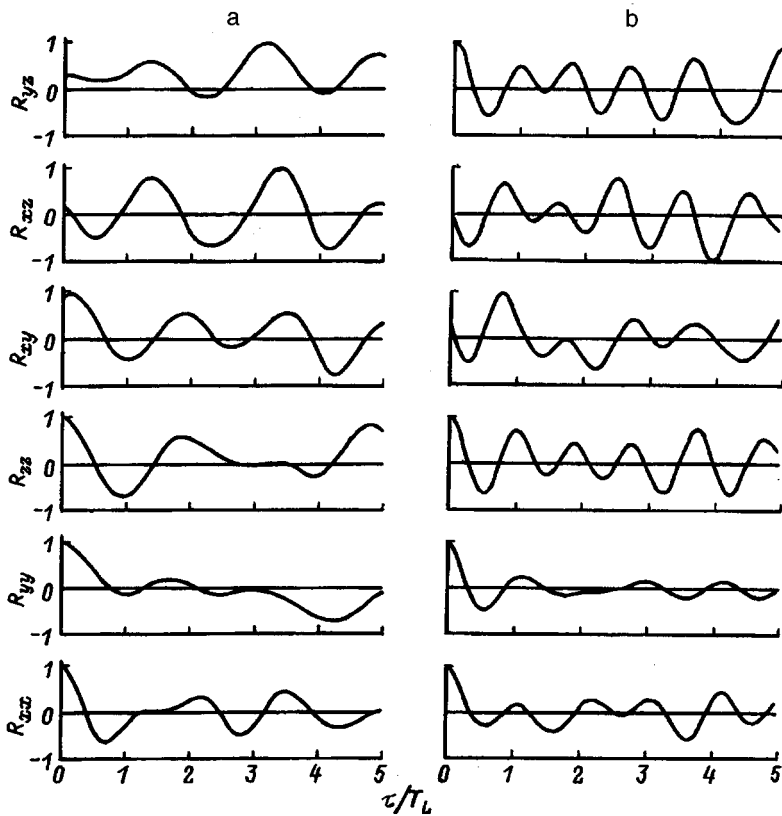
FIG. 5. Time dependence of the correlation functions of the projections of the total dipole moment of a system of Coulomb particles. The plasma and calculation parameters are the same as those in Figs. 4a and 4b; a — $m_i = 1836 m_e$, b — $m_i = m_e$.

frequency of the observed oscillations was 1.5–2 times lower than the frequency of the Langmuir oscillations for a H plasma. For equal masses, the frequency of the collective oscillations was approximately $\sqrt{2}$ times higher than that for heavy ions, as was to be expected, and was also 1.5–2 times lower than the frequency of the Langmuir oscillations which should be found for a plasma with equal masses of positive and negative particles (in this case, in the expression for the frequency of the Langmuir oscillations m_e should be replaced by the reduced mass $m_e/2$ of a particle pair).
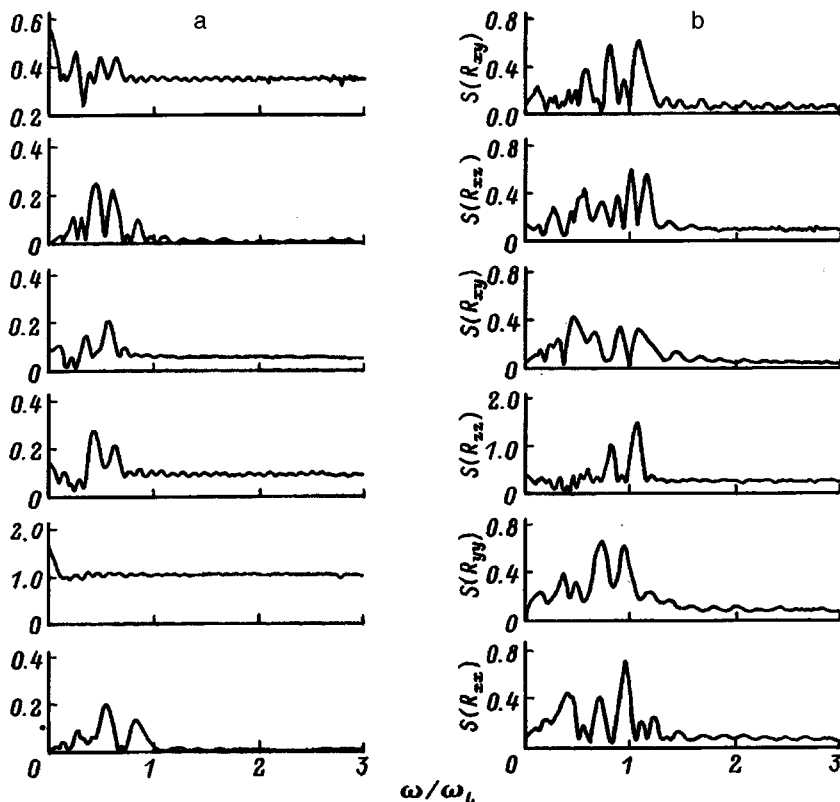

FIG. 6. Spectra of correlation functions of the dipole moment of the system. The plasma and calculation parameters are the same as those in Fig. 5: a — H plasma, b — $m_i = m_e$.

b) Dipole moment correlation functions. Although the time dependence of the dipole moment is conspicuously cyclic, objective characteristics are required to demonstrate the collective oscillations of the system. In this context, calculations were made of the time correlation functions of various projections of the dipole moment

$$R_{\alpha\beta}(\tau) = \frac{\frac{1}{t}\int_0^t D_\alpha(t')D_\beta(t'+\tau)dt'}{\max\left(\frac{1}{t}\int_0^t D_\alpha(t')D_\beta(t'+\tau)dt'\right)}. \quad (9)$$

Here α and β pass through the values x, y, and z corresponding to the projections of the dipole moment on different coordinate axes. In view of the symmetry with respect to α and β, it is sufficient to calculate only six of the nine correlation functions for each calculation. The results (Fig. 5) show that the correlation functions have a periodic character and the period of their oscillations is also slightly greater than the Langmuir period

$$T_L = \frac{2\pi}{\omega_L} = \sqrt{\frac{2\pi m_e}{e^2 N_e}}.$$

It is particularly noticeable that the cross correlation functions R_{xy}, R_{xz}, and R_{yz} are periodic, as are the autocorrelation functions R_{xx}, R_{yy}, and R_{zz}, with the periodicity of the cross correlation functions being even more defined. It may be stated that the vector of the dipole moment of the system of Coulomb particles behaves as a harmonic oscillator under an external influence.

c) Collective oscillation spectrum. The Fourier transformations of the correlation functions[2] were calculated to obtain the collective oscillation spectrum

$$S_{\alpha\beta}(\omega) = \left|\frac{1}{2\pi t}\int_0^l d\tau e^{i\omega\tau}R_{\alpha\beta}(\tau)\right| = \sqrt{\left(\frac{1}{2\pi t}\int_0^l d\tau R_{\alpha\beta}(\tau)\cos\omega t\right)^2 + \left(\frac{1}{2\pi t}\int_0^l d\tau R_{\alpha\beta}(\tau)\sin\omega t\right)^2}. \quad (10)$$

The spectra obtained (Fig. 6) demonstrate that collective oscillations are observed at a frequency slightly lower than the Langmuir frequency. However, the presence of noise and the comparatively few particles in the calculations means that the spectral line profile of the dipole radiation of a metastable plasma cannot yet be reliably determined

2. Collective oscillations

It should first be noted that in this energy range and for a nonideal plasma with $n_D < 1$, $\gamma \sim 1$, there is no small parameter required to construct a quantitative theory. The results of the numerical calculations must therefore be discussed on the basis of qualitative reasoning near the limits of validity of the adopted theoretical concepts.

a) Langmuir oscillations and dipole moment oscillations. For a plane longitudinal wave in a plasma the Vlasov equation yields a dispersion relation linking the wave number $k = 2\pi/\lambda$ (λ is the wavelength) with the frequency ω of the collective electron oscillations (see, for instance, Refs. 13 and 14)

$$\omega^2 = \omega_L^2 + 3k^2 v_{T_e}^2 + i\gamma_L\omega_L^2 = \omega_L^2 \cdot (1 + 6k^2 r_D^2 + i\gamma_L). \quad (11)$$

Here

$$\gamma_L = \frac{\sqrt{\pi}}{(2k^2 r_{D_e}^2)^{3/2}}\exp\left(-\frac{1}{2k^2 r_{D_e}^2}\right)$$

is the collisionless damping rate determined by Landau (the expression is given for a Maxwellian electron distribution). The dispersion equation (11) is valid for fairly long-wavelength Langmuir oscillations where

$$k^2 r_{D_e}^2 \sim k^2 v_{T_e}^2/\omega_L^2 \ll 1.$$

When this condition is satisfied, the real correction to the Langmuir frequency is small and the collisionless damping rate is exponentially small. All the collective electron oscillations take place at a frequency close to, but slightly higher than, the Langmuir frequency.

Thus, the small correction merely increases the oscillation frequency, whereas the simulation indicates that the frequency of the dipole moment oscillations is lower than the Langmuir frequency. This may be associated with the boundedness of the volume under study. However, for the cube edge length $a = (n/N_e)^{1/3} = 1.7 \times 10^{-5}$ cm, the correction to the Langmuir frequency in the dispersion equation is still fairly small

$$\omega = \omega_L\sqrt{1 + 6[(2\pi/a)r_D]^2} \approx 1.17 \cdot \omega_L.$$

It may well be that, for the Langmuir oscillations, the plasma studied cannot be considered to be ideal. Although the parameter $\delta = 2e^6 N_e/T_e^3$ characterizing the degree of ideality of the thermodynamic quantities is still small in this case, $\delta = 0.075$, the number of electrons in the Debye sphere is still less than one $n_D = (4\pi/3)r_D^3 N_e = 1/\sqrt{36\pi\delta} = 0.34$, and the Debye length is appreciably smaller than the average interparticle spacing $r_D N_e^{1/3} \sim 0.35$. Thus, if the Debye length in the estimate is replaced by the average interparticle spacing as the larger quantity, the correction to the oscillation frequency is then comparable with the Langmuir frequency $\omega \approx 2 \cdot \omega L$.

The fact that the opposite effect is observed — that the frequency of the collective oscillations is lower, not higher, than the Langmuir frequency — may be attributed to damping. It is known that for a harmonic oscillator with the natural frequency ω_0 and damping factor $\omega_0\alpha$, the oscillation frequency decreases with increasing α: $\omega = \omega_0\sqrt{1 - \alpha^2}$. In
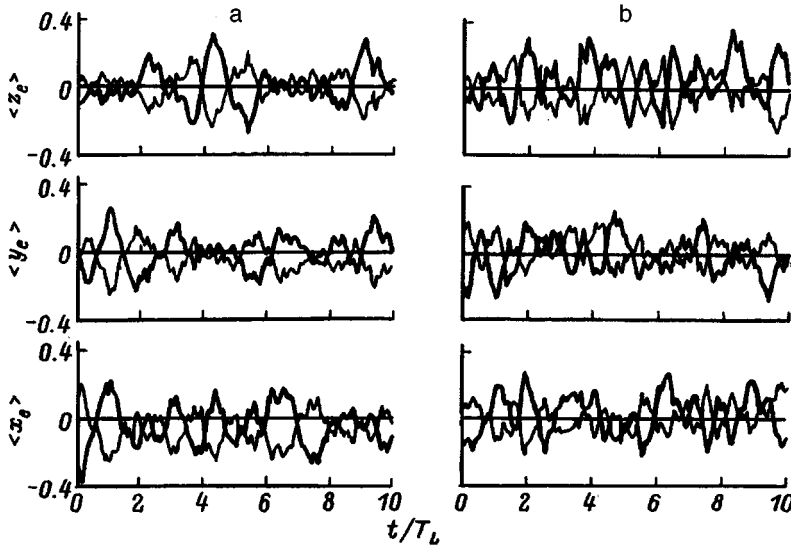
FIG. 7. Time dependence of the projections of the radius vector of the electron center of gravity on the x, y, and z axes. Solid heavy curves — center of gravity of free electrons, fine curves — center of gravity of electrons of energy $\varepsilon < 1.5\gamma T_\varepsilon$; the plasma and calculation parameters are the same as in Figs. 5a and 5b.

practice however, no decrease in the oscillation amplitude with time can be detected, even for a fairly long time interval (Fig. 5b). It may thus be postulated that the oscillations take place under the action of some external stimulating force.

The question naturally arises as to the nature of the damping which shifts the oscillation spectrum into the low-frequency range. Since the collisionless damping decrement is comparable to unity in the range of parameters studied, the decrease in the frequency of the electron collective oscillations may be caused by collisionless damping. However, this collisionless damping may be caused not by interaction between the wave and cophasal fast electrons, as is the case for Landau damping, but by interaction between the collective field and quasibound particles.

b) Kepler frequency. It is known that the period of particle rotation in a Coulomb field only depends on the binding energy $|\varepsilon|$ and specifically does not depend on the angular momentum. The motion then takes place with the Kepler angular frequency

$$\omega_\varepsilon = \sqrt{\frac{8|\varepsilon|^3}{e^4 m_e}} = \sqrt{\frac{2}{\pi}}\omega_L \left(\frac{|\varepsilon|}{e^2 N_e^{1/3}}\right)^3 .$$

Thus, when the binding energy of an electron is of the order of the interaction energy with other plasma electrons $|\varepsilon| \sim e^2 N_e^{1/3}$ (in this case the radius of the Kepler orbit is of the order of the average interparticle spacing), the electron rotates at a frequency of the order of the Langmuir frequency $\omega_\varepsilon \sim \omega_L$. This electron should therefore interact effectively with the collective oscillations of the plasma electrons — exciting oscillations in the plasma and being exposed to the action of the plasma oscillations.

The dispersion equation given above clearly cannot be used directly to describe the interaction between bound electrons and collective oscillations. This is because the characteristic size of the interaction zone (the radius of the electron orbit) is smaller than, or of the order of, the average distance between the charges. However, some results of numerical calculations can be put forward as indirect evidence to support the view that the oscillations of quasibound and free

electrons interact strongly. It can be seen from Fig. 7 that the centers of mass of the free and quasibound electrons oscillate in antiphase, so that the contribution of the free-electron oscillations to the dipole moment is compensated to a considerable extent by the reverse motion of electrons having the energy $\varepsilon < 1.5\gamma$. In our view, the antiphase nature of the oscillations indicates that the oscillations of the quasibound electrons excited by the Langmuir oscillations of free particles are forced.

c) Simplest model of the effect of antiphase oscillations. We first note that the main contribution to the oscillations of the dipole moment and the center of mass is made by perturbations of infinite wavelength (the entire spatial electron distribution is shifted in some direction). We postulate that the free electrons in the system are shifted by x_1. This shift leads to the generation of the electric field $E_1 = 4\pi e N_{e1} x_1$. For $N_{e1} = N_e$, i.e., when only free electrons are taken into account, the dynamics of the center-of-mass motion of the free electrons is determined by

$$m\ddot{x}_1 = -eE_1, \quad \text{or} \quad \ddot{x}_1 + \omega_L^2 x_1 = 0.$$

As was to be expected, the free electrons will oscillate at the Langmuir frequency. We now postulate that these free-electron oscillations cause a change in the motion of the bound electrons. In this case, the polarization of the bound electrons will partially compensate for the motion of the free electrons and the resulting field is

$$E = 4\pi N_e (\alpha_1 x_1 - \alpha_2 x_2).$$

Here x_2 is the shift of the bound electrons, and $\alpha_1 = N_{e1}/N_e$ and $\alpha_2 = N_{e2}/N_e$ are the fractions of free and bound electrons, respectively. The equation of motion for an individual bound electron then has the form

$$\ddot{x} - F/m_e = -eE,$$

where F is the projection of the Coulomb force acting on an individual electron, on the x axis.

If, for simplicity, a bound electron in the absence of an external field is treated as an oscillator oscillating at the Ke-

pler frequency, the equations of motion for the centers of mass of the bound and free electrons have the form

$$\ddot{x}_1 = -\omega_L^2(\alpha_1 x_1 - \alpha_2 x_2),$$

$$\ddot{x}_2 + \omega_\varepsilon^2 = -\omega_L^2(\alpha_1 x_1 - \alpha_2 x_2).$$

We now consider the simplest situation where few bound electrons participate in the collective oscillations and the shift of their center of gravity relative to the equilibrium position is small. Then, assuming $\alpha_1 = 1$ and $\alpha_2 = 0$, we obtain

$$x_1 = a_1 \cos(\omega_L t + \varphi_1),$$

$$x_2 = a_2 \cos(\omega_\varepsilon t + \varphi_2) - \frac{a_1}{1 - \omega_L^2/\omega_\varepsilon^2} \cos(\omega_L t + \varphi_1).$$

Here a_1 and a_2 are the amplitudes, and φ_1 and φ_2 are the initial phases of the oscillations. The ''minus'' sign in front of the second term in the second expression reflects the fact that the free and bound electrons oscillate in antiphase. This antiphase behavior is most noticeable for electrons close to resonance, i.e., for those rotating at the Kepler frequency close to the Langmuir frequency. The binding energy of these electrons is $\varepsilon \approx 1.9 \cdot e^2 N_e^{1/3}$.

The analysis can evidently be made more complex: a) by performing more accurate averaging which gives oscillating equations for the centers of gravity, b) by introducing damping, and c) by accurately solving the oscillating equations (without using the approximation of few bound electrons), and so forth. However, it is not yet clear whether this will give more accurate quantitative results. Nevertheless the essence of the antiphase oscillation effect can be seen from our very simple model.

CONCLUSIONS

The main conclusions of this study are given as follows.

1. At a low initial electron temperature corresponding to a strongly nonideal initial plasma state, a stage of temperature relaxation to a higher value takes place. This relaxation stage is characterized by the universal function (5).

2. There is a limiting degree of plasma nonideality ($\gamma \approx 0.4$) which may be achieved in the metastable state in the absence of any external influence.

3. The correlation functions of the positions of the charged particles obtained by calculations of many-particle dynamics reveal reasonable agreement with the expressions obtained using the Bogolyubov theory in the Debye approximation, where agreement between the results of the simulation and the theory is found outside the limits of validity of the Debye model: for a nonideal plasma and up to distances smaller than the Debye length and the average interparticle spacing. This observation agrees with the results of previous studies[9,10] (see also Refs. 7 and 8) where substantial ''tightening'' of the validity of the Debye concepts was observed.

4. The results of calculations of the total dipole moment of a system of charged particles show that the center of gravity of the electrons (and, if the masses are equal, the center of gravity of the ions) undergoes periodic oscillations. The characteristic time scale of the collective motion is of the

same order of magnitude as the Langmuir frequency. However, these values do not agree exactly: the frequency of the observed oscillations is 1.5–2 times lower than the Langmuir frequency. Moreover, the centers of mass of the free electrons and the quasibound electrons of energy $\varepsilon < 1.5\gamma T_\varepsilon$ oscillate in antiphase, so that the contribution of the free electron oscillations to the dipole moment is compensated to a considerable extent by the reverse motion of the quasibound electrons.

The results of calculations of the time correlation functions show that these are also periodic and their period coincides with the period of the oscillations of the total dipole moment of the system. The cross correlation functions of the different projections of the dipole moment are also periodic, like the autocorrelation functions, and the periodicity of the cross correlation functions is even more clearly defined.

The spectra of the correlation functions also demonstrate that collective oscillations take place at a frequency slightly lower than the Langmuir frequency.

5. On the basis of these results, it may be hypothesized that the decrease in the oscillation frequency compared with the Langmuir frequency is caused by collisionless damping which takes place when oscillations of quasicontinuum electrons describing Kepler orbits, whose radii are of the order of the interparticle spacing, interact with collective oscillations of free plasma electrons. This allows us to hypothesize that the observed metastable recombination behavior of a system of classical Coulomb particles is caused by interaction between quasicontinuum electrons and collective oscillations of free electrons.

[1] If periodic boundary conditions are used to solve the dynamic problem, this is stated separately. For a discussion of the unreliability of using periodic boundary conditions to investigate the fundamental properties of a system of many Coulomb particles see Ref. 6.

[2] Calculations of the Fourier components directly from the projections of the total dipole moment did not give interesting results: the spectrum disappeared almost completely against the noise background.

[1] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Usp. Fiz. Nauk **164**, 298 (1994).

[2] S. A. Mayorov, A. N. Tkachev, and S. I. Yakovlenko, Phys. Scripta **51**, 498 (1994).

[3] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Pis'ma Zh. Tekh. Fiz. **17**(23), 33 (1991) [Sov. Tech. Phys. Lett. **17**, 836 (1991)].

[4] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Mat. Model. No. 7, 3 (1992).

[5] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Izv. Vyssh. Uchebn. Zaved. Fiz. **35**(11), 76 (1992).

[6] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Izv. Vyssh. Uchebn. Zaved. Fiz. **34**(11), 3 (1991).

[7] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Izv. Vyssh. Uchebn. Zaved. Fiz. **35**(2), 10 (1992).

[8] Tr. Inst. Obshch. Fiz. Ross. Akad. Nauk **46**, 142 (1992).

[9] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Kratk. Soobshch. Fiz. No. 12, 33 (1987).

[10] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Pis'ma Zh. Tekh. Fiz. **14**(4), 354 (1988) [Sov. Tech. Phys. Lett. **14**, 156 (1988)].

[11] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Parts 1 and 2, 3rd ed. (Pergamon Press, Oxford, 1980) [Russian original, Parts 1 and 2, 3rd ed., Nauka, Moscow, 1976].

[12] S. A. Maĭorov and S. I. Yakovlenko, Izv. Vyssh. Uchebn. Zaved. Fiz. **37**(11), 44 (1994).

[13] V. D. Shafranov, *Questions of Plasma Theory* [in Russian], Part 3, Gosatomizdat, Moscow (1963), pp. 3–140.

[14] B. B. Kadomtsev, *Collective Phenomena in a Plasma* [in Russian], Nauka, Moscow (1976).

Translated by R. M. Durham

Characteristics of sound wave generation in a gas discharge

A. R. Aramyan and G. A. Galechyan

Institute for Applied Problems in Physics, Academy of Sciences of Armenia, 375014 Erevan, Armenia

An analysis is made of the generation of sound waves in a dc gas-discharge tube by applying an alternating component, and the optimum generation conditions are determined. © *1997 American Institute of Physics.* [S1063-7842(97)00708-3]

Continuous control of the plasma parameters at constant electric current and gas pressure in a discharge is one of the major problems in plasma physics and the application of gas discharges. An effective method of varying the plasma parameters is to generate acoustic waves in the plasma.[1–3]

Sound waves propagating along a positive column may compress and expand the plasma column, they may induce the formation of ionization waves,[4,5] they may create a uniform distribution of excited gas at high pressures, and may alter the electron and gas temperature in the plasma, the electric field, and other parameters over a wide range.[1–3]

Sound waves can be excited in a gas discharge by two methods: 1) by introducing sound into the discharge chamber from outside, by means of an electrodynamic generator clamped to one end of the tube, 2) by generating sound waves inside the chamber and adding an alternating component to the dc discharge.[6]

The results of experimental investigations of the influence of sound waves on the plasma parameters obtained by these two methods show that they completely agree.[1] The second method (alternating current) is the more effective method of generating sound waves in a discharge chamber. In many cases, this method is the only one, particularly for controlling the plasma parameters in gas-discharge lasers, in which the optical cavity mirrors are mounted at the ends of the tube.

Here we consider the processes of sound wave generation in a dc gas-discharge tube by the application of an alternating component and we determine the optimum generation conditions. A method is proposed to measure the gas temperature in a low-temperature plasma by determining the resonant frequency of sound in the discharge chamber.

It has been shown that the maximum amplitude of a sound wave generated (at the resonant frequency) in a discharge tube by an alternating current component depends on the ratio of the discharge length to the tube length (at constant pressure). The maximum amplitude is established when the length of the plasma column is half the tube length and the discharge is located in one half of the chamber. In this case, matching of the resonances of the two cavities (one nested inside the other) leads to an increase in the amplitude of the sound wave.

The investigations were carried out using the apparatus shown schematically in Fig. 1. The measurements were made in a quartz discharge tube *1* with an inner diameter of 6 cm and length 100 cm. Six tungsten electrodes 1 cm in diameter were soldered to the side wall, three on one side

and three on the opposite side, as shown in Fig. 1. The distances between the electrodes were 16, 26, 53, 69, and 85 cm. A high dc voltage was supplied from a power supply *3* via a high-voltage transformer winding *4* to create a plasma at the electrodes. A sound wave was generated in the discharge tube by sine-wave modulating the dc current by means of a discharge current modulation unit, consisting of a generator *5*, an amplifier *6*, and a transformer *4*. A microphone *7* was attached to one of the plane ends of the tube and the signal was fed to an oscilloscope *8*. The discharge tube was evacuated by a pump *9* to a vacuum of 10^{-2} Torr and was then filled with gas to the working pressure, which was measured with a U-shaped mercury manometer *10*.

A gas discharge was created at various distances between the electrodes at a fixed argon pressure of 80 Torr and a dc current of 50 mA. When the alternating component was switched on (applied to the dc current), a sound wave began to be generated in the discharge tube and was detected with the microphone. The measurements were made at frequencies near the first resonance at which the tube length corresponded to half the wavelength.

Figure 2 gives the amplitude of the sound wave in the tube as a function of the length of the discharge gap at various frequencies near resonance, from which it can be seen that the largest amplitude of sound is established at a resonant frequency of 184 Hz and a plasma column length of 53 cm, i.e., approximately half the length of the discharge tube. This corresponds to a quarter of the wavelength λ. When the length of the discharge region is greater or less than $\sim 0.25\lambda$, the amplitude of the sound wave decreases. This effect can evidently be explained as follows. If a discharge of length 0.25λ is positioned at the end of the tube, a cylindrical cavity is formed with one open end.[7] From the open end of the cavity, sound is emitted into that part of the tube where there is no discharge, i.e., into the other cavity. When their resonances are the same, the amplitude of the sound wave shows the greatest increase.

Thus, optimum conditions for the generation of a sound wave by an alternating current component are established in a discharge tube in which the length of the plasma column is half the tube length and this discharge is located at the end of the tube.

It should be noted that generating the discharge at the center of the 53 cm long tube (between the second and fifth electrodes) did not produce such abrupt amplification of the sound as between the first and fourth electrodes, and an appreciably lower sound amplitude was established, i.e., in this
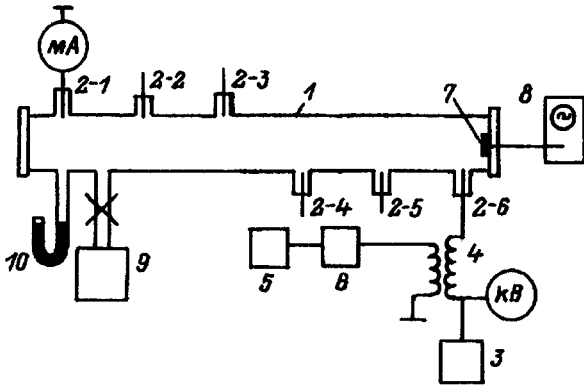
FIG. 1. Schematic of apparatus: *1* — quartz discharge tube, *2* — electrodes, *3* — high-voltage dc source, *4* — transformer, *5* — sound generator, *6* — amplifier, *7* — microphone, *8* — oscilloscope, *9* — vacuum pump, and *10* — manometer.

FIG. 3. Amplitude–frequency characteristic of a sound wave in an argon gas discharge for interelectrode gaps of 85 (*1*), 53 (*2*), and 26 cm (*3*) in a tube of length 100 cm with an inner diameter of 6 cm at 80 Torr pressure and 50 mA current.

case, only the resonance corresponding to the tube length was effective.

Figure 3 gives the sound wave amplitude as a function of the frequency near resonance for discharge gaps of 26, 53, and 85 cm, from which it can be seen that when the current and gas pressure in the tube are fixed, different discharge gaps correspond to different resonant frequencies of the sound wave. For instance, for gaps of 85, 53, and 26 cm, the resonant frequencies are 188, 184, and 180 Hz. This dependence of the sound wave resonant frequency on the length of the discharge gap at constant pressure and current can be explained as follows. As the discharge gap increases, the discharge voltage increases at the electrodes and this is accompanied by a rise in the gas temperature in the tube. This leads to an increase in the sound wave velocity which is accompanied by an increase in the resonant frequency.

We shall now determine the average gas temperature in an argon discharge for the three discharge gaps given in Fig. 3 for constant argon pressure and current (80 Torr and 50 mA) using the resonant frequencies and the well-known expression[8]

$$c_s = \sqrt{\gamma \frac{RT}{\mu}},$$

where $\gamma = \dfrac{c_p}{c_v}$, R is the gas constant, and μ is the molecular weight.

For a discharge gap $l_1 = 85$ cm and a resonant frequency $f_1 = 188$ Hz, the velocity of sound is $c_s = 376$ m/s and the gas temperature $T_g = 375$ K, for $l_2 = 53$ cm and $f_2 = 184$ Hz, we find $c_s = 368$ m/s and $T_g = 359$ K, and for $l_3 = 26$ cm and $f_3 = 180$ Hz, we find $c_s = 360$ m/s and $T_g = 343$ K.

The gas temperature is one of the main parameters of a partially ionized plasma. It depends on the frequency of elastic collisions between electrons and gas atoms, since most of the electron energy, and specifically $2m/M$ (where m is the electron mass and M is the atomic mass) is transferred to neutral particles.

An increase in the discharge current is accompanied by an increase in the electron density and in the frequency of their collisions with atoms, which leads to an increase in the gas temperature in the plasma. A sound wave propagating along a positive column reduces the gas temperature in the plasma (at constant discharge current and pressure)[1–3] and allows it to reach a particular value required for a specific application or for a particular experiment.

Measurement and monitoring of the gas temperature in the plasma during an experiment is one of the most important tasks in studies of gas discharges or their practical application. It is particularly valuable to develop and monitor the gas-discharge parameters of lasers in which the gas temperature in the plasma determines the optimum lasing conditions. Any change in this parameter leads to a substantial drop in the laser radiation intensity (such processes are typical of carbon dioxide lasers, metal vapor lasers, and so forth).
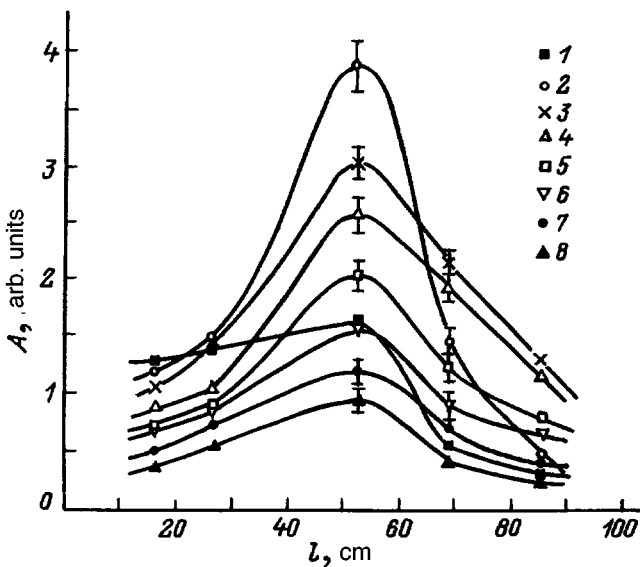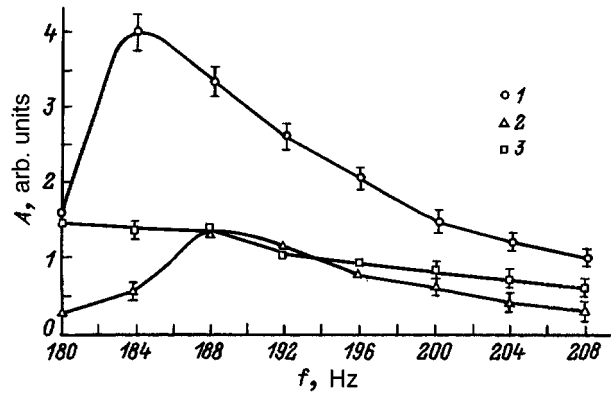

FIG. 2. Amplitude of sound wave versus length of discharge gap at an argon pressure of 80 Torr, current of 50 mA, and different frequencies near resonance: *1* — 180, *2* — 184, *3* — 188, *4* — 192, *5* — 196, *6* — 200, *7* —204, and *8* — 208 Hz.

The most widely used method of measuring the gas temperature in a low-temperature plasma is the thermocouple method, which has various disadvantages. These include: response time, contact, i.e., some foreign body — a thermocouple — must be inserted in the plasma at a suitable point to measure the gas temperature, and locality — the gas temperature is only measured in a small region where the thermocouple is located.

Methods of measuring the gas temperature in a plasma by determining the velocity of sound were considered in Ref. 9, where the velocity of sound was determined from the measured time taken for propagation of an acoustic pulse between two identical Langmuir probes positioned a specific distance apart in the plasma column. In another experiment[10] the sound wave velocity was determined by an optical method. The velocity of sound was studied in Ref. 11 by inducing localized breakdown in a discharge in various gases.

The method of measuring the gas temperature in a plasma by determining the resonant frequency of a sound wave in the discharge tube (described above) does not have the disadvantages inherent in the thermocouple method and others, it is simple to implement, can easily be automated, and can give information on the average gas temperature in the discharge, which is very useful in many cases, particularly for gas lasers. The resonant frequency can be determined very accurately from the amplitude-frequency characteristic. In addition, the resonant frequency can also be determined from the profile of the sound wave oscillations (the microphone signal), i.e., at resonance, the oscillations are sinusoidal since the dispersion is zero. At nonresonant frequencies, the profile of the oscillations differs appreciably from sinusoidal. It should be noted that when sound is excited in a discharge by an alternating current whose oscillation profile differs substantially from sinusoidal, sinusoidal sound waves are established (in resonance).

It is desirable to measure the gas temperature in a weakly ionized plasma in resonance with the frequency of the sound by using a low intensity (up to 75 dB) in order to avoid strong interaction between the acoustic waves and the plasma. This is because when the sound wave intensity increases above 80–85 dB, the gas temperature in the plasma column decreases under the influence of the sound,[1-3] and this is accompanied by a decrease in the velocity of sound. This leads to a decrease in the resonant frequency which must be continuously tracked by a frequency change at the master oscillator.

Thus, the interaction between the sound wave and the plasma operates on a feedback principle, i.e., an increase in the sound intensity causes a decrease in the gas temperature in the plasma (as a result of the formation of vortex motion which transfers heat to the walls), which leads to a change in the sound wave velocity and the resonant frequency. In other words, the parameters of the sound and the plasma vary consistently. This process is clearly demonstrated by the interaction between a sound wave pulse and a plasma which we shall analyze briefly.

Figure 4 gives the amplitude-frequency characteristic of an acoustic cavity consisting of a quartz tube 100 cm long
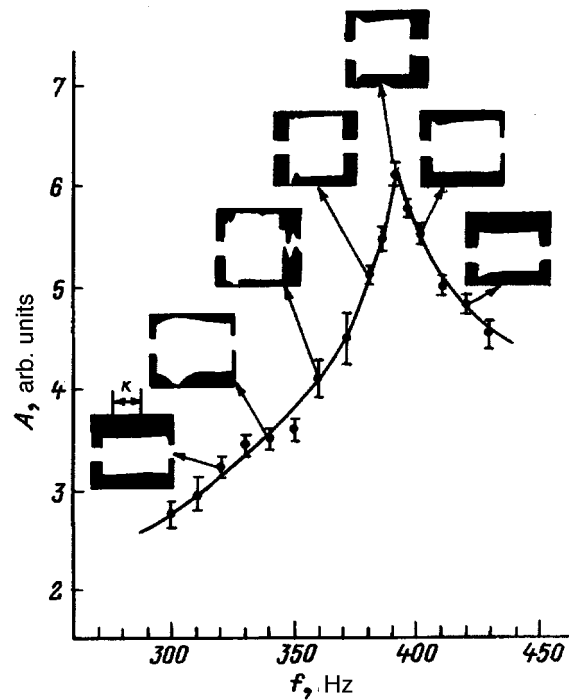

FIG. 4. Amplitude–frequency characteristic of acoustic cavity formed by a discharge tube 6 cm in diameter and 100 cm long at an argon pressure of 120 Torr and discharge current of 50 mA, with oscilloscope traces of sound wave pulse.

with an inner diameter of 6 cm, for an argon discharge at a pressure of 120 Torr and current of 50 mA. The characteristic was obtained for a constant sound wave intensity supplied by an emitter attached to the end of the tube, in the frequency range shown. This intensity was low, i.e., it was such that no interaction took place between the wave and the plasma. The maximum of the sound intensity corresponds to a resonance at 390 Hz, for which the wavelength is equal to the length of the discharge tube. Figure 4 shows oscilloscope traces of the sound wave pulse propagating along the positive discharge column and reaching the microphone. The intensity of the sound wave supplied from the emitter to the discharge is constant in all the traces.

It can be seen from the trace at 320 Hz that the pulse amplitude remained constant over this length, i.e., the wave amplitude is not exposed to any plasma influences. The pulse length was ∼2.5 s and the sound wave intensity was 84 dB. The trace at 340 Hz shows that the amplitude of the sound in the tube increased, although the intensity of the wave from the emitter was the same as that at 320 Hz. In addition, and this may well be the main factor, the wave amplitude begins to increase with time from the beginning of the pulse and reaches a maximum after approximately 0.7 s, before decaying. The wave intensity under these conditions was 86 dB. The amplitude of the sound wave increases as it propagates along the discharge because its wavelength decreases and the conditions in the tube approach resonance. This decrease in wavelength occurs because the sound wave induces vortex motion in the discharge, which increases the transfer of heat from the discharge to the tube walls, reduces the gas temperature, and decreases the velocity of sound. On the third

A. R. Aramyan and G. A. Galechyan   903

trace, obtained at 360 Hz, the amplitude increases to its maximum within 0.3 s since 360 Hz is closer to the 390 Hz resonant frequency and the time taken to reach it is shorter as a result of the decrease in wavelength. It should be noted that on the trace at 360 Hz the amplitude of the sound wave is approximately 1.3 times that at 340 Hz and this has the result that the gas absorbs that amount of energy which relaxes at a frequency of 2 kHz after the sound is switched off. These oscillations were indicated at the end of the trace. The sound intensity under these conditions was 88 dB. On the fourth trace, obtained at 390 Hz, the amplitude of the sound wave reaches a maximum 0.1 s after the beginning of the pulse and then decays. The wave amplitude at 380 Hz is lower than that in the third trace at 360 Hz and corresponds to 87 dB. This circumstance arises because as a result of interaction with the plasma, the sound wave becomes so retuned that it is far from resonance to the right of the maximum. On the oscilloscope trace obtained at the resonant frequency the amplitude of the wave pulse decreases almost from the very beginning. This occurs because as a result of negligible cooling of the gas, the wavelength becomes shorter than the tube length, the resonance condition is violated, and the sound amplitude decreases. The processes on the next two traces at 400 and 420 Hz can be explained as for 390 Hz.

These experimental results indicate that the relative influence of the sound wave and the plasma becomes stronger, the closer the sound frequency to the resonant frequency.

The following conclusions may be drawn:

1. It has been established that the optimum conditions for the generation of a sound wave in a dc discharge are formed when an alternating current (with a resonant frequency) is superposed on this current, the length of the plasma column is half the tube length, and the discharge is located at the end of the chamber. In this way, two cavities are created, one (a cylindrical cavity with an open end) acting inside the other. When the two resonances are matched, the amplitude of the sound wave increases.

2. A method has been proposed to measure the gas temperature in a low-temperature plasma by determining the resonant frequency of a sound wave propagating along the discharge tube.

[1] A. R. Aramyan, G. A. Galechyan, and A. R. Mkrtchyan, Akust. Zh. **37**, 213 (1991).
[2] M. A. Antinyan *et al.*, Laser Phys. **2**, 685 (1992).
[3] G. A. Galechyan, Usp. Fiz. Nauk. **165**, 1357 (1995).
[4] S. Subertova, Czech. J. Phys. B **15**, 701 (1965).
[5] A. R. Mkrtchyan *et al.*, Acustica **69**, 124 (1989).
[6] G. A. Galechyan, A. R. Aramyan, and A. R. Mkrtchyan, Zh. Tekh. Fiz. **60**, 207 (1990) [Sov. Phys. Tech. Phys. **35**, 264 (1990)].
[7] Z. I. Avdus' *et al.*, *Practical Course in General Physics* [in Russian], Prosveshchenie, Moscow (1971), p. 85.
[8] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon Press, Oxford, 1976) [Russian original, 3rd ed., Nauka, Moscow, 1986].
[9] K. W. Gentle and U. Ingard, Appl. Phys. Lett. **5**, 105 (1964).
[10] C. G. Suits, Physics **6**, 190 (1935).
[11] L. Goldstein *et al.*, in *Proceedings of the Sixth International Conference on Ionization Phenomena in Gases, Paris, 1963*, Vol. 3, pp. 115–116.

Translated by R. M. Durham

Stochastic dynamics of a Josephson fluxon in an inhomogeneous junction

F. Kh. Abdullaev and É. N. Tsoĭ

Physicotechnical Institute, Fizika–Solntse Scientific–Industrial Consortium,
Academy of Sciences of Uzbekistan, 700084 Tashkent, Uzbekistan

A study is made of fluxon motion in a long Josephson junction with a local inhomogeneity (microresistor) in the presence of an alternating external current. Nonlinear resonances in the fluxon dynamics are considered. The process by which a fluxon trapped at an inhomogeneity is stochastically depinned under the influence of a periodic current is investigated analytically and numerically. © *1997 American Institute of Physics.* [S1063-7842(97)00808-8]

INTRODUCTION

The dynamics of Josephson vortices in long inhomogeneous junctions are attracting considerable attention among researchers because of the possible physical applications such as a parametric Josephson oscillator, tunable generators of submillimeter radiation, logic devices,[1,2] and so on. The analyses deal with the influence of two types of inhomogeneities on the fluxon motion: distributed (frequently in the form of periodic reliefs) and point (microshorts or microresistors) inhomogeneities. It was shown in Ref. 3 that a fluxon may become trapped alongside an inhomogeneity under the influence of a dc current. The establishment of stochastic fluxon dynamics under the influence of an alternating current was studied in Ref. 4. Stationary fluxon states in the field of two microshorts were studied in Ref. 5 while the stochastic acceleration (an analog of the Ulam problem) of a fluxon under the action of an oscillating impurity was investigated in Ref. 6. The analysis of stochastic fluxon dynamics under the action of an alternating current, reported in Ref. 4, involved studying a special type of mapping valid for motion near the separatrix on the phase plane of the fluxon parameters in the absence of damping. Here, we report an analytic and numerical study of fluxon dynamics in a wide range of nonlinear resonances and we obtain a criterion for stochastic motion of a Josephson fluxon, which also allows for dissipation effects.

This problem is also of interest for nonlinear wave theory. The mechanisms for the onset of dynamic chaos of nonlinear waves have not been sufficiently well studied and the proposed simple model may serve as a basis for comparing the theoretical conclusions and the experiment.

FLUXON MOTION IN THE FIELD OF A POINT INHOMOGENEITY

We consider the dynamics of a fluxon in a long Josephson junction under the action of an alternating external current. In standard dimensionless variables, the equation for the phase difference of the wave functions $\varphi(x,t)$ has the form[1–3]

$$\varphi_{tt} - \varphi_{xx} + \sin\varphi = -\alpha\varphi_t + \beta\varphi_{xxt} - F(t)$$

$$+ \left[\sum_{i=1}^{N} \varepsilon_i \delta(x - a_i)\sin\varphi_i \right], \quad (1)$$

where α and β are damping parameters associated with tunneling of normal electrons across the barrier and the flow of normal electrons parallel to the barrier, the term $F = F_0 \sin(\Omega_0 t)$ describes the influence of a distributed external current, and the spatial variable x and time t are normalized to the Josephson penetration length λ_J and the reciprocal plasma frequency $1/\omega_p$ of the long Josephson junction, respectively.

The junction has point inhomogeneities positioned at the distance a_i. The case $\varepsilon_i < 0$ corresponds to microshorts while $\varepsilon_i > 0$ corresponds to microresistors. We shall subsequently study the case $\varepsilon_i > 0$, where the distance between inhomogeneities is assumed to be much greater than the fluxon dimension, i.e., $a_i \gg d$, where $d \sim (1 - v^2)^{1/2}$. Then, the interaction between the fluxon and each inhomogeneity can be analyzed separately, since the influence of the neighboring inhomogeneity can be neglected. For small ε, γ, β, and $F \ll 1$, a solution may be sought in the form of a kink (antikink) with variable parameters (the center coordinate $\zeta(t)$ and the velocity $v(t)$)

$$\varphi(x,t) = 4\arctan\left[\exp\left(\sigma\frac{x - \zeta(t)}{\sqrt{1 - v^2(t)}}\right)\right], \quad (2)$$

where $\sigma = \pm 1$ is the polarity corresponding to a kink (antikink).

Applying perturbation theory to the sine–Gordon equation,[3,7] we can derive equations for $\zeta(t)$ and $v(t)$

$$\frac{dv}{dt} = -\alpha v(1 - v^2) - \beta v/3 + (\pi\sigma F_0/4)\sin(\Omega_0 t)$$

$$\times (1 - v^{3/2}) - (\varepsilon/2)(1 - v^2)\sinh(y)/\cosh^3(y),$$

$$\frac{d\zeta}{dt} = v + (\varepsilon/2)v\zeta\,\sinh(y)/\cosh^3(y), \quad y = \zeta/(1 - v^2)^{1/2}. \quad (3)$$

We shall subsequently investigate the case of low velocities ($v^2 \ll 1$), since this is of interest for small perturbations. In this case, system (3) may be written as[3,4]

$$\frac{d^2\zeta}{d\tau^2} + \frac{\sinh(\zeta)}{\cosh^3(\zeta)} = -\Gamma\frac{d\zeta}{d\tau} + i\,\sin(\Omega\tau), \quad (4)$$

where $\Gamma=(\alpha+\beta/3)/\nu_0$, $f=\pi\sigma F_0/(4\nu_0^2)$, $\Omega=\Omega_0/\nu_0$, $\tau=t\nu_0$, and $\nu_0=(\varepsilon/2)^{1/2}$ is the frequency of small oscillations of the kink alongside an inhomogeneity.

Thus, the problem of fluxon dynamics in a long Josephson junction reduces to studying the motion of a particle of unit mass in the field of an attractive anharmonic potential $U(\zeta)=-(1/2)\mathrm{sech}^2\zeta$ under the influence of an external variable force. Note that in addition to changes in the fluxon parameters, small perturbations also cause the fluxon to emit linear electromagnetic waves. However, at low velocities the radiation energy is low ($\sim\exp(-1/v^2)$) and equations in the adiabatic approximation may be used for times $t<f^{-2}$. An equation similar to (4) is also obtained by studying the interaction between a soliton of the nonlinear Schrödinger equation and an oscillating impurity.[8] In addition, this problem is similar to that of the beam propagation in a medium whose refractive index depends on the spatial coordinates.[9]

ANALYSIS OF FLUXON MOTION IN THE ABSENCE OF DISSIPATION. CRITERION FOR STOCHASTICITY OF THE MOTION

In this section we use Eq. (4) to study fluxon motion for the case of negligibly small dissipation when $\Gamma=0$ can be assumed. The Hamiltonian of the system can then be written as follows:

$$H=H_0+fH_1$$
$$=(1/2)[(d\zeta/d\tau)^2-\mathrm{sech}^2(\zeta)]-f\sin(\Omega\tau)\zeta. \qquad (5)$$

We first consider the main features of the unperturbed particle motion ($f=0$). The phase portrait for this case is given in Fig. 1. The separatrix (curve 1) determined by the relations

$$\zeta=\pm\mathrm{arcsinh}(\tau), \quad v=\pm(1+\tau)^{-1/2} \qquad (6)$$

separates the region of periodic (finite) and drift (infinite) trajectories (curves 2 and 3, respectively). To analyze motion in the region $H_0<0$, it is convenient to use the action–angle variables (J,Θ). As a result, we obtain

$$H_0(J)=-(1-J)^2/2, \quad \omega_0(J)=dH_0/dJ=1-J, \qquad (7)$$

where $\omega_0(J)$ is the frequency of the unperturbed oscillations.

The value of the action J varies between 0 for oscillations alongside the bottom of the well and 1 near the separatrix. The frequency ω_0 varies accordingly between 1 and 0. As the particle energy increases further (H_0), the motion changes to drift motion.

An external force destroys the integrability of the system and leads to an additional term in the Hamiltonian. Adopting the approach used in Ref. 9, we can derive the following relations:

$$fH_1(J,\Theta)=-f\sin(\Omega\tau)\mathrm{arcsinh}[a\sin(\Theta)/(1-a^2)^{1/2}]$$

$$=-f\sin(\Omega t)\sum_{k=0}^{\infty}H_{1,2k+1}(J)\sin[(2k+1)\Theta],$$
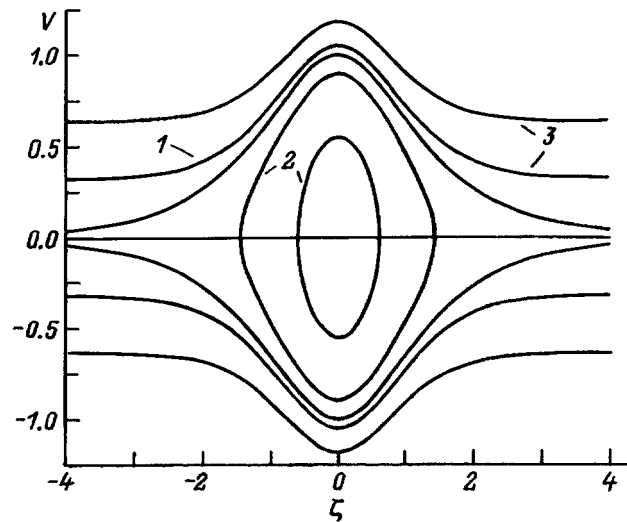
$$\qquad (8)$$

where

FIG. 1. Phase portrait of the unperturbed system (4). The separatrix (1) divides the region of periodic (2) and drift (3) trajectories.

$$H_{1,2k+1}(J)=\frac{2\Gamma(k+1/2)\cdot\Gamma(k+3/2)a^{2k+1}}{\pi(2k+1)\cdot\Gamma(2k+2)}$$
$$\times F(k+1/2,k+3/2,2k+2,a^2),$$

$$a^2=J(2-J),$$

where $\Gamma(x)$ is a gamma function and $F(a,b,c,d)$ is a hypergeometric function.

It can be seen from expression (8) that the perturbation only has odd harmonics in Θ and thus the trajectories resonant with the external perturbation will be those for which the following relation is satisfied

$$\omega_0(J)=\Omega/n, \quad n=2k+1. \qquad (9)$$

When an external influence is taken into account, the character of the motion can be assessed most easily by means of a Poincaré map. This map determines a sequence of points corresponding to the cross sections of some trajectory in the phase space of Eq. (4) at the time intervals $2\pi/\Omega$. For small f, according to the Kolmogorov–Arnol'd–Moser theorem, the nonresonant trajectories are only negligibly deformed (Fig. 2, region 1). Near resonant trajectories whose unperturbed frequency satisfies condition (9), a nonlinear resonance — splitting of a given trajectory — occurs. This leads to the appearance of pairs of singular points of the Poincaré map — elliptic and hyperbolic. Figure 2 (region 2) shows the separatrix of the nonlinear resonance for $n=3$. Using a standard nonlinear resonance technique,[10] we can estimate the width of the nth resonance

$$\Delta\omega=2[fH_{1,n}(J)|\omega_0'|]^{1/2}, \quad \omega_0'=d\omega_0(J)/dJ. \qquad (10)$$

The frequency spacing between the resonances is determined by

$$|\delta\omega|=|\omega_{2k+1}-\omega_{2k+3}|=2\Omega[(2k+1)(2k+3)]. \qquad (11)$$

If ε increases, the width will increase resonantly and may overlap neighboring resonances. In this case, the motion becomes chaotic, i.e., similar to the motion of a Brownian
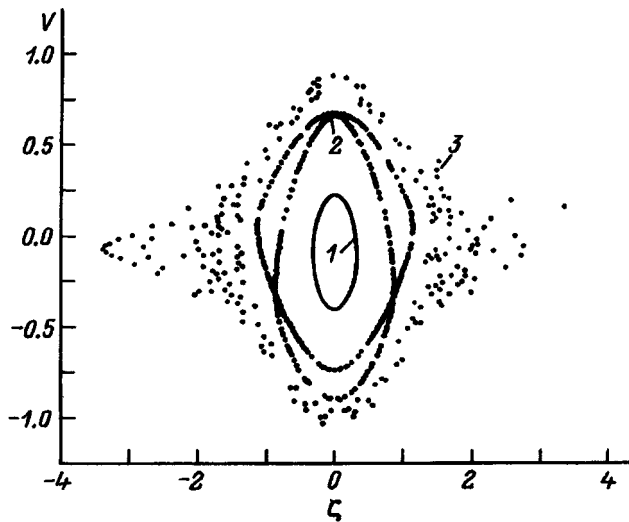
FIG. 2. Poincaré map for the case $\Gamma=0$, $\Omega=2$, $\varepsilon=0.14$: *1* — nonresonant trajectory, initial coordinates and velocity equal $(0, -0.4)$, *2* — motion near resonance with $n=3$ $(0, -0.898)$, and *3* — stochastic motion and fluxon depinning $(0, -0.957)$.

particle (Fig. 2, region *3*). According to the Chirikov criterion,[11,12] stochasticity occurs near the nth resonance when the condition $K=(\Delta\omega/\delta\omega)^2 \geqslant 1$ is satisfied or, if Eqs. (10) and (11) are used, when

$$f \geqslant f_{th} = \Omega^2/[H_{1,n}n^2(n+2)^2]. \quad (12)$$

Thus, the fluxon motion will be stochastic for almost all initial conditions and parameters of the problem satisfying condition (12). At this point, it should be noted that even in a region of advanced stochasticity, there are always regions of regular motion.

The analytic results were checked by a numerical simulation of Eq. (4) using the fourth-order Runge–Kutta–Merson method. The results of the numerical calculations for $\Omega=2$ are given in Table I.

The degree of chaos was estimated quantitatively by determining the largest Lyapunov exponent[13] characterizing the local instability of the trajectories in the phase space.[10] If we take initial points positioned at the distance $d(0)$, the spacing between the trajectories at time t may be written as $d(t) \cong d(0)\exp(\lambda t)$. If $\lambda < 0$, the motion is locally stable and regular (periodic or quasiperiodic), whereas if $\lambda > 0$, the motion is locally unstable (chaotic). Table I gives values of λ for various initial conditions and parameters. The results of the numerical simulation show that the resonance overlap criterion gives a satisfactory estimate of the onset of stochasticity. The difference between the analytic and numerical

values arises because, when deriving condition (12), we neglected the shift and width of the neighboring $(n+2)$th resonance.

Note that chaotic motion in this system is a transient process. A fluxon which has acquired a sufficient amount of energy during diffusion via resonances becomes separated from the inhomogeneity and then moves as a free particle. Thus, the stochasticity condition (12) may be considered as the condition for depinning of a fluxon from a local inhomogeneity.

INFLUENCE OF DISSIPATION ON FLUXON MOTION

In this section we consider the influence of dissipation ($\Gamma \neq 0$) on the characteristics of the establishment of chaos. When damping is ''switched on'' in system (4) (near the points $(0,0)$, $(\pm\infty,0)$), attractive points of the Poincaré map — attractors — appear, whose position and parameters depend on the ratio of the parameters Γ, f, and Ω. Most of the trajectories are attracted to the main attractor near $(0, 0)$. As the amplitude of the external force f increases, the neighborhood of this point may become unstable as a result of the formation of a homoclinic structure as follows. The equilibrium $(\pm\infty,0)$ and separatrix points (Fig. 1, curve *3*) of the unperturbed system (3) generate fixed mapping points and invariant stable and unstable manifolds related to the given points, in the perturbed system. For specific values of the parameters, these manifolds may intersect at a so-called homoclinic point. The existence of one homoclinic point gives rise to the existence of an infinite set of similar points (Refs. 10 and 13, p. 88). The presence of these points or the existence of a homoclinic trajectory implies local instability of the phase space, i.e.,sensitivity of the system to the initial conditions. In this case we talk of the establishment of homoclinic chaos.[10,13] From this point of view, Eq. (4) is interesting because the unperturbed system has no hyperbolic equilibrium positions, although a homoclinic structure does occur, as we shall see subsequently. The Mel'nikov method[14] is used to obtain the conditions for the formation of such a structure in the system, that is stochasticity. This method involves calculating the Mel'nikov function $M(t_0)$ characterizing the distance between perturbed invariant manifolds. If $M(t_0)$ has simple zeros, there exists a homoclinic point at which the manifolds intersect.[14] For Eq. (4) $M(t_0)$ is

$$M(t_0) = \int_{-\infty}^{\infty} [-\Gamma v_s(t) + f \sin(\Omega(t+t_0))]v_s(t)dt, \quad (13)$$

where v_s is the velocity at the unperturbed separatrix (6).

TABLE I.

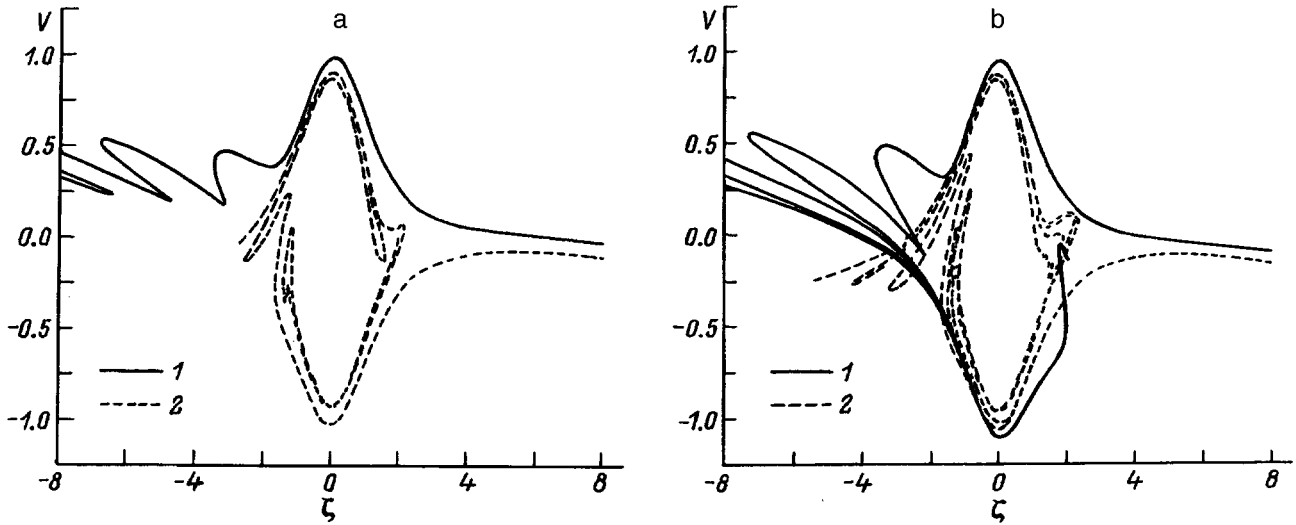| Resonance number n	Initial velocity $(x=0)$	Threshold amplitude from (12) f_{th}	Threshold amplitude from numerical analysis f_{th}	Largest Lyapunov exponent λ
9	0.9750	0.0262	0.011	3.52×10^{-2}
7	0.9584	0.0506	0.022	2.77×10^{-2}
5	0.9166	0.1226	0.058	5.80×10^{-2}
3	0.7454	0.5106	0.253	3.94×10^{-2}

FIG. 3. Invariant stable (*1*) and unstable (*2*) manifolds of Eq. (4) for the stationary point $x \to +\infty$ $\Omega = 1$, $\Gamma = 0.02$, threshold amplitude $f_c = 0.075$: a — nonintersecting manifolds for $f = 0.06$, b — homoclinic structure for $f = 0.1$.

From the condition $M(t_0) \leqslant 0$, we obtain the criterion of the onset of stochastic fluxon motion

$$f \geqslant f_c = \pi \Gamma / [2K_0(\Omega)], \tag{14}$$

where $K_0(\Omega)$ is the Macdonald (modified Bessel) function.

Using the asymptotic form $K_0(\Omega)$, we can obtain the following conditions:

$$f \geqslant \Gamma(\pi\Omega/2)^{1/2}\exp(\Omega) \qquad \text{for } \Omega \gg 1,$$
$$f \geqslant \Gamma(\pi\Gamma/2)[\ln(\Omega/2)]^{-1} \qquad \text{for } \Omega \ll 1. \tag{15}$$

Figure 3 shows invariant manifolds corresponding to the fixed point at $\zeta \to +\infty$. It can be seen that for below-critical values of f, no homoclinic points exist (Fig. 3a). For $f > f_c$ we have an intersection of invariant manifolds. Figure 3 was plotted using a Poincaré map of a fairly small segment near $\zeta \geqslant 10$, where an unstable manifold was calculated for $\tau > 0$ and a stable manifold was calculated for $\tau < 0$. Note that we are only aware of one study in which the existence of a homoclinic structure is demonstrated analytically and numerically for an infinite potential (Morse potential).[15]

The numerical simulation shows that condition (14) fairly accurately determines the intersection of the manifolds. This intersection initially occurs near the unperturbed separatrix and then, with increasing f, covers larger and larger regions of the phase space. As a result, the neighborhood of the main attractor also becomes unstable, which gives rise to randomness over almost the entire space. As in the case $\Gamma = 0$, the chaos is transient. In this case however, a particle having acquired energy does not go to infinity but "sticks" at large ζ ($d\zeta/d\tau \approx 0$) as a result of dissipation and stays there for a long time. Since these distances are much greater than the characteristic dimension of influence of the inhomogeneity, this drift may be interpreted as depinning of a fluxon from a microresistor.

CONCLUSIONS

It has been shown that the application of an external current $J(t)$ leads to nontrivial fluxon dynamics. In the absence of a current, the fluxon will be trapped by a microresistor and will undergo damped oscillations alongside it. The action of an alternating current induces periodic, quasiperiodic, and stochastic motion. Stochastic motion is observed when a fluxon, having undergone chaotic oscillations alongside an inhomogeneity and having acquired a sufficient amount of energy, becomes detached from this inhomogeneity. We estimate the external current density required for stochastic fluxon depinning. We take the following values of the typical parameters of a long Josephson junction:[1,2] $\omega_p \sim 10^8$ rad/s, $\Gamma \sim 0.01$ ($\alpha, \beta \sim 10^{-3} - 10^{-2}$), critical current density $J_c \sim 10^8$ A/m², and $\varepsilon \sim 0.1$. Then, from conditions (15) we obtain the amplitude of the external current density J:

1) $\omega_{\text{in}} \sim (10^{-3} - 10^{-2})\omega_p \ll \omega_p \nu_0$ — low frequencies of the external current:

$$J \geqslant 2\nu_0 \Gamma |\ln(\Omega/2)|^{-1} J_c \sim 10^5 - 10^6 \text{ A/m}^2,$$

2) $\omega_{\text{in}} \sim \omega_p \gg \omega_p \nu_0$ — high frequencies of the external current:

$$J \geqslant 4\nu_0^2 \Gamma(2\pi)^{1/2}\Omega^{1/2}\exp(\Omega)J_c \sim 10^7 \text{ A/m}^2,$$

$$\Omega = \omega_{\text{in}}/(\omega_p \nu_0).$$

This effect of stochastic fluxon depinning may well be useful for determining the quantitative characteristics of local inhomogeneities in long Josephson junctions. For this purpose, a slow Josephson fluxon is generated which becomes trapped by a microresistor and localized at it as a result of damping. The inhomogeneity parameter ε can then be estimated from the amplitude of the alternating current at which the kink becomes detached from the impurity.

[1] K. K. Likharev, *Introduction to the Dynamics of Josephson Junctions* [in Russian], Nauka, Moscow (1985).

[2] A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect* (Wiley, New York, 1982) [Russian translation, Mir, Moscow, 1984].

[3] D. W. McLaughlin and A. C. Scott, Phys. Rev. A **18**, 1652 (1978).

[4] B. A. Malomed, Phys. Lett. A **120**, 28 (1987); Physica D **27**, 113 (1987).

[5] T. Bountis and St. Pnevmatikos, Phys. Lett. A **143**, 221 (1990).

[6] F. Kh. Abdullaev, S. A. Darmanyan, and B. A. Umarov, Phys. Rev. A **41**, 4498 (1990).

[7] V. I. Karpman and E. M. Maslov, Zh. Éksp. Teor. Fiz. **73**, 537 (1977) [Sov. Phys. JETP **46**, 281 (1977)].

[8] F. Kh. Abdullaev and A. A. Abdumalikov, Phys. Status Solidi B **167**, 503 (1991).

[9] S. S. Abdullaev and G. M. Zaslavskiĭ, Zh. Éksp. Teor. Fiz. **80**, 524 (1981) [Sov. Phys., JETP **53**, 265 (1981)].

[10] A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion* (Springer-Verlag, New York, 1982) [Russian translation, Mir, Moscow, 1984].

[11] G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to Nonlinear Physics* [in Russian], Nauka, Moscow (1988).

[12] B. V. Chirikov, Phys. Rep. **52**, 265 (1979).

[13] *Stochastic Systems* (Special issue), Proc. IEEE **75**, No. 8 (1987).

[14] V. K. Mel'nikov, Tr. Mosk. Mat. Ob-va **12**, 3 (1963).

[15] G. P. Tsironis, *Dynamical Studies of the Discrete Nonlinear Schrödinger Equation*, Series of Scientific Monographs (Crete, 1994).

Translated by R. M. Durham

A new universal method of monitoring layer parameters and surface roughness in vacuum deposition and etching processes

A. M. Baranov, S. A. Tereshin, and I. F. Mikhaĭlov

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 103907 Moscow, Russia

Tunnel-thin films of various materials around 2–10 nm thick have recently become increasingly widely used in optoelectronics and microelectronics. These films have extremely extensive fields of application, ranging from nonvolatile memory elements to applications as functional layers in multilayer structures. Metals, semiconductors, and dielectrics have all been used. When such ultrathin films are used, their parameters (thickness, density, and surface roughness) must be strictly monitored directly during the technological process — so-called *in situ* monitoring.

At present, laser interferometry and ellipsometry are most commonly used for this purpose.[1,2] Disadvantages of laser interferometry are that it cannot be used to analyze coatings of submicron thickness ($d < 0.1 \mu$m) and materials whose refractive index varies as a function of the deposition conditions (such as diamond-like films[3,4]). Thus, ellipsometry is usually used to analyze ultrathin layers. However, ellipsometry is extremely sensitive to the state of the substrate surface. This makes it difficult to use when transition layers are present at the film–substrate interface, since the optical constants of the material on its surface vary during the initial growth of the film. In addition, ellipsometry cannot be used to study absorbing materials (such as metals).

Thus, a method of monitoring the reflection coefficient in the x-ray wavelength range between 0.05 and 0.3 nm was proposed to analyze the properties of ultrathin films.[5,6] In this range, all materials have an extremely low absorption coefficient so that the three most important parameters of the films can be determined: the thickness, surface roughness, and refractive decrement, which is directly related to the density of the material.

An important advantage of this method is its universal applicability, i.e., it is not sensitive to the parameters of the vacuum technological process (such as the pressure, working gas composition, and plasma parameters) or to the type of deposited material. In this case, the x-ray source and detector are located outside the working chamber.

Here we present results of using x-ray *in situ* monitoring to monitor the growth and etching processes of multilayer carbon structures, and also for monitoring the surface quality during treatment.

Figure 1 shows a schematic of the vacuum chamber with the in situ x-ray measuring system. The angle of incidence of the x-ray beam is $\Theta = 1$ deg. The x-ray radiation source was a tube with a copper anode ($\lambda = 0.154$ nm). The beam reflected specularly from the surface was recorded by the detector. The results were fed to a computer via a data processing system and interface. The substrates were polished

silicon wafers with an initial rms roughness $\sigma = 0.4$ nm. Carbon films were deposited by two methods: by magnetron sputtering of a graphite target in argon and by gas-phase deposition in a plasma excited by an rf discharge. The working gas was cyclohexane, C_6H_{12}, vapor.

Figure 2a gives the time dependence of the reflection coefficient $I_0(t)$ for the deposition of carbon layer I on silicon from cyclohexane (C_6H_{12}) vapor in an rf discharge plasma (Fig. 2a, curve *1*) and for layer II obtained by magnetron sputtering of a graphite target in Ar (Fig. 2a, curve *2*) during its deposition on the first layer (the dependence is reversed). The arrows in Fig. 2a indicate the beginning (b) and end (e) of the deposition process.

The initial intensity of the reflected beam was 1700 pulse/s. Growth of the carbon layer causes the intensity of the x-ray beam reflected by the film–substrate system to oscillate. The period of the oscillations L is determined by the Bragg condition $2L \sin \Theta = \lambda$. In accordance with the Bragg condition, both layers have the same thickness $d = 8.8$ nm. It can be seen from Fig. 2a that the growth rate V_0 of the first layer is slower than that of the following layer. The minimum observed on the curve $I_0(t)$ in the first interference order indicates that the density of the film material is lower than that of the substrate material.[5] The smaller amplitude of the peak on curve *2* (Fig. 2a) corresponds to the higher density of the second layer relative to the first. The densities and roughness of the growing layer may be determined by analyzing the behavior of $I = f(t)$. The film parameters averaged

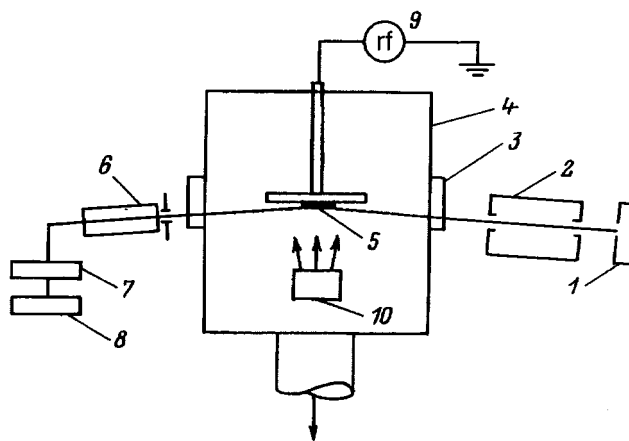FIG. 1. Schematic diagram of apparatus: *1* — radiation source, *2* — collimating system *3* — entry window, *4* — vacuum chamber, *5* — sample, *6* — detector, *7* — recording unit, *8* — computer, *9* — rf generator, and *10* — magnetron source.
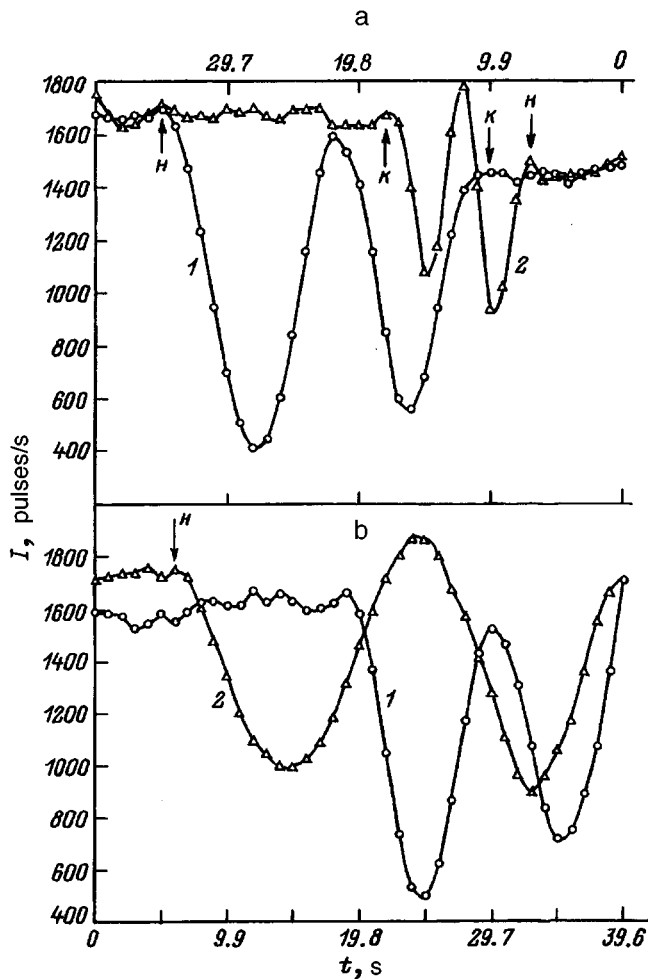
FIG. 2. Experimental time dependences of the reflection intensity for deposition of carbon layers on a silicon substrate (a) and etching (b). *1* — layer obtained in a plasma excited by an rf bias, *2* — layer obtained by magnetron sputtering.
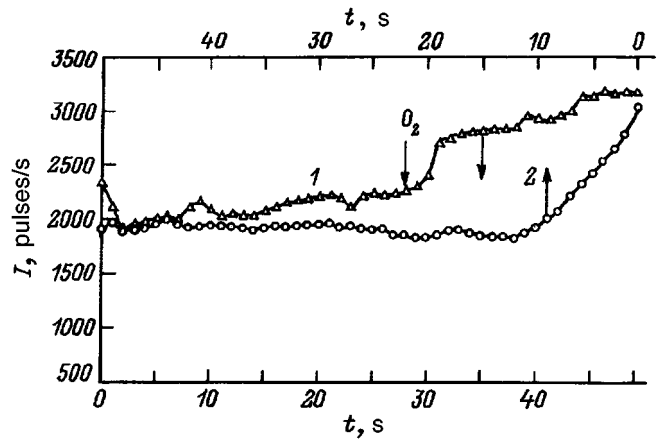

FIG. 3. Experimental curves of the reflection intensity versus treatment time for a silicon wafer. *1* — in an Ar + O_2 mixture and *2* — in Ar.

in the range $L/2$ are presented in Table I. The average densities and rms roughness of the films are calculated in the thickness range $d = L/2$ because then, by converting from $I = f(t)$ to $I = f(d)$, we can obtain analytic expressions to calculate ρ and σ (Refs. 5 and 6).

It can be seen from Table I that the density and roughness of the first layer increase during its growth. The density of layer II is higher than that of layer I and is close to the

TABLE I. Parameters of layers I and II obtained during deposition and etching.

Layer No.	Thickness d	Deposition			Etching		
		ρ, g/cm^3	σ, Å	V_0, Å/s	ρ, g/cm^3	σ, Å	V_{et}, Å/s
I	$0-L/2$	1.78	4.00	3.7	1.84	4.01	4.4
	$L/2-L$	1.81	4.32	3.7	1.88	4.58	4.4
	$L-3/2L$	1.90	4.31	3.7	2.01	4.54	4.4
	$3/2L-2L$	1.96	4.96	3.7	1.96	4.81	4.4
II	$0-L/2$	2.11	4.90	11	2.05	3.82	2.8
	$L/2-L$	2.04	3.44	11	2.00	2.90	2.8
	$L-3/2L$	2.05	3.10	11	2.02	2.90	2.2
	$3/2L-2L$	2.11	3.70	11	2.09	3.68	2.5

density of graphite (2.26 g/cm^3). In addition, the surface of layer II has a lower value of σ. The difference in the densities of the carbon layers is a consequence of the different contents of sp^2 and sp^3 states and hydrogen.[3,4]

This two-layer coating was etched by a standard technological process in an oxygen plasma excited by supplying rf power to the cathode on which the silicon substrate was mounted. The curve $I_{et}(t)$ obtained for etching is plotted in Fig. 2b. It can be seen (in Fig. 2b, curves *1* and *2* correspond to Fig. 2a) that, in terms of absolute values of the recorded signal, the curve $I_{et}(t)$ accurately reproduces $I_0(t)$ but in reverse order. As a result, the dependences $\rho(d)$ and $\sigma(d)$ calculated from $I_{et}(t)$ show a good correlation with the values obtained from $I_0(t)$. However, the rate of etching (V_{et}) of the layer obtained by magnetron sputtering in argon is slower (Fig. 2b, curve *2*) than that for the layer obtained by deposition in an rf discharge (Fig. 2b, curve *1*). This is clearly the result of a difference in the densities of the first and second layers.

The substrate surface is usually subjected to ion-plasma cleaning before the thin-film coating is deposited. Physical sputtering, ion-chemical etching, and plasma-chemical etching are used to clean the surface.[7] In many cases, it is important that the substrate cleaning process should not increase the surface roughness or cause oxidation or adsorption of carbon.

Figure 3 shows the behavior of $I(t)$ when an x-ray beam is reflected by a silicon wafer undergoing treatment in an argon plasma with added oxygen. It can be seen that at the beginning of the treatment process in Ar, the intensity of the reflected signal from the surface of the Si wafer falls sharply (Fig. 3, curve *1*). The reflected signal then remains constant. However, when oxygen is added to the Ar, the reflection coefficient increases abruptly, reaching a constant value. When the silicon wafer then undergoes further treatment in pure argon, the intensity falls again and reaches a level corresponding to the refection before the oxygen was added (Fig. 3, curve *2*). This behavior of the reflection coefficient may be attributed to changes in the surface roughness of the silicon.

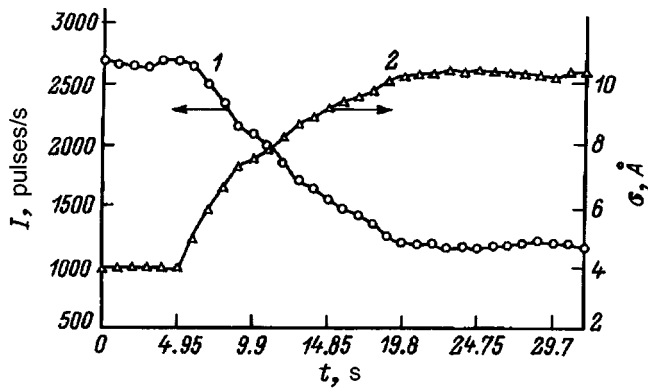Figure 4 shows the behavior of $I(t)$ during the treatment

FIG. 4. Experimental curve $I(t)$ for the treatment of a silicon wafer in CF_4 (1) and $\sigma(t)$ (2).

of a silicon wafer in a CF_4 plasma. It can be seen that the reflected signal is more than halved. In accordance with the Debye–Waller condition $I = I_s \exp[-(4\pi\sigma \sin \Theta/\lambda)^2]$ (where I_s is the intensity before the treatment process),[8] this corresponds to an increase in the rms surface roughness of the wafers from 0.4 to 1.3 nm. The observed increase in the silicon roughness has no single mechanism. It is the combined result of various processes taking place on the silicon surface in a CF_4 plasma (for example, etching of an SiO_x surface layer, etching of the silicon surface with fluorine radicals at sites of emergence of dislocations, deposition of carbon atoms on Si, and so on).

To sum up, it has been shown that in situ x-ray monitoring at 0.154 nm can be used to monitor the surface quality and to study the growth kinetics of single-layer and multilayer structures to within 0.1 nm. This method has the advantages that it is easy to implement, it can be universally applied to study any substrates and films, and can also be used to determine the most important parameters of thin-film coatings (thickness, density, and roughness) at the initial stages of film growth.

[1] M. S. Sterheim, W. Gelder, and A. W. Hartman, J. Electrochem. Soc. **130**, 655 (1983).
[2] B. Hong, W. Wakagi, R. W. Collins *et al.*, Diamond Related Mater. **3**, 431 (1994).
[3] H. Tsai and B. Body, J. Vac. Sci. Technol. A **5**, 3287 (1987).
[4] V. V. Sleptsov, A. A. Kuzin, G. F. Ivanovsky *et al.*, J. Non-Cryst. Solids **136**, 53 (1991).
[5] I. F. Mikhailov, V. I. Pinegin, V. V. Sleptsov, and A. M. Baranov, Cryst. Res. Technol. **30**, 643 (1995).
[6] A. M. Baranov, S. A. Tereshin, I. F. Mikhailov *et al.*, Proc. SPIE **2519**, 108 (1995).
[7] G. F. Ivanovskiĭ and V. I. Petrov, *Ion-Plasma Treatment of Materials* [in Russian], Radio i Svyaz', Moscow (1986).
[8] A. V. Vinogradov and I. V. Kozhevnikov, Tr. Fiz. Inst. Akad. Nauk SSSR **196**, 63 (1989).

Translated by R. M. Durham

Sources of 1/F noise in gallium arsenide IMPATT diodes

S. A. Kornilov, K. D. Ovchinnikov, and É. B. Kislitsyn

M. A. Bonch-Bruevich State University of Telecommunications, 191186 St. Petersburg, Russia

The quasistatic approximation is used to analyze 1/F noise in IMPATT diodes in the static and dynamic (self-oscillating) modes. Sources of 1/F noise are defined in accordance with the fluctuator model: allowance is made for fluctuations of the charge of traps and fluctuations of the electron drift velocity caused by their scattering by traps and metastable neutral centers. It is shown that the fluctuations of the voltage across the diode and the fluctuations of the oscillation frequency are mainly determined by the fluctuations of the trap charge, while the fluctuations of the oscillation amplitude are determined by scattering by neutral centers. A method is developed to determine the intensity of noise sources using the results of measurements of the fluctuations in the static and dynamic modes of IMPATT diodes and a method of checking the model as a whole is checked. Experimental results are presented and these show satisfactory agreement with the calculations. © *1997 American Institute of Physics.*
[S1063-7842(97)01008-8]

INTRODUCTION

Information on sources of 1/F noise in semiconductors with hot carriers is scant and frequently contradictory. However, this problem is interesting from the physical point of view and has important applications — many semiconductor microwave devices utilize hot electrons. IMPATT diodes are an attractive research prospect because they operate in extremely high fields (hundreds of kV/cm) and can expand the experimental base by including dynamic measurements. The aim of the present paper is to select an adequate model of 1/F noise sources in IMPATT diodes, to check this model experimentally, and to develop a method of determining the intensity of the sources using the results of the measurements.

MODEL OF 1/F NOISE SOURCES IN IMPATT PHOTODIODES AND AN APPROACH TO CHECK THIS MODEL

We shall briefly consider models of 1/F noise sources which have been used to describe this noise in high fields. In Refs. 1–6 it is assumed that bulk 1/F noise is caused by fluctuations in the carrier mobility (drift velocity). The origin of these fluctuations is either not specified or is attributed to scattering by acoustic phonons[2–4] in the spirit of the Hooge–Vandamme concept.[7] This last concept presupposes that the phonon concentration exhibits 1/F noise, which is difficult to reconcile because the phonon frequency has a lower limit in samples of finite dimensions.

The authors of Refs. 8 and 9 proposed an alternative model of 1/F noise, where this noise is attributed to fluctuations of the energy level population in the density-of-states ''tail'' extending into the semiconductor band gap. In other words, the 1/F noise is attributed to the behavior of the traps, which leads to fluctuations in their total charge and the free carrier concentration. This model, which is similar to the well-known McWhorter model,[10] was used in Ref. 8 to calculate the voltage fluctuations at IMPATT diodes. The trap model can explain the 1/F spectrum and, which is particu-

larly important, has direct experimental confirmation of the involvement of traps in the formation of 1/F noise (sample illumination experiments[9]). However, this model does not solve all the problems. The authors of Ref. 9 note that sample illumination experiments indicate that in addition to the ''trap'' component, the 1/F noise has another component whose nature has not yet been identified.

The concept of the origin of 1/F noise known as the fluctuator model has recently attracted serious attention.[11] It was demonstrated in Ref. 12 that this model may be applied to disordered regions of high-temperature semiconductors. This model assumes that the 1/F noise is caused by interaction between charge carriers and metastable centers possessing a relaxation time distribution (fluctuators). In semiconductors, these may be neutral formations such as two-level systems and generation-recombination centers. In the first case, fluctuations of the carrier mobility (drift velocity) occur as a result of scattering and in the second case, fluctuations of the trap charge and free carrier concentration are also added. The fluctuator model is supported by strong experimental evidence, including the observation of the behavior of single fluctuators and the transition from a Lorentzian to a 1/F spectrum with increasing number of fluctuators (see the review presented in Ref. 13), the influence of light on the 1/F noise already mentioned,[9] and the demonstration that the low-frequency noise spectrum in ultrapure semiconductors is a superposition of Lorentzians.[14]

An important factor is that the fluctuator model in a certain sense combines the Lorentzian model and the mobility fluctuation model, but attributes the latter to scattering by metastable centers rather than by phonons. This model was used in Ref. 15 to analyze the influence of electron heating on 1/F noise in a homogeneous sample, but the influence of traps was neglected.

Here we analyze 1/F noise in IMPATT diodes using the fluctuator model with allowance for interaction of charge carriers with charged and neutral fluctuators.[16] Thus, the following sources of 1/F noise are introduced in the theory:

fluctuations in the concentration of charged traps δN_t, drift velocity fluctuations δv_t correlated with δN_t (the result of scattering by traps), and drift velocity fluctuations δv_n uncorrelated with δN_t (the result of scattering by neutral fluctuators).

A widely used method of checking models of $1/F$ noise sources involves comparing calculations of the current noise with an experiment conducted in the static operating mode of the samples or devices (dc measurements). The method proposed here combines dc measurements with measurement of the oscillation fluctuations in the generation regime of an IMPATT diode. This obviously makes the calculations and experiment more complicated, but enlarges the scope of the model and allows more comprehensive information to be obtained on the noise sources.

ANALYSIS OF 1/F NOISE IN IMPATT DIODES

The theory is developed for single-drift, unannealed, GaAs IMPATT diodes with a uniformly donor-doped base, which were used in the experiment. The slowness of the $1/F$ noise compared with the transient processes in the diode and the oscillator allowed the fluctuation problem to be solved in the quasi-steady-state approximation using equations describing steady-state modes. The processes in the diode were analyzed using the equivalent multiplication layer model.[17]

dc regime

In the absence of any oscillations, the processes in IMPATT diodes are described by a system of equations consisting of the solution of the Poisson equation for the depletion layer,

$$u = \left(1 + \frac{l_d}{l_a}\right) u_a - \frac{q(N_D + N_t)l_d^2}{2\varepsilon}\left(1 + \frac{l_a}{l_d}\right) + \frac{il_d^2}{2\varepsilon s v_s}\left(1 + \frac{l_a}{3l_d}\right),$$

(1)

the avalanche equation

$$i = i_s\left[1 - l_a\bar{\alpha}\left(\frac{u_a}{l_d}\right)\right]^{-1},$$

(2)

the supply circuit equation

$$E_b = u + iR_b,$$

(3)

and an equation derived from the condition of zero electric field at the end of the depletion layer, obtained, like Eq. (1), from the Poisson equation

$$\frac{u_a}{l_a} - \frac{q(N_D + N_t)l_d}{\varepsilon}\left(1 + \frac{l_a}{2l_d}\right) + \frac{il_d}{\varepsilon s v_s}\left(1 + \frac{l_a}{6l_d}\right) = 0, \quad (4)$$

where u is the voltage across the depletion layer, u_a is the voltage across the multiplication layer, i is the avalanche current, which is equal to the diode supply current, E_b and R_b are the electromotive force and the internal resistance of the supply source, l_a and l_d are the lengths of the multiplication and drift layers, s is the cross-sectional area of the structure, i_s is the saturation current, $\bar{\alpha}(u_a/l_d)$ is the impact ionization coefficient averaged over the length of the multiplication layer, q is the electron charge, ε is the permittivity,

N_D and N_t are the donor and charged trap concentrations, and v_s is the saturated electron drift velocity.

In accordance with the adopted model of $1/F$ noise sources, the perturbations are introduced into Eqs. (1)–(4) by the spontaneous fluctuations δN_t and $\delta v_s = \delta v_t + \delta v_n$ which lead to fluctuations of the variables u, u_a, i, and l_d. Expressing the time-dependent quantities by sums of the averages and the fluctuations ($N_t = \bar{N}_t + \delta N_t$, $u = \bar{u} + \delta u$, and so on) and using the smallness of these fluctuations, we can linearize the system (1)–(4) with respect to the fluctuations and find the relation between δu, δl_d, and the perturbations

$$\frac{\delta u}{ir_d} = -\left(h_t\frac{N_D}{n_0} - \eta h_v\right)\frac{\delta N_t}{N_D} - h_v\frac{\delta v_n}{v_s},$$

(5)

$$\frac{\delta l_d}{l_d} = -\left(\frac{H_t N_D}{N_D - n_0} - \eta\frac{H_v n_0}{N_D - n_0}\right)$$

$$\times \frac{\delta N_t}{N_D} - \frac{H_v n_0}{N_D - n_0}\frac{\delta v_n}{v_s}.$$

(6)

Here $h_t = 1 + l_a/l_d$, $h_v = 1 + l_a/3l_d$, $H_t = 1 + l_a/2l_d$, $H_v = 1 + l_a/6l_d$, n_0 is the electron concentration in the depletion layer, $\eta = (\tau_p/2\tau_{pt})(N_D/N_t)$,

$$r_d = l_d^2/(2\varepsilon s v_s)$$

(7)

is the differential resistance of the junction.

In Eqs. (5) and (6), the component of the velocity fluctuations caused by scattering by traps is expressed in terms of δN_t: $\delta v_t/v_s = -(\tau_p/2\tau_{pt})(\delta N_t/N_t)$, where τ_p and τ_{pt} are the resultant momentum relaxation time and that associated with scattering by traps. To simplify the notation in Eqs. (5)–(7) and subsequently, the average notation is omitted, i.e., it is assumed that $\bar{N}_D$ is written as N_D, $\bar{l}_d$ as l_d, and so on.

The following approximations were used to derive Eqs. (5) and (6): a) $R_b \gg r_d$ (the diode power supply has a high internal resistance); b) $N_D - n_0 \gg N_t$; c) l_a = const (it can be shown that $\delta l_a \ll \delta l_d$, at least for $R_b \gg r_d$).

Oscillation generation mode

Representing an IMPATT-diode self-excited oscillator near the oscillation frequency as a singly resonant oscillatory circuit (Fig. 1), we can write the equations for the steady-state oscillations as follows:

$$R + R_e(B, \omega_0) = 0,$$

(8)

$$\omega = \omega_0 + \frac{1}{2L}x_e(B, \omega_0).$$

(9)

Here $\omega_0 = (LC)^{-1/2}$, where C is the diode capacitance, L is the equivalent inductance of the cavity, R is the loss resistance (including the losses in the load), R_e is the diode resistance for the first harmonic and x_e is the "electron" component of the reactance for the first harmonic (the total reactance is $1/\omega C - x_e$). According to Refs. 17 and 18, R_e and x_e are given by

$$R_e = -\frac{(\Omega_a/\omega_0)^2\Phi(B)(1 - \cos\Theta_d)/\Theta_d}{\omega_0 C_d[1 - (\Omega_a/\omega_0)^2\Phi(B)]},$$
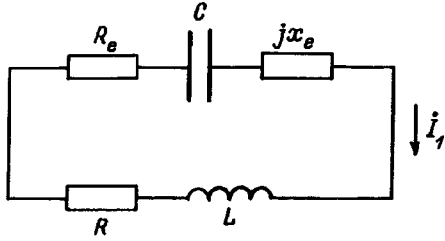
(10)

FIG. 1. Circuit diagram of IMPATT-diode self-excited oscillator (for the first harmonic).

$$x_e = -\frac{(\Omega_a/\omega_0)^2\Phi(B)}{1-(\Omega_a/\omega_0)^2\Phi(B)}\left[\frac{1}{\omega C_a}+\frac{\sin\Theta_d}{\omega C_d\Theta_d}\right]. \qquad (11)$$

Here C_a and C_d are the capacitances of the multiplication and drift layers, $\Theta_d = \omega_0 l_d/v_s$ is the drift angle in the drift layer, $\Omega_a = [2\bar{\alpha}' i/C_a\tau_a]^{1/2}$ is the avalanche frequency (τ_a is the transit time in the multiplication layer, $\bar{\alpha}'$ is the derivative of $\bar{\alpha}$ with respect to the electric field), $\Phi(B) = 2I_1(B)/[BI_0(B)]$, where $I_0(B)$ and $I_1(B)$ are modified Bessel functions, $B = 2\bar{\alpha}'U_a/\Theta_a$ ($\Theta_a = \omega_0\tau_a$), where U_a is the amplitude of the voltage across the multiplication layer. The quantity B and the amplitude A of the first harmonic of the current $\dot{I}_1$ (Fig. 1) are related by[17]

$$A = \left[\frac{\omega_0 C_a\Theta_a}{2\bar{\alpha}'}-i\Phi(B)\right]B. \qquad (12)$$

Equations (8), (9), and (12) with allowance for Eqs. (10) and (11) form a system determining the steady-state quantities B, A, and ω. After linearizing with respect to all the stochastic variables, we can find a relation in the quasi-steady-state approximation between the fluctuations $\delta\omega$, δA, and the primary perturbations δN_t, δv_t, and δv_n, which modulate the parameters dependent on them. The fluctuations of the length of the drift region must also be taken into account using Eq. (6). Calculations carried out using the same assumptions as those used to derived Eqs. (5) and (6) yield the results

$$-\frac{2\delta\omega}{\omega_0} = \left(b_1-\frac{b_2}{Q_L}\right)\frac{\delta N_t}{N_D}-\left(b_3-\frac{b_4}{Q_L}\right)\frac{\delta v_n}{v_s}, \qquad (13)$$

where

$$b_1 = \frac{C}{C_d}\left(H_t-\eta H_v\frac{n_0}{N_D}\right)\frac{N_D}{N_D-n_0},$$

$$b_2 = \frac{\Theta_d}{1-\cos\Theta_d}\left\{\left(1+\frac{C_d}{C_a}\frac{\Theta_d\sin\Theta_d}{1-\cos_d}\right)\right.$$
$$\left.\times\left[\frac{H_tN_D}{N_D-n_0}-\eta\left(1+\frac{H_vn_0}{N_D-n_0}\right)\right]+\eta\frac{C_d}{C_a}\right\},$$

$$b_3 = \frac{C}{C_d}\frac{H_vn_0}{N_D-n_0},$$

$$b_4 = \frac{\Theta_d}{1-\cos\Theta_d}$$

$$\times\left\{\left(1+\frac{C_d}{C_a}\frac{\Theta_d\sin\Theta_d}{1-\cos\Theta_d}\right)\left(1+\frac{H_vn_0}{N_D-n_0}\right)-\frac{C_d}{C_a}\right\},$$

$$\frac{1}{P(B)}\frac{\delta A}{A} = a_1\frac{\delta N_t}{N_D}-a_2\frac{\delta v_n}{v_s}, \qquad (14)$$

where

$$a_1 = \frac{\Theta_d\sin\Theta_d}{1-\cos\Theta_d}\frac{H_tN_D}{N_D-n_0}$$
$$+\eta\left[1+\frac{1+P(B)}{P(B)[1-(\Omega_a/\omega_0)^2\Phi(B)]}\right.$$
$$\left.-\frac{\Theta_d\sin\Theta_d}{1-\cos\Theta_d}\left(1+\frac{H_vn_0}{N_D-n_0}\right)\right],$$

$$a_2 = 1+\frac{1+P(B)}{P(B)[1-(\Omega_a/\omega_0)^2\Phi(B)]}$$
$$-\frac{\Theta_d\sin\Theta_d}{1-\cos\Theta_d}\left(1+\frac{H_vn_0}{N_D-n_0}\right).$$

In Eq. (13), $Q_L = \omega_0 L/R$ is the Q factor of the loaded cavity and in Eq. (14), $P(B) = -[1-2(\Omega_a/\omega_0)^2\varphi']\times(1-B\varphi'/\varphi)^{-1}$, where $\varphi = I_1(B)/I_0(B)$ and $\varphi' = d\varphi/dB$. An analysis of expressions (5), (13) and (14) shows that the influence of the trap charge fluctuations δN_t on δu, $\delta\omega$, and δA is considerably stronger than the influence of the drift velocity fluctuations δv_t correlated with δN_t. In all these expressions the action of δv_t is mapped by terms proportional to $\eta = 0.5(\tau_p/\tau_{pt})(N_D/N_t) = 0.5(\tau_p/\tau_{pD})(\tau_{pD}/\tau_{pt})\times(N_D/N_t)$, where τ_{pD} is the momentum relaxation time, associated with scattering by donors since $\tau_{pt}/\tau_{pD} = N_D/N_t$ and $\eta = 0.5\tau_p/\tau_{pD}$. From this it follows that $\eta \ll 1$ since at high temperature (around 450 K for IMPATT diodes) scattering by phonons predominates ($\tau_p \ll \tau_{pD}$). Assuming that the coefficients h_t, h_v, H_t, and H_v are of the order of unity, and n_0/N_D is of the order 10^{-1}, terms proportional to η can be neglected in Eqs. (5), (13), and (14), and this will be done subsequently. We merely note that, as applied to Eq. (14), this approximation is valid provided that $\Theta_d \leq 0.9\pi$, which is satisfied in most cases.

Figure 2 illustrates the channels of action of the noise sources on δu, $\delta\omega$, and δA. The dashed lines indicate channels whose influence is known to be weak and is neglected in the calculations. These include modulation of the parameters by the source δv_t and modulation of the frequency and amplitude of the oscillations by the current, whose fluctuations according to Eq. (3) are $\delta i = -\delta u/R_b$ and are small because R_b is assumed to be large.

We now consider the energy spectra of the fluctuations, writing them in normalized form. Assuming that δN_t and δv_t are uncorrelated, it follows from Eqs. (5), (13) and (14) that

$$S'_u = (h_tN_D/n_0)^2S'_{Nt}+h_t^2S'_{vn}, \qquad (15)$$
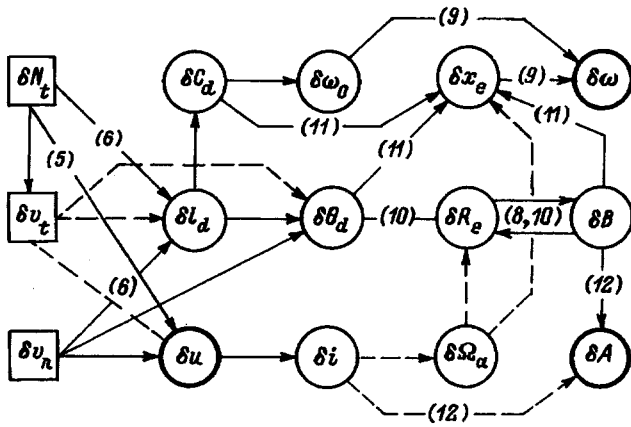
$$S'_\alpha = a_1^2S'_{Nt}+a_2^2S'_{vn}, \qquad (16)$$

FIG. 2. Channels of action of $1/F$ noise sources δN_t, δv_t, and δv_n on fluctuations of the amplitude δA, frequency $\delta \omega$ of the oscillations, and the voltage across the diode δu. The numbers of the equations controlling the coupling are given in parentheses.

$$S_f' = \left(b_1 - \frac{b_2}{Q_L} \right)^2 S_{Nt}' + \left(b_3 - \frac{b_4}{Q_L} \right)^2 S_{vn}'. \tag{17}$$

Here $S_u' = S_u/(ir_d)^2$, $S_\alpha' = S_\alpha/P^2(B)$, and $S_f' = 4S_f/f_0^2$, where S_u is the spectrum of the voltage fluctuations across the diode δu, S_f is the spectrum of the fluctuations of the cyclic frequency $\delta f = \delta \omega / 2\pi$, and S_α is the spectrum of the relative amplitude fluctuations $\delta A/A$. These spectra are directly measured experimentally. The normalized fluctuation spectra δN_t and δv_n were determined as follows: $S_{Nt}' = S_{Nt}/N_D^2$ and $S_{vn}' = S_{vn}/v_s^2$.

For a given supply current to the diode the oscillation mode of the oscillator depends on the loss resistance R (Fig. 1), which may be varied experimentally by changing the relation between the cavity and the load. This will be accompanied by variations in the cavity Q factor Q_L, the amplitude of the current oscillations A, and the amplitude of the voltage across the multiplication layer, which is characterized by the parameter B. The theory allows Q_L and B to be related to the coupling parameter $\beta = i/i_{th}$, where i is working supply current and i_{th} is the threshold supply current (corresponding to the self-excitation threshold of the oscillator)[17,6]

$$\beta = BI_0(B)/2I_1(B), \tag{18}$$

$$Q_L = \frac{C_d}{C} \frac{\Theta_d}{1 - \cos \Theta_d} \left(\frac{\omega_0^2}{\Omega_a^2} \beta - 1 \right). \tag{19}$$

The threshold current is easily measured. Thus, Eqs. (18) and (19) link the calculations made using Eqs. (16) and (17) to the experiment.

PROCEDURE FOR DETERMINING THE INTENSITY OF 1/F NOISE SOURCES FROM EXPERIMENTAL DATA

We use the system (15)–(17) to determine the spectral densities (intensities) of the $1/F$ noise sources S_{Nt}' and S_{vn}', using the results of measurements of the spectral densities of the fluctuations S_u, S_α, and S_f. Two equations, (15) and (16), are adequate for this purpose. The remaining equation (17) can be used to check the overall fit of the model.

Assuming that the left-hand sides of Eqs. (15) and (16) are known ($S_u' = S_u/(ir_d)^2$, $S_\alpha' = S_\alpha/P^2(B)$, where S_u and S_α are measured quantities), we find

$$S_{Nt}' = \frac{a_2^2 S_u' - h_v^2 S_\alpha'}{(a_2 h_t N_D/n_0)^2 - a_1^2 h_v^2}, \tag{20}$$

$$S_{vn}' = \frac{(h_t^2 N_D/n_0)^2 S_\alpha' - a_1^2 S_u'}{(a_2 h_t N_D/n_0)^2 - a_1^2 h_v^2}. \tag{21}$$

Obviously these relations are only effective when the minuends in the numerators of Eqs. (20) and (21) and in their common denominator are greater than the subtrahends. This is achieved in particular when $N_D/n_0 \gg 1$, although the possibility of a combination of parameters and mode for which this condition is not satisfied cannot be excluded. At any rate, the relation between the minuends and subtrahends should be monitored when making calculations using Eqs. (21) and (22).

Equation (17) was chosen from the three possible equations to check the model because it contains the cavity Q factor Q_L, which varies substantially with varying β, i.e., the coupling between the resonator and the load. By comparing the dependence calculated using the values of S_{Nt} and S_{vn} determined from Eqs. (20) and (21) with the experimental curve, we can assess the accuracy of the adopted noise source model and the procedure for determining their intensity.

EXPERIMENTAL RESULTS AND DISCUSSION

The experiment was carried out using unannealed GaAs IMPATT diodes in the 10 GHz range with a Schottky barrier and a uniformly doped base. The diode parameters were: $N_D = 10^{16}$ cm^{-3}, $n_0 = 7 \times 10^{14}$ cm^{-3}, $l_a = 0.9$ μm, $l_d = 2.1$ μm, $\Theta_d = 0.9\pi$, and $r_d = 25$ Ω. All the measurements were made at the analysis frequency of 30 Hz, at which the $1/F$ noise predominated over the avalanche noise over a wide range of β for $R_b = 1.3$ kΩ and a supply current $i = 80$ mA. The value of β was regulated by changing the relation between the resonator and the load.

TABLE I. Average values of S_u, S_α, S_u', S_α' and S_{Nt}', S_{vn}' for two batches of like IMPATT diodes.

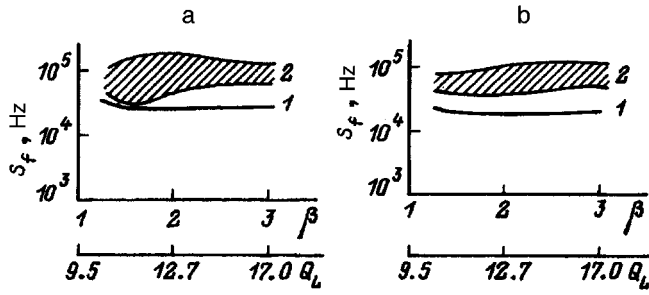| Batch No.	S_u, V$^2 \cdot$ Hz^{-1}	S_α, Hz^{-1}	S_u', Hz^{-1}	S_α', Hz^{-1}	S_{Nt}', Hz^{-1}	S_{vn}', Hz^{-1}
1 (9 diodes)	4.3×10^{-12}	1.5×10^{-13}	1.1×10^{-12}	8.4×10^{-14}	1.6×10^{-15}	1.3×10^{-13}
2 (5 diodes)	2.9×10^{-12}	10^{-13}	7.1×10^{-13}	5.6×10^{-14}	1.1×10^{-15}	8.6×10^{-14}

FIG. 3. Curves $S_f(\beta)$ for diodes in batch 1 (a) and batch 2 (b): $i = 80$ mA, $F = 30$ Hz; 1 — calculated results, 2 — scatter zone of experimental data.

FIG. 4. Curves of $S_\alpha(\beta)$ (a) and $S_f(\beta)$ (b). 1 — calculations for model with $1/F$ noise source δN_t: 2 —— the same with the source δv_n, 3 — scatter zone of experimental data, $i = 80$ mA, $F = 30$ Hz.

Determination of the intensity of 1/F noise sources.

Table I gives the measured values of S_u and S_α, their normalized values S'_u and S'_α, and also the noise source intensities S_{Nt} and S_{vn} calculated using Eqs. (20) and (21). The measurements and the calculations were made for $\beta = 1.85$.

Check of the model

Figure 3 shows the curves $S_f(\beta)$ calculated using Eq. (17) for both batches of diodes using the previously determined values of S'_{Nt} and S'_{vn} (Table I). Also shown are the scatter zones of the experimental values of $S_f(\beta)$. It can be seen that the theory agrees with the experiment, showing a weak dependence of S_f on β. This result is by no means trivial, since Eq. (17) contains the Q factor Q_L which varies substantially with β; agreement can only occur if S'_{Nt} and S'_{vn} are correctly determined. The calculated and measured levels of S_f are also fairly close.

We compare these results with those obtained by simpler models which only assume one source of $1/F$ noise — velocity fluctuations or trap charge fluctuations. These variants are described by Eqs. (15)–(17) in which $S'_{Nt} = 0$ should be assumed for the velocity fluctuation model or $S'_{vn} = 0$ for the charge fluctuation model. In both cases, the noise source intensity can be determined from Eq. (15): $S'_{vn} = S_u h_v^{-2}$ or $S'_{Nt} = S_u (h_t N_D / n_0)^{-2}$. For diodes in batch 1 we have $S'_{vn} = 8 \times 10^{-13}$ Hz^{-1} and $S'_{Nt} = 1.9 \times 10^{-15}$ Hz^{-1}. In this case, Eqs. (16) and (17) remain independent and may be used to check the models. The results are plotted in Fig. 4. It can be seen that the velocity fluctuation model leads to a serious qualitative discrepancy between theory and experiment in terms of the dependence $S_f(\beta)$ while the charge fluctuation model leads to a less serious quantitative discrepancy in the S_α level.

Thus, the model assuming the simultaneous action of two uncorrelated sources of $1/F$ noise shows considerably better agreement with the 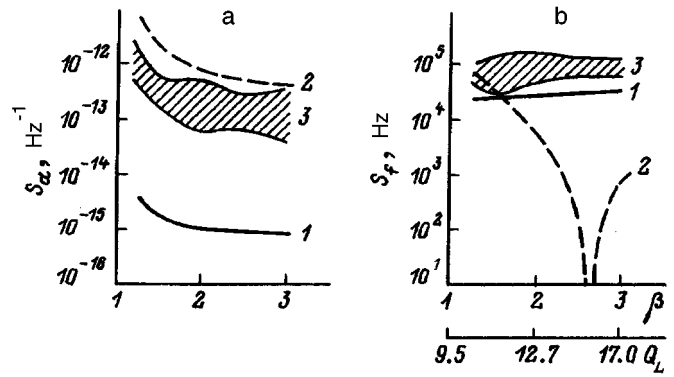experiment. In this model, the volt-age fluctuations across the diode and the fluctuations of the oscillation frequency are mainly determined by trap charge fluctuations, while the fluctuations of the oscillation amplitude are determined by fluctuations of the electron drift velocity associated with their scattering by neutral metastable centers. This is demonstrated by the results of the calculations made using Eqs. (15)–(17), which are given in Table II, and by the data given in Table I.

Note that the conclusion that fluctuations of the trap charge play a dominant role in the formation of $1/F$ noise in the diode voltage essentially agree with the viewpoint put forward by the authors of Ref. 8.

CONCLUSIONS

It has been shown that a model assuming two uncorrelated sources of $1/F$ noise in an IMPATT photodiode (fluctuations of the trap charge and fluctuations of the scattering by neutral metastable centers) satisfactorily agrees with the experiment. It has been established that the level of low-frequency fluctuations of the voltage across the diode and the frequency fluctuations in the oscillation mode are mainly determined by fluctuations of the trap charge, whereas the amplitude fluctuations are determined by fluctuations of the scattering by neutral centers. In all cases, fluctuations of the scattering by traps play a minor role. A method has been developed to determine the intensity of the $1/F$ noise sources from the results of measurements of the fluctuations in static and oscillatory modes.

It is envisaged that these results, including the method of investigation, may prove useful not only for the physics of IMPATT diodes but also for the more general problem of studying $1/F$ noise in high fields.

TABLE II. Contribution of noise sources S'_{Nt} and S'_{vn} to S'_u, S'_α, and S'_f for diodes in batch 1. Data for S'_f and S'_α calculated for $\beta = 1.85$.

$S'_u = 1.1 \times 10^{-12}$ Hz^{-1}		$S'_f = 10^{-15}$ Hz^{-1}		$S'_\alpha = 1.1 \times 10^{-12}$ Hz^{-1}	
S'_{Nt} contribution	S'_{vn} contribution	S'_{Nt} contribution	S'_{vn} contribution	S'_{Nt} contribution	S'_{vn} contribution
9×10^{-13}	2×10^{-13}	9.3×10^{-16}	7×10^{-17}	5×10^{-16}	8.35×10^{-14}

[1] T. G. M. Kleinpenning, Physica B **103**, 340 (1980).

[2] T. G. M. Kleinpenning, Physica B **113**, 189 (1981).

[3] Th. G. van de Roer, Solid–State Electron. **23**, 695 (1980).

[4] Th. G. van de Roer, Physica B **168**, 53 (1991).

[5] S. A. Kornilov, K. D. Ovchinnikov, and V. M. Pavlov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **28**, 607 (1985).

[6] S. A. Kornilov and K. D. Ovchinnikov, *Noise in Oscillators, Amplifiers, and Frequency Multipliers Using IMPATT Diodes* [in Russian], St. Petersburg University Press, St. Petersburg (1993).

[7] F. N. Hooge, T. G. M. Kleinpenning, and L. K. J. Vandamme, Rep. Progr. Phys. **44**, 479 (1981).

[8] N. V. D'yakonova and M. E. Levinshteĭn, Fiz. Tekh. Poluprovodn. **23**, 1187 (1989) [Sov. Phys. Semicond. **23**, 743 (1989)].

[9] N. V. D'yakonova, M. E. Levinshteĭn, and S. L. Rumyantsev, Fiz. Tekh. Poluprovodn. **25**, 2065 (1991) [Sov. Phys. Semicond. **25**, 1241 (1991)].

[10] A. L. McWhorter, in *Semiconductor Surface Physics*, edited by R. H. Kingston (Pennsylvania University Press, Philadelphia, 1957), p. 207.

[11] Sh. M. Kogan and K. E. Nagaev, Solid State Commun. **49**, 387 (1984).

[12] Yu. M. Gal'perin and V. I. Kozub, in *Proceedings of the Fifth All-Union Conference on Fluctuation Phenomena in Physical Systems, Palanga, 1988* [in Russian], p. 7.

[13] M. J. Kirton and M. J. Uren, Adv. Phys. **38**, 367 (1989).

[14] V. Palenskis, K. Maknys, and A. Stadalnikas, in *Proceedings of the 13th International Conference on Noise in Physical Systems and* 1/*f Fluctuations, Palanga, 1995*, p. 285.

[15] V. B. Orlov and A. V. Yakimov, Physica B **154**, 175 (1987).

[16] S. A. Kornilov, K. D. Ovchinnikov, and I. Corbella, in *Proceedings of the 13th International Conference on Noise in Physical Systems and* 1/*f Fluctuations, Palanga, 1995*, p. 319.

[17] A. S. Tager and V. M. Val'd-Perlov, *IMPATT Diodes and Their Application in Microwave Technology* [in Russian], Sov. Radio, Moscow (1986).

[18] D. Delagebeadeuf, Rev. Tech. Thomson-CSFI, **1**, 309 (1969).

Translated by R. M. Durham

Use of the speckle effect to analyze vibrations of a rough surface

Yu. P. Presnyakov and V. P. Shchepinov

Moscow Engineering Physics Institute, 115409 Moscow, Russia

An analysis is made of a contactless method of recording the vibrations of a rough surface using the speckle effect. Dependences of the percent modulation of the photodetector electric signal on the vibration amplitude and the parameters of the optical system are given and this optical dependence is checked experimentally. Results of a spectral analysis of the vibrations of an electric motor casing are presented for illustration purposes. It is noted that this method may be used to study vibrations of heated surfaces and transparent media. © *1997 American Institute of Physics.* [S1063-7842(97)01108-2]

Vibrations of a solid cause elements of its surface to rotate through angles which are greatest at the nodes and smallest at the antinodes. Let us assume that Θ is the angle of rotation of such an element relative to an axis tangent to its surface. Then, when a section of the vibrating rough surface substantially smaller than the spatial period of the vibrations is illuminated by a laser beam, the reflected radiation is turned through the angle 2Θ. As a result, the random distribution of the reflected radiation intensity in the plane of the photodetector diaphragm positioned at a distance z from the illuminated section of the surface is shifted by $h(t)$, which is a periodic function of time

$$h(t) = 2\Theta(t)z, \qquad (1)$$

where $\Theta(t) \ll 1$.

The intensity of laser radiation reflected by a rough surface is described by a random functions of the coordinates[1] whose correlation interval δ (speckle size) is given by

$$\delta \approx (\lambda z)/d, \qquad (2)$$

where λ is the wavelength of the laser radiation and d is the size of the illuminated part of the surface.

When the speckles are displaced in the plane of the diaphragm as a result of vibration of the surface, the photodetector electric signal is a periodic function with a period equal to that of the surface vibrations. If the size of the diaphragm is $\Delta \ll \delta$, the amplitude of the electric signal for a photodetector with a linear characteristic is proportional to the optical intensity $J(t)$

$$u(t) \sim J(t). \qquad (3)$$

This approach to study the vibrations of a rough surface was considered in Ref. 2 where the diaphragm size $\Delta = 5$–10 μm satisfies the condition $\Delta \ll s$. In this case however, calibration measurements of the electric signal $u(t)$ must be made for a known displacement D of the diaphragm relative to the fixed speckle structure in order to determine the amplitude of the function $h(t)$. In accordance with formula (3), we have $h(t) = D$ for $u(t) = u_0(D)$.

Here we examine a method of investigating the vibrations of a rough surface by means of an integrated speckle effect, whereby the diaphragm area is $s \gg \delta^2$.

The proposed method is based on using a random dependence of the intensity, averaged over the diaphragm, on a specific realization of the speckle structure. This effect was analyzed theoretically in Refs. 3 and 4, where it was shown that the probability density of the distribution of the intensity I, averaged over the diaphragm, is approximately described by the gamma distribution

$$p(I) = \frac{1}{\Gamma(m)} \left(\frac{m}{J_0} \right)^m I^{m-1} \exp \left(\frac{-mI}{J_0} \right), \qquad (4)$$

where $I \geq 0$, J_0 is the average intensity in the plane of the diaphragm, $\Gamma(m)$ is a gamma function, $m \geq 1$ is the number of independent correlation cells, where $m \cong s/\delta^2$ for a two-dimensional diaphragm and $m \cong L/\delta$ for a slit diaphragm of length L, whose width is $\Delta \ll \delta$.

Periodic vibrations of the surface lead to periodic displacements of the speckle structure in the plane of the diaphragm as a result of which the output electric signal from the photodetector also varies periodically and the spectrum of this signal is proportional to the vibration spectrum of the surface.

The method of spectral analysis of the vibrations of a solid surface is shown schematically in Fig. 1a. Radiation from a cw laser *1* is focused by a lens *2* onto the section of surface being studied *3*. The reflected radiation is passed through a diaphragm *4* onto the sensitive surface of a photodiode *5*, from which the output electric signal is amplified by an amplifier *6* and passed via an analog-to-digital converter *7* to a computer *8*. Figure 1b show the plane of the diaphragm, which comprises a rectangular aperture with side lengths $2a$ and $2b$ in a nontransmitting screen. The axes of the Cartesian coordinate system (x, y) are parallel to the sides of the rectangular aperture and the origin is located at the center.

When the speckle structure is displaced in the plane of the diaphragm, the output electric signal from the photodetector is given by

$$u(t) = A \int_{-a}^{a} \int_{-b}^{b} J[x + h_x(t), y + h_y(t)] dx dy, \qquad (5)$$

where A is a constant which depends on the power of the reflected radiation in the region of the diaphragm and on the
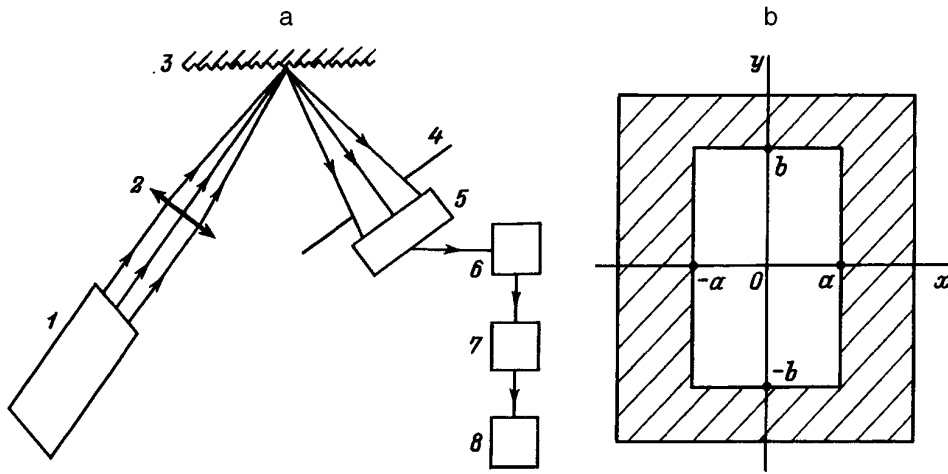
FIG. 1. Schematic of apparatus (a) and plane of photodetector diaphragm (b).

sensitivity of the photodetector, and $h_x(t)$ and $h_y(t)$ are the components of the displacement vector of the speckle structure $\mathbf{h}(t)$.

The amplitude h_0 of the displacement vector is related to the amplitude of the angle of rotation Θ_0 of the illuminated part of the surface by formula (1). For the displacement of speckles of amplitude $h_0 \ll \delta$, i.e., substantially smaller than the speckle size, we can confine ourselves to the first two terms of the series expansion of the intensity

$$J[x+h_x(t),y+h_y] \cong J(x,y) + \nabla \cdot J(x,y)\mathbf{h}, \qquad (6)$$

where ∇ is the gradient operator.

Substituting expression (6) into Eq. (5) and dropping the constant A, we obtain

$$u(t) = \int_{-a}^{a}\int_{-b}^{b} J(x,y)dxdy + \int_{-a}^{a}\int_{-b}^{b} \mathbf{h}\cdot\nabla J(x,y)dxdy. \qquad (7)$$

We write expression (7) in the form

$$u(t) = u_0 + u_\sim(t),$$

where $u_0 = 4abI$ is the constant component of the electric signal and $u_\sim(t)$ is the varying component, which has the form

$$u_\sim(t) = h_x \int_{-a}^{a}\int_{-b}^{b} \frac{\partial J(x,y)}{\partial x} dxdy + h_y \int_{-a}^{a}\int_{-b}^{b} \frac{\partial J(x,y)}{\partial y} dxdy. \qquad (8)$$

After integrating, expression (8) has the form

$$u_\sim(t) = \mathbf{h}(t)\mathbf{g}, \qquad \mathbf{g} = \{g_x, g_y\}, \qquad (9)$$

where

$$g_x = \int_{-b}^{b} [J(a,y) - J(-a,y)]dy,$$

$$g_y = \int_{-a}^{a} [J(x,b) - J(x,-b)]dx.$$

For a given amplitude h_0, the value of $|\mathbf{g}|$ determines the random amplitude of the varying signal $u_\sim(t)$, which is invariably nonzero since the probability that, for instance,

$$\int_{-b}^{b} J(a,y)dy = \int_{-b}^{b} J(-a,y)dy,$$

is zero according to the definition of the probability density function.

For terms forming the components of the vector $\mathbf{g}$ we have

$$I = \frac{1}{2b}\int_{-b}^{b} J(\pm a,y)dy$$

and $m = m_b = (2b)/\delta$ for

$$\int_{-b}^{b} J(\pm a,y)dy.$$

For the terms

$$\int_{-a}^{a} J(x,\pm b)dx$$

in formula (4), we have

$$I = \frac{1}{2a}\int_{-a}^{a} J(x,\pm b)dx$$

and $m = m_a = (2a)/\delta$.

It follows from Eqs. (1) and (9) that the spectrum of the electric signal $u_\sim(t)$ is proportional to the spectrum of the angle of rotation $\Theta(t)$. This result forms the basis of the proposed method of analyzing the mechanical vibrations of a rough surface.

The expression for the percent modulation γ of the electric signal for a specific realization of the speckle structure as given by Eq. (9) has the form

$$\gamma = \frac{|\mathbf{h}_0\cdot\mathbf{g}|}{u_0}.$$

The rms percent modulation f is given by

$$f^2 = \langle \gamma^2 \rangle = \left\langle \frac{(\mathbf{h}_0\cdot\mathbf{g})^2}{u_0^2} \right\rangle, \qquad (10)$$

where the angular brackets denote averaging over an ensemble of random intensity distributions inside the diaphragm.

In expression (10) the numerator and the denominator are statistically independent, so that

$$f^2 = \frac{\langle (\mathbf{h}_0 \cdot \mathbf{g})^2 \rangle}{\langle u_0^2 \rangle}.$$

Formula (7) gives

$$\langle u_0^2 \rangle = 16a^2 b^2 \langle I^2 \rangle,$$

where the random quantity

$$I = \frac{1}{4ab} \int_{-a}^{a} \int_{-b}^{b} J(x,y) dx dy$$

is distributed according to the law (4), from which it follows that

$$\langle I^2 \rangle = \frac{m+1}{m} J_0^2, \quad m = \frac{ab}{\delta^2}.$$

From the previously assumed condition $F \gg \delta^2$ it follows that $m \gg 1$ and

$$\langle I^2 \rangle \approx J_0^2, \quad \langle u_0^2 \rangle \approx 16a^2 b^2 J_0^2.$$

The average value of the numerator in Eq. (10) is

$$\langle (\mathbf{h}_0 \cdot \mathbf{g})^2 \rangle = h_{0x}^2 \langle g_x^2 \rangle + h_{0y}^2 \langle g_y^2 \rangle + 2h_{0x} h_{0y} \langle g_x g_y \rangle. \quad (11)$$

The random quantities g_x and g_y are statistically independent and their averages are zero, so that the last term on the right-hand side of Eq. (11) is zero. Using the statistical independence of g_x and g_y and the distribution (4), we easily find that

$$\langle g_x^2 \rangle = \frac{8b^2}{m_b} J_0^2, \quad \langle g_y^2 \rangle = \frac{8a^2}{m_a} J_0^2.$$

The final expression for the rms percent modulation has the form

$$f = \frac{\sqrt{\delta}}{2ab} \sqrt{bh_{0x}^2 + ah_{0y}^2}. \quad (12)$$

When the diaphragm is oriented orthogonal to its initial position, the rms percent modulation $f_\perp$ is

$$f_\perp = \frac{\sqrt{\delta}}{2ab} \sqrt{ah_{0x}^2 + bh_{0y}^2}. \quad (13)$$

Thus, by measuring the rms percent modulation for two orthogonal orientations of the diaphragm ($a \neq b$), we can determine the components of the displacement vector of the speckle structure using formulas (12) and (13) and consequently, the components of the vector of the angle of rotation of the surface point.

For a circular diaphragm of radius R, the amplitude of the displacement vector is determined from

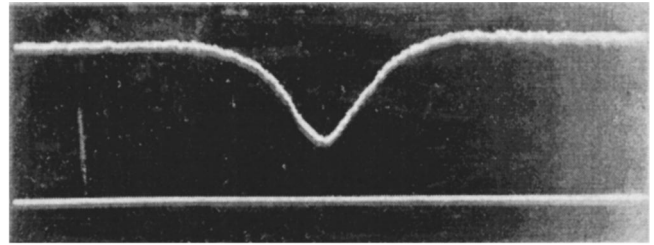$$f_0 = \frac{h_0}{R} \sqrt{\frac{\delta}{R}}. \quad (14)$$

FIG. 2. Oscilloscope trace of intensity correlation function.

Formula (14) was checked experimentally for $h_{0x} = h_0$, $h_{0y} = 0$. In this case, we have

$$f = \frac{h_0}{2a} \sqrt{\frac{\delta}{b}}. \quad (15)$$

The speckle size is determined by means of a speckle transparency[5] mounted on a two-coordinate micrometer table. The value of δ was determined from the e^{-1} level of the correlation signal at the photodetector exit as the speckle transparency was displaced. The measurement error for the displacement was $1\,\mu$m.

Figure 2 shows an oscilloscope trace of the intensity correlation function of the speckle structure with its negative image, obtained by displacement of the speckle transparency at a constant rate from one position of total decorrelation to another via the correlation extremum. In this case, the extremum is a minimum since the speckle transparency is a negative. The lower horizontal line was recorded with the laser switched off.

The percent modulation γ_i for the ith realization of n independent speckle structures was measured using a digital oscilloscope, with periodic transverse displacement of a photodiode mounted on the two-coordinate stand. For measurements of the speckle size, the plane of the diaphragm was matched with the plane of the speckle transparency. The amplitude h_0 was monitored from the digital readout of the displacement. An independent realization of the speckle structure for alternate measurements of γ_i was achieved by transverse displacement of the diaphragm and the photodetector by an amount greater than the speckle size δ. The rms percent modulation was determined by

$$f^2 = \frac{1}{n} \sum_{i=1}^{n} \gamma_i^2. \quad (16)$$

As a result of $n = 10$ measurements of γ_i for $h_0 = 0.3$ mm and $s = 0.9$ mm, the rms percent modulation determined using formula (16) was $f = 0.120$. The theoretical value of f obtained from formula (15) for $a = 0.5$ mm and $b = 5$ mm was 0.127. The difference between the experimentally measured value and the theoretical value is less than 6%. Thus, in order to make calibration measurements of the amplitude of the angle of rotation using this method, there is no need to bring the vibrating surface to a state of rest, which is important when studying the vibrations of industrial equipment with a continuous operating cycle.

A series of measurements of the correlation depth for various positions of the photodetector in the recording plane
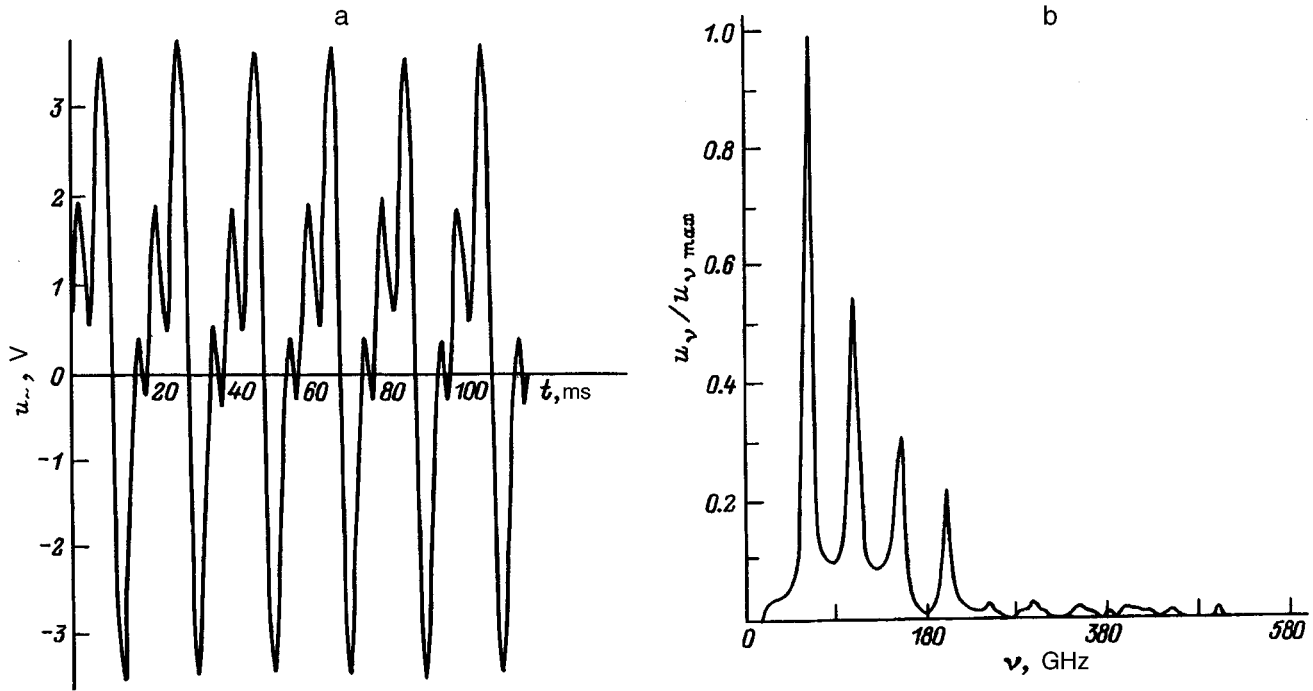
FIG. 3. Oscilloscope trace of photoelectric signal of the motor vibration (a) and its spectrum (b).

must be made to calibrate a vibration speckle detector, using a circular diaphragm for instance. Then, h_0 should be determined using formulas (16) and (14) and Θ_0 should be determined using formula (1). The condition $s=$const must be satisfied for the measured γ_i.

This method of analyzing vibrations was used to investigate the vibrations of the casing of an electric motor rotating at 50 Hz. The results are plotted in Fig. 3. An oscilloscope trace of the varying component of the photodetector electric signal, observed on the screen of an IBM PC used as an oscilloscope is shown in Fig. 3a. Figure 3b shows the vibration spectrum calculated by means of a fast Fourier transformation. The high-intensity lines correspond to frequencies of 50, 100, 150, and 200 Hz. The peaks correspond to the nonlinear vibration process of the part of the motor casing being studied.

It should be noted that, unlike the method using piezo-electric detectors, this method is contactless so that the vibrations of heated surfaces can be studied. The method can also be used to study the vibrations of transparent media inserted between mat glass and the photodetector. In this case, the displacement of the speckles in the region of the diaphragm is caused by refraction of the laser radiation in an optically inhomogeneous medium.

[1] *Laser Speckle and Related Phenomena*, edited by J. C. Dainty (Springer-Verlag, Berlin, 1975).
[2] K. J. Ebeling, Optik **54**, 295 (1979).
[3] A. A. Scribot, Opt. Commun. **11**, 238 (1974).
[4] R. Barakat, Opt. Acta **20**, 729 (1973).
[5] V. M. Dobrido, V. V. Manikalo, and Yu. P. Presnyakov, Opt. Spektrosk. **65**, 212 (1988) [Opt. Spectrosc. (USSR) **65**, 125 (1988)].

Translated by R. M. Durham

FIG. 3. Oscilloscope trace of photoelectric signal of the motor vibration (a) and its spectrum (b).

must be made to calibrate a vibration speckle detector, using a circular diaphragm for instance. Then, h_0 should be determined using formulas (16) and (14) and Θ_0 should be determined using formula (1). The condition $s=$const must be satisfied for the measured γ_i.

This method of analyzing vibrations was used to investigate the vibrations of the casing of an electric motor rotating at 50 Hz. The results are plotted in Fig. 3. An oscilloscope trace of the varying component of the photodetector electric signal, observed on the screen of an IBM PC used as an oscilloscope is shown in Fig. 3a. Figure 3b shows the vibration spectrum calculated by means of a fast Fourier transformation. The high-intensity lines correspond to frequencies of 50, 100, 150, and 200 Hz. The peaks correspond to the nonlinear vibration process of the part of the motor casing being studied.

It should be noted that, unlike the method using piezo-electric detectors, this method is contactless so that the vibrations of heated surfaces can be studied. The method can also be used to study the vibrations of transparent media inserted between mat glass and the photodetector. In this case, the displacement of the speckles in the region of the diaphragm is caused by refraction of the laser radiation in an optically inhomogeneous medium.

[1] *Laser Speckle and Related Phenomena*, edited by J. C. Dainty (Springer-Verlag, Berlin, 1975).
[2] K. J. Ebeling, Optik **54**, 295 (1979).
[3] A. A. Scribot, Opt. Commun. **11**, 238 (1974).
[4] R. Barakat, Opt. Acta **20**, 729 (1973).
[5] V. M. Dobrido, V. V. Manikalo, and Yu. P. Presnyakov, Opt. Spektrosk. **65**, 212 (1988) [Opt. Spectrosc. (USSR) **65**, 125 (1988)].

Translated by R. M. Durham

Some characteristics of relief formation on photothermoplastic carriers used in double-exposure interferometry

L. M. Panasyuk and I. V. Chapurin

Moldavian State University, 277009 Kishinev, Moldavia

An analysis is made of the main trends in the development of a surface relief on photothermoplastic carriers of optical information used to record double-exposure holographic interferograms. It is shown that high-quality interferograms require recording conditions (taking into account the composition of the material and the heating temperature of the thermoplastic carrier layer, as well as the parameters of the sensitizing corona discharge) where the surface of the photothermoplastic carrier can be treated in the equipotential approximation. © *1997 American Institute of Physics.* [S1063-7842(97)01208-7]

INTRODUCTION

We report results of investigations of the processes of relief formation during the recording of double-exposure interferograms on a photothermoplastic carrier (PTPC). This carrier consists of a double-layer structure, formed by a photosensitive semiconductor and a thermoplastic polymer coating layer, deposited on a metallized substrate. This type of PTPC is charge-sensitized in the field of a corona discharge, accompanied by simultaneous heating and exposure.[1] The mechanism of relief formation in this method of recording optical information has been studied in fairly great detail by various authors, as in Refs. 2–5 for instance.

One of the important practical applications of PTPCs is in double-exposure holographic interferometry,[6,7] where two holographic images of a test object in different states are recorded consecutively on a continuously heated PTPC. This process incorporates two corona charging cycles of the PTPC and the time interval between charging is utilized to create conditions conducive to high-quality recording of the second hologram on the surface relief formed on the PTPC by recording the first hologram. In general, the problem of simulating the processes taking place in a PTPC system during the formation of double-exposure interferograms, reduces to identifying how the relief formed initially on the PTPC surface tends to vary under additional exposure to deforming forces.[8]

If a surface relief having the spatial frequency k and depth $A = A_0 \cos kx$, simulating the initially recorded holographic image, is formed on the PTPC surface before this undergoes repeated charging, the corona charging of the PTPC also modulates the surface charge density harmonically $\sigma = \sigma_0 + \sigma_1 \cos kx$. The deformation amplitude A_0 is then clearly much smaller than the layer thickness d. The equation of motion of the thermoplastic layer is then written (as in Ref. 9) as

$$\tau_m \frac{dB}{dt} + (1 - Fe^{-\omega_f t})B + Ye^{-\omega_y t} = 0, \tag{1}$$

where $B = Ak$ is the normalized depth of the surface relief, τ_m is the characteristic time for mechanical relaxation of a viscoelastic liquid modeling the thermoplastic layer, the coefficients Y and ω_y characterize the evolution of the relief of the thermoplastic layer exposed to a density-modulated charge of amplitude σ_1 and frequency k, and the coefficients F and ω_f characterize the evolution of the relief under uniform charging of the thermoplastic layer with the surface charge density σ_0.

INVESTIGATION OF THE INFLUENCE OF CORONA CHARGING OF A DEFORMED PHOTOTHERMOPLASTIC CARRIER SURFACE ON THE EVOLUTION OF RELIEF DEPTH

We shall assume that the charged PTPC surface is equipotential. This approximation characterizes a thermoplastic carrier layer with a low surface resistivity, allowing the thermoplastic surface to be treated as ''metallized.'' We examine the main trends systematically characterizing the time evolution of the PTPC surface relief.

1. Corona charging of a PTPC surface exhibiting some initial relief is accompanied by effective modulation of the surface charge as a result of redistribution of the corona current in the indentation of the surface relief.[10] This increases the density of the deforming forces acting on the initially formed surface deformations of the PTPC and increases the depth of the initial relief as the surface potential reaches its equilibrium value over the entire PTPC free surface.

2. After the PTPC surface has reached an equipotential state, an appreciable surface charge density σ_0, much greater than the modulated density σ_1, builds up on the surface. Since the latter is proportional to Y, it may be assumed that $Y = 0$ and Eq. (1) may be rewritten as

$$\tau_m \frac{dB}{dt} + (1 - Fe^{-\omega_f t})B = 0. \tag{2}$$

The solution of Eq. (2) for times $t \ll \omega_f^{-1}$ has the form

$$B(t) = B(0)e^{-\frac{(1-F)t}{\tau_m}}. \tag{3}$$

These calculations show that the relation $F > 2$ is usually satisfied for these types of PTPC. Thus, the depth of the surface relief continues to increase exponentially with growth rate $(F - 1)$ times greater than τ_m^{-1}.

3. As the depth of the surface relief increases, so does the parameter ω_f (Ref. 9). At a certain time corresponding to

the optimum duration of corona charging of the PTPC, the parameter ω_f increases to such an extent that the exponential term in Eq. (2) becomes negligible and deepening of the surface relief is replaced by infilling with a characteristic depth reduction time equal to τ_m.

4. Cessation of corona charging of the PTPC with continued heating slows the reduction in the depth of the surface relief, with the time constant $\tau_m^* > \tau_m$. The time constant τ_m^* is increased because in the absence of corona charging, the thermoplastic layer is simulated, not by a viscoelastic fluid, but by a Newtonian viscous fluid.

The choice of thermoplastic layers with enhanced surface resistivity, where a ''dielectric'' model representation of the system is achieved, determines the value of $\omega_f < 0$, and the increase in the surface deformations accompanying corona charging of the PTPC is not limited by the factor described above. In this case, the relief develops over the total thickness of the thermoplastic layer (''funnel'' deformation). However, at high spatial frequencies deformations from neighboring interference lines begin to partially overlap in the recorded holographic images. This leads to a reduction of the spatial frequency in the recorded image, thus limiting the mechanical breakdown of the thermoplastic layer but at the same time, appreciably reducing the quality of the image recorded in the PTPC.

INFLUENCE OF UNIFORM ILLUMINATION OF A CHARGED PHOTOTHERMOPLASTIC CARRIER ON THE EVOLUTION OF ITS SURFACE RELIEF

If a deformed PTPC surface is exposed to uniform illumination at the same time as corona charging, the influence of this illumination reduces the surface potential V_s, as a result of increased leakage currents across the PTPC, and increases its differential capacitance C because of the accumulation of negative charges at the interface between the semiconducting and thermoplastic layers. Our experiments to study the kinetics of the variation in V_s, using the apparatus described in Ref. 11, showed that under illumination the decrease in V_s is on a larger scale than the increase in C, and this tendency is enhanced with increasing illumination E. Thus, both the total surface charge $\Omega = V_s C$, and the surface charge density σ_0 determined by it decrease with increasing illumination of the PTPC. Bearing in mind that $F \simeq \sigma_0^2$ (Ref. 9), we note that an increase in E appreciably reduces F. If the relation $F > 1$ is satisfied for certain values of E, the main laws governing the evolution of the PTPC relief described above are conserved. However, from a certain illumination the situation $F < 1$ begins to arise. We shall consider this in greater detail as a function of the sign of the parameter ω_f.

a) The case $\omega_f > 0$, characterizing a thermoplastic with an equipotential (''metallized'') surface. Transition processes leading to the establishment of an equipotential state and characterized by appreciable deepening of the relief on the PTPC are followed by a stage involving exponential infilling of surface deformations, as given by Eq. (3), with time constants exceeding τ_m by the amount by which F is close to unity. Thus, the role of uniform illumination accompanying charging of a PTPC with an initial surface relief reduces to changing this relief by the required amount.

b) The case $\omega_f < 0$, characterizing a thermoplastic with a high surface resistivity (''dielectric'' approximation). In this case, the situation is similar to the charging of an unilluminated PTPC analyzed above, and leads to mechanical breakdown of the thermoplastic layer. The depth of the surface relief is not modulated as a function of the illumination. In addition, it is preferable to use a mechanism of recording on the PTPC for which the growth of surface deformation is limited (or controlled) in some way.

EVOLUTION OF A SURFACE RELIEF FORMED INITIALLY ON A PHOTOTHERMOPLASTIC CARRIER DURING REPEATED RECORDING OF A HOLOGRAPHIC IMAGE

Since the surface potential V_s decreases with increasing illumination of a charged PTPC, the additional charging current becomes redistributed on the most strongly illuminated parts of the PTPC.[12] Bearing in mind that the total charging current is constant over the entire PTPC, we note that this causes a decrease in the charging current in the less illuminated parts of the PTPC, which then reduces the values of σ_0 and the related coefficient F in these sections. In addition, the parameter ω_f increases as a result of decrease in the resistivity of the semiconducting layer of the PTPC as the illumination is increased.[9] All these factors interact to modulate the depth of the initially formed PTPC relief as it evolves as a function of the distribution of the illumination in the recorded second image. The recording of a double-exposure interferogram may involve the simultaneous reconstruction of the initially recorded hologram and modulation of its surface relief by the holographic image recorded by the second. This is the mechanism for the recording of a double-exposure interferogram on a PTPC.

We consider how the parameters of the thermoplastic layer influence the quality of recording a double-exposure interferogram.

a) Thermoplastic in the dielectric approximation ($\omega_f < 0$). We have shown above that the initially formed relief first undergoes partial infilling, and only for $t > |\omega_f^{-1}|$ does it begin to deepen. Since ω_f depends on the illumination of the different parts of the PTPC, the transition to enhancement of a partially infilled relief image occurs at different times for parts of the PTPC with different recorded illuminations. In this case however, the first of the recorded holographic images will predominate in the resultant interferometric image and the double-exposure interferogram will possess low contrast.

b) Thermoplastic in the metallic approximation ($\omega_f > 0$). If there is a fairly large difference between the minimum and maximum illuminations in the recorded second holographic image, we observe a situation where each of the dark and light interference lines may be represented as an equipotential surface characterized by different values of the coefficient F: $F > 1$ and $F < 1$, respectively. In the first case, the depth of the relief first increases and then decays exponentially with the characteristic time τ_m. In the second situation, this exponential decay is preceded by a section where the reduction in the depth of the surface relief is slowed without being preliminarily deepened. We then have a situation where reconstruction of the first holographic image is ac-

companied by the formation of an image of the hologram exposed by the repeated corona charging. Thus, a double-exposure interferogram is formed throughout the PTPC charging process and the quality of the recorded double-exposure interferogram is almost independent of the corona charging duration.

CONCLUSIONS

These characteristic features observed in the formation of a double-exposure interferogram on an PTPC suggest that the evolution of an initially applied surface relief under repeated charging of the PTPC is influenced by the heating parameters of the thermoplastic layer and by the conditions of charge sensitization and exposure of the PTPC. A decisive factor for the high-quality recording of a double-exposure interferogram on a PTPC is the choice of recording parameters which allows the surface of the thermoplastic layer to be treated as metallized in the corona charging regime.

[1] L. M. Panasyuk, *Data Storage Methods in Silverless Media*, Vol. 8 [in Russian], Vishcha Shkola, Kiev (1977), pp. 12–24.

[2] A. Handojo, J. Appl. Phys. **50**, 886 (1979).

[3] E. Storck and U. Wolff, Siemens Forsch. Entwickl. **6**, No. 2, 63 (1977).

[4] V. I. Vybornov, L. M. Panasyuk, and M. M. Rusanov, Zh. Tekh. Fiz. **54**, 929 (1984) [Sov. Phys. Tech. Phys. **29**, 554 (1984)].

[5] V. I. Anikin, *Optical Memory and Media: Basic Principles*, Vol. 20 [in Russian], Vishcha Shkola, Kiev (1989), pp. 9–13.

[6] V. I. Anikin, O. Ya. Korshak, L. M. Panasyuk *et al.*, *Properties of Light-Sensitive Materials and Their Application in Holography* [in Russian], Nauka, Leningrad (1987), pp. 100-105.

[7] L. M. Panasyuk and I. V. Chapurin, Opt. Spektrosk. **81**(1), 140 (1996) [Opt. Spectrosc. **81**, 126 (1996)].

[8] I. V. Chapurin and L. M. Panasyuk, *ICO-17 Congress, Optics for Science and New Technology*.

[9] Yu. P. Gushcho, *Properties and Scope of Silverless Photography* [in Russian], Khimiya, Leningrad (1988), pp. 196–239.

[10] V. I. Anikin, I. L. Zhurminskiĭ, and L. M. Panasyuk, *Optical Memory and Media: Basic Principles*, Vol. 18 [in Russian], Vishcha Shkola, Kiev (1987), pp. 41–43.

[11] L. B. Pogorel'skiĭ and L. M. Panasyuk, Inventor's Certificate (USSR) No. 1341593.

[12] Yu. I. Denisenko, L. M. Panasyuk, and P. N. Yakovlenko, Zh. Nauchn. Prikl. Fotogr. Kinematogr. **27**, 241 (1982).

Translated by R. M. Durham

Kinetics of luminescence flashes accompanying the formation of microdamage in alkali halide crystals exposed to CO_2 laser radiation pulses

V. N. Smirnov

S. I. Vavilov State Optical Institute All-Russia Science Center, 199034 St. Petersburg, Russia

Since the seventies through to the present day, considerable attention has been devoted in the literature to studying the evolution of optical breakdown in the bulk of alkali halide crystals initiated by heating of absorbing inhomogeneities contained in these crystals, under the action of CO_2 laser radiation with different pulse parameters.[1–18] The breakdown thresholds according to data given by various authors cover a wide range of values between $\sim 1-10$ (Refs. 1–11) and 300 MW/cm^2 (Ref. 13). One of the main reasons for such a large discrepancy is the well-known dependence of the breakdown threshold on the diameter D of the effective radiation spot which leads to an increase in the experimentally measured threshold with decreasing D. In fact, in Refs. 1–11 the measurements were made for $D \geq 0.2$ mm and in Ref. 13, for $D \sim 0.070$ mm. Another factor may be that the breakdown threshold also depends on the concentration and absorption cross sections of the inhomogeneities in the samples.

The results of Refs. 2 and 3 indicate that the number of damage sites formed within an irradiated zone and the intensity of the accompanying luminescence formation depend not only on the laser radiation intensity q but also on the number N of pulses acting at the same point on the sample. For $N=1$ within the caustic of the lens used to focus the laser radiation into the sample, it is observed that a large number (for $D \geq 0.2$ mm it is several tens) of damage sites is formed and the formation of each one is accompanied by luminescence. As a result of multiple exposure ($N \geq 2$) of the same part of the crystal to pulses of constant radiation intensity ($q = $const), the vast majority of the previously formed damage sites do not luminescence. However, new damage sites are formed, accompanied by luminescence, and some luminescence is observed from a negligible number of the already formed damage sites. Typically, the morphology and dimensions of the damage sites which do not luminesce under repeated exposure, do not vary with increasing N whereas those damage sites which do luminesce under each exposure expand and their morphology changes from one pulse to another. As N increases, the number of newly formed damage sites decreases to zero. The integrated intensity of the luminescence flash either also decreases to zero (under conditions where no new damage sites are formed and the old ones luminesce repeatedly) or it undergoes irregular oscillations (with repeated luminescence from previously formed damage sites) about some average, which is between two and three orders of magnitude lower than the intensity of the first flash.[3] Results of studying changes in the damage morphology under multiple irradiation of a single damage site were described in Ref. 14.

Particular attention was paid in Refs. 1–18 to studying the various phenomena accompanying the evolution of breakdown, particularly the kinetics[3–5,17,18] and luminescence spectra,[6,7,15,17] and also the dynamics of the growth of breakdown centers[9,16] and the kinetics of the scattering caused by these centers.[4,18] The kinetics and spectra of the luminescence flashes were investigated in Refs. 3–7, for example, at radiation intensities (1–10 MW/cm^2) possibly closer to the threshold q^* for the appearance of luminescence flashes, whereas the intensities used in Refs. 15–18 were 100–1000 MW/cm^2 and the value of q^* was not indicated.

An analysis of these experimental conditions,[1–18] and of their results and conclusions on the luminescence mechanism indicates that two main viewpoints were adopted as to the nature of the luminescence flashes accompanying breakdown. One of these viewpoints was based on experimental data obtained for $D \geq 0.2$ mm and $q \sim q^* \sim 1-10$ MW/cm^2. In this case, the luminescence flashes may consist of individual more or less resolved pulses[3–5] and the luminescence spectra exhibit narrow peaks (in Refs. 6 and 7, for instance, of $\sim 1 \mu$m width which is determined by the spectral resolution of the apparatus). The size of the damage sites under these conditions is $\sim 10 \mu$m. The authors of Refs. 3–6 put forward the hypothesis that the observed luminescence is mainly attributable to the crystal triboluminescence caused by cracking under the influence of thermoelastic stresses created near heated inhomogeneities. The authors of Ref. 6 discuss possible reasons for the differences in the spectra of the triboluminescence accompanying crystal damage near laser-heated absorbing inhomogeneities and excited by conventional methods, such as grinding the samples.[19] According to estimates made for $q \sim q^* = 10$ MW/cm^2 (Ref. 3), the heating temperature of the most dangerous absorbing inhomogeneities only reaches ~ 1000 K by the end of the laser pulse.

Another viewpoint is based on the results of measurements made for small D and $q \sim 100-100$ MW/cm^2 (with no indication of q^*) (Refs. 13–18). Under these conditions the luminescence flashes are fairly smooth and no narrow peaks are observed in the spectra. Some of the observed luminescence spectra are well approximated by a Planck curve with temperatures of 5000–15 000 K, and the dimensions of the damage sites are $\sim 100 \mu$m. The authors of this series of studies hold the view that the luminescence flashes are of a thermal nature. It was also emphasized in Ref. 15
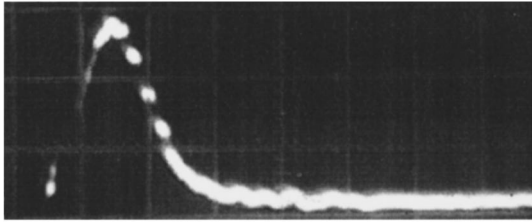
FIG. 1. Laser pulse. Horizontal scale sweep rate 50 ns/div.

FIG. 3. Laser pulse (*1*), and kinetics of luminescence flashes (*2, 3*). Horizontal scale 500 ns/div: $q=11$ (*2*) and 10 MW/cm^2 (*3*), $N=1$ (*2*) and 3 (*3*).

that triboluminescence should be discarded as a possible major factor in the luminescence observed when absorbing inhomogeneities are heated by laser radiation. Such a statement made without specifying the range of q values for which the results are valid is obviously unjustifiably categorical.

The contradictory nature of these results and treatments of the nature of the luminescence accompanying laser breakdown damage in crystals provided the basis for the present study. Here we present results of investigations of the kinetics of luminescence flashes carried out for $q \sim q^*$ which indicate that the observed form of the flashes under these conditions cannot be explained using concepts whereby the recorded luminescence is of a thermal nature.

The experiments were carried using an apparatus which differed little from that described in Ref. 3. The laser radiation was focused into the bulk of the crystals by an NaCl lens of ~ 250 mm focal length. The diameter of the effective radiation spot at $1/e$ of the maximum intensity was ~ 0.2 mm. Signals from an FD-0.5 detector using carrier increase, which recorded the shape of the CO_2 laser pulses, and from an FÉU-84-3 photomultiplier, which recorded the luminescence flashes, were fed to an S8-14 oscilloscope. The laser pulse, typical of TEA CO_2 lasers, consisted of a short leading spike of ~ 50 ns half-height duration (Fig. 1) and an ~ 1.5 μs tail (Fig. 2). The spike contained approximately $\sim 50\%$ of the total pulse energy. The measurements were made at $q \sim 10$–12 MW/cm^2 whereas the threshold is $q^* \sim 9 \pm 1$ MW/cm^2.

The shape of the luminescence flashes for single (a new point in the bulk of the sample was irradiated each time) and multiple (successive irradiation of the same site) exposure to laser pulses, respectively, was investigated for arbitrary values of q and for $q=$const. Only the flashes corresponding to
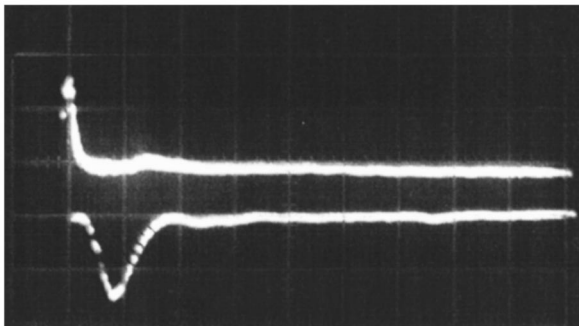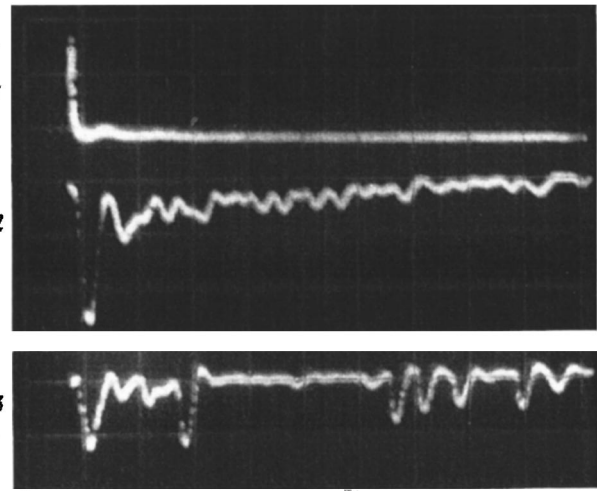
the formation and luminescence of one damage zone were recorded. For cases of multiple exposure luminescence was observed from the same zone with successively changing features.

The intensity and shape of the flashes depends on q and on the number of pulses which had previously acted on the irradiated site. Oscilloscope traces of the observed luminescence flashes are shown in Figs. 2–4. Figures 3 and 4 show the kinetics of the luminescence flashes for the successive action of several laser pulses ($q=$const) when the radiation is focused at the same point for two samples. On exposure to the first laser pulse, the flash intensity is higher than that for the subsequent pulses. This flash consists of a larger number of more or less overlapping spikes than those for the subse-
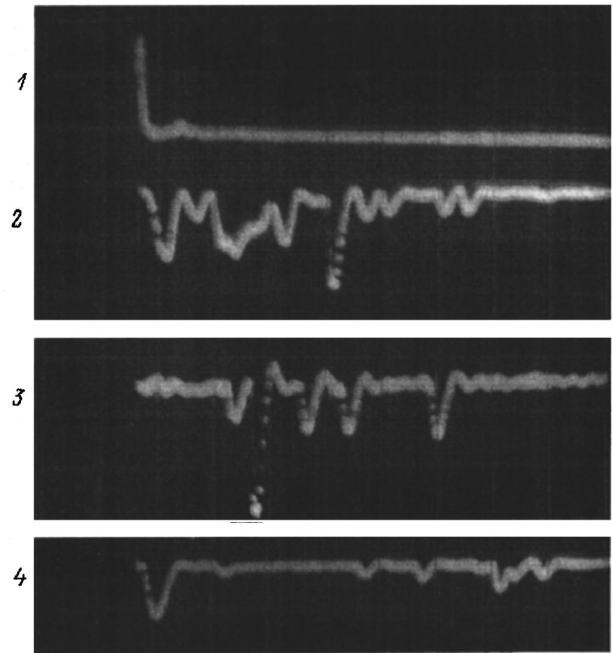

FIG. 4. Laser pulse (*1*) and kinetics of luminescence flashes (*2– 4*). Horizontal scale 500 ns/div: $q=10$ MW/cm^2, $N=3$ (*2*), 4 (*3*), and 5 (*4*).

FIG. 2. Laser pulse (upper trace) and kinetics of luminescence flash: $q=10$ MW/cm^2, $N=5$. Horizontal scale 200 ns/div.

quent pulses. An increase in q increases the number and amplitude of the spikes, and also their overlap. In most cases, the first spike, whose onset approximately corresponds to the maximum of the leading laser pulse spike, has a larger amplitude than the others but the reverse situation is also possible. In some cases, the amplitude of the first flash spike may be close to zero (Fig. 4).

The amplitude of the following spikes and their position within the flash varies when the irradiated site is changed and for multiple irradiation of the same site, which is responsible for the irreproducibility of the shape of the luminescence flashes from one pulse to another. When the crystals are irradiated by pulses with $q \sim q^*$, the spikes with the largest amplitudes are generally found in a range of ~ 2 μs from the beginning of the laser pulse. Outside this range, the intensity and number of the spikes decreases. Nevertheless, their intensity within a range (~ 3 μs) twice the laser pulse length remains comparable with the intensity of the spikes at the beginning of the flash. The authors of Refs. 4, 5, and 14 merely reported the observation of delayed luminescence pulses within the flash. The nature of the observed kinetics of the luminescence flashes cannot be attributed to a thermal mechanism. However, these results agree with the previously developed concepts whereby the flashes are caused by the triboluminescence[2,3] accompanying cracking in the barely heated zone surrounding a heated inhomogeneity. These cracks should be formed under the influence of thermoelastic stresses at the initial stage of the damage formation. In most cases, a further supply of laser pulse energy leads to crack propagation, the appearance of new cracks, and ultimately to cavity formation.[1–3,10–14]

An estimate of the heating temperature of absorbing inhomogeneities with an absorption coefficient $\beta \sim 10^3 - 10^4$ cm^{-1} at the beginning of a luminescence flash corresponding to the maximum of the leading spike of the laser pulse for $q \sim 12$ MW/cm^2, made in accordance with Ref. 20, gives $T = (3/4)(\beta q \tau / 2c) \sim 100 - 1000$ K, where $\tau = 50$ ns is the half-height duration of the leading spike of the laser pulse and $c = 2$ J/cm$^3 \cdot$K is the specific heat per unit volume of the inhomogeneity, which is assumed to be the same as the specific heat of the crystal. The thermoelastic stresses formed near the heated inhomogeneity may be estimated as $\sigma \sim \alpha E T$ (Ref. 20), which gives $\sigma = 1600 - 16\,000$ kg/cm$^2 \gg \sigma^* \sim 20$ kg/cm^2, where $\alpha = 4 \times 10^{-5}$ 1/K is the coefficient of linear expansion, $E = 4 \times 10^5$ kg/cm^2 is Young's modulus, and σ^* is the macroscopic tensile strength of the material. The use of the macroscopic tensile strength is clearly justified here, at least for the crystal zone surrounding the inhomogeneity.

Changes in the morphology of the damage zone can be identified by examining under a microscope samples exposed to successive laser pulses: a primary damage center is formed, consisting of a micrometer-sized cavity from which cracks emanate along the cleavage planes, new cracks appear and old cracks show systematic growth, accompanied by luminescence. In some cases, under the action of multiple pulses, luminescence can be detected from a growing crack propagating in the opposite direction to the laser radiation. The distance between the luminescent region in the crack and the initial damage site may be tens of micrometers. The

mechanism for the growth of such a crack, separated from the initially absorbing region of the crystal has not been fully clarified. However, from these observations the observed kinetics of the luminescence flashes may be attributed to irregularity of the triboluminescence, by ascribing the appearance of each spike of a flash to the expansion of an individual crack or ascribing its active section to an enhanced density of electric charges at the walls.[21]

A comparison between data reported here and the results of Refs. 15–18 indicates that here, as in Ref. 2–7, where the authors advanced the hypothesis that the luminescence accompanying the evolution of breakdown in crystals is of a triboluminescent nature, attention was focused on the near-threshold luminescence, whereas in Refs. 15–18, the luminescence accompanying the advanced stage of breakdown was most likely studied. This is particularly evidenced by the large dimensions of the damage centers (up to 100 μm), which are almost an order of magnitude greater than those observed in Refs. 2–7. Since an increase in q is accompanied by an increase in the number of spikes in the luminescence flash and by their overlap, it is highly unlikely that resolved spikes could be observed for a considerable excess over q^*. This should also be promoted by a systematic increase (with increasing q) in the contribution made by the thermal radiation from the absorbing inhomogeneities to the recorded luminescence flashes as their heating temperature increases. The contribution of this thermal radiation in the range $q \sim 10^8 - 10^9$ W/cm^2, for which the investigations were made in Refs. 15–18, was clearly a decisive factor. In this context, the fact that fairly smooth luminescence flashes were observed in Refs. 17 and 18, whereas in Refs. 15 and 17 the spectra of the luminescence flashes were accurately described by a Planck curve with temperatures of 5000–15 000 K, is not unexpected and does not contradict the present results or those observed in Ref. 12 since in this case, the spectra of the luminescence flashes in KCL, for instance, were investigated for $q \sim 3$ MW/cm^2.

To sum up, a comparison between the published data and our results indicates that the observed luminescence flashes are initiated by heating of absorbing inhomogeneities contained in the crystals, and the characteristics of the recorded luminescence should be determined by the relation between the contributions of the triboluminescence and the thermal radiation. At known above-threshold laser radiation intensities, the contribution of the thermal radiation predominates whereas at near-threshold intensities, the triboluminescence prevails. Luminescence flashes consisting of time-resolved short (~ 100 ns) spikes have been observed for the first time, indicating that these flashes are of a nonthermal nature.

[1] I. Apostol, L. C. Arsenovici, I. M. Mihailescu *et al.*, Rev. Roum. Phys. **20**, 665 (1975).
[2] V. P. Krutyakova and V. N. Smirnov, Pis'ma Zh. Tekh. Fiz. **2**, 1111 (1976) [Sov. Tech. Phys. Lett. **2**, 436 (1976)].
[3] V. P. Krutyakova and V. N. Smirnov, Zh. Tekh. Fiz. **48**, 844 (1978) [Sov. Phys. Tech. Phys. **23**, 495 (1978)].

[4] V. P. Krutyakova and V. N. Smirnov, Zh. Tekh. Fiz. **49**, 2647 (1979) [Sov. Phys. Tech. Phys. **24**, 1495 (1979)].

[5] V. P. Krutyakova and V. N. Smirnov, Kvantovaya Elektron. (Moscow) **8**, 2017 (1981) [Sov. J. Quantum Electron. **11**, 1225 (1981)].

[6] V. P. Krutyakova and V. N. Smirnov, Zh. Prikl. Spektrosk. **32**, 1002 (1980).

[7] L. M. Faĭnberg, V. P. Krutyakova, and V. N. Smirnov, Zh. Prikl. Spektrosk. **41**, 736 (1984).

[8] Ya. A. Imas, T. I. Kalugina, V. P. Krutyakova, and V. N. Smirnov, Pis'ma Zh. Tekh. Fiz. **9**(3), 129 (1983) [Sov. Tech. Phys. Lett. **9**, 57 (1983); **9**(21), 1343 (E)].

[9] V. P. Krutyakova and V. N. Smirnov, Zh. Tekh. Fiz. **54**, 323 (1984) [Sov. Phys. Tech. Phys. **29**, 189 (1984)].

[10] V. S. Golubev and V. N. Snopko, Fiz. Tverd. Tela (Leningrad) **19**, 293 (1977).

[11] V. E. Rogalin, T. I. Samoĭlova, N. A. Tishchenko, and M. A. Shaskol'skaya, Fiz. Tverd. Tela (Leningrad) **22**, 3549 (1980) [Sov. Phys. Solid State **22**, 2079 (1980)].

[12] V. E. Rogalin, N. A. Tishchenko, and M. P. Shaskol'skaya, Zh. Tekh. Fiz. **50**, 1077 (1980) [Sov. Phys. Tech. Phys. **25**, 648 (1980)].

[13] A. V. Gorbunov, E. M. Nadgornyi, and S. N. Valkovskii, Phys. Status Solidi A **66**, 53 (1981).

[14] A. V. Gorbunov, V. Ya. Emelin, and N. V. Klassen, Zh. Tekh. Fiz. **54**, 1983 (1984) [*sic*].

[15] A. V. Gorbunov, N. V. Klassen and M. Yu. Maksimuk, Zh. Tekh. Fiz. **62**(12), 39 (1992) [Sov. Phys. Tech. Phys. **37**, 1152 (1992)].

[16] A. V. Gorbunov and M. Yu. Maksimuk, Fiz. Tverd. Tela (St. Petersburg) **36**, 1416 (1994) [Phys. Solid State **36**, 2599 (1994)].

[17] A. V. Gorbunov and M. Yu. Maksimuk, Fiz. Tverd. Tela (St. Petersburg) **36**, 1430 (1994) [Phys. Solid State **36**, 2624 (1994)].

[18] A. V. Gorbunov and M. Yu. Maksimuk, Fiz. Tverd. Tela (St. Petersburg) **37**, 2568 (1995) [Phys. Solid State **37**, 1410 (1995)].

[19] V. P. Krutyakova and V. N. Smirnov, Zh. Prikl. Spektrosk. **30**, 846 (1979).

[20] M. Sparks and C. J. Duthler, J. Appl. Phys. **44**, 3038 (1973).

[21] L. M. Belyaev and V. V. Nabatov, Kristallografiya **8**, 927 (1963) [Sov. J. Crystallogr. **8**, 744 (1963)].

Translated by R. M. Durham

Influence of nonuniformities of a magnetic-mirror field on the space–time characteristics of a long-pulse relativistic electron beam

L. Yu. Bogdanov, N. V. Dvoretskaya, G. G. Sominskiĭ, and A. Ya. Fabirovskiĭ

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

An experimental investigation is made of the influence of local nonuniformities of a mirror-configuration magnetic field on oscillations of the space charge and the structure of a long-pulse relativistic electron beam. It is found that the outcome depends on the axial configuration of the nonuniformity. A nonuniformity near the cathode can substantially reduce the amplitude of the oscillations and improve the beam transport. The creation of a nonuniformity far from the cathode leads to an accelerated increase in the oscillations and causes spreading of the transverse structure of the beam. A possible explanation is given for the mechanism responsible for the influence of these local magnetic field nonuniformities assuming reflection of the cathode plasma and electron flux from the magnetic mirror, and also allowing for a jump in the drift velocity. © *1997 American Institute of Physics.* [S1063-7842(97)01408-6]

INTRODUCTION

The possibility of using relativistic electron beams to produce superpower microwave devices (see Refs. 1–3, for example) has stimulated interest in studying their space–time characteristics. Unfortunately, only fragmentary information is available on the space charge oscillations and structure of a relativistic electron beam. Data on long-pulse relativistic electron beams is particularly scarce. This is mainly because of the difficulties involved in experimental studies of high-current, high-energy, electron beams. A method developed by us earlier[4–7] provides information on the ''instantaneous'' characteristics of space charge oscillations in a wide frequency band and in different sections of the beam transport channel. This method was used to determine the characteristics of the oscillation spectrum and to identify their spatial variations.[6,7] In order to refine this model of collective processes, we consider how local nonuniformities of a mirror-configuration magnetic field in the beam transport channel influence the space-charge oscillations and structure of a long-pulse relativistic electron beam.

MEASUREMENT METHOD AND APPARATUS

The measurements were made using the SÉR-1 system described in Ref. 4 and 8. A cross section through the beam shaping and transport system is shown schematically in Fig. 1a. Electrons from a 20 mm diameter, explosive-emission, edge cathode *1*, made of stainless steel, are accelerated by a voltage *U* in the gap between the cathode and an anode *3*, then drift in the magnetic field inside a beam transport channel *4* connected electrically to the anode, before being deposited on a collector *6* in a decaying magnetic field. The transport channel had a diameter of 32 mm and a length of ~ 1.2 m. The cathode was positioned 27 mm from the anode.

The collector current I_{col} reached 1.1 kA per pulse. Its time variation was monitored by studying the space charge oscillations using the x-ray emission from the collector.[7] Since the discharge circuit of the pulse voltage generator had a finite resistance, the voltage *U* was reduced from 220 to 180 kV at the leading edge of the I_{col} current pulse.

The magnetic field was generated using four coaxial main solenoids *7*, together with coils *9* and *10* positioned 0.2 m and 0.6 m from the cathode, respectively. When the coils were switched off, a pulsed magnetic field B_0 of approximately 10 ms duration, was generated with nonuniformities not exceeding 10% far from the edges of the solenoid system. The distribution of the magnetic induction B_0 along the *z* axis is shown by the solid curve in Fig. 1b. The coils connected in series with the solenoids were used to introduce local nonuniformities in the mirror-configuration magnetic field (dashed curve). The variations in the magnetic field $\Delta B(z)$ reached a maximum ΔB_m below the center of the coils. The nonuniformities had a half-width of ~ 0.1 m and the measurements were made for fixed $B_0 = 1$ T.

The characteristics of the space charge oscillations were determined using probes *11* and *12* positioned 0.4 and 0.8 m from the cathode, respectively, which almost exclusively recorded the induced signals, according to an analysis made in Refs. 5–7. In special cold measurements the sensitivity of the probes to space charge oscillations in the beam was set approximately the same (to within ~ 20–30%) by altering the insertion depth of the probes relative to the apertures in the transport channel. The detected probe signals were fixed so that time variations of the oscillation amplitude could be identified.

The transverse structure of the relativistic electron beam was determined by observing the x-ray emission from the end collector. The x-ray emission in the collector plane was determined using a special x-ray image converter similar to that described in Ref. 9. The radiation from the collector was fed into the image converter via a pinhole camera with a ~ 1 mm diameter aperture. Supplying the image converter with ~ 100 ns pulses yielded information on the structure of the relativistic electron beam at different times, by changing the delay of the supply pulse relative to the beam current pulse. By using a dc supply voltage, it was also possible to determine the beam structure averaged over the pulse time τ_b. A

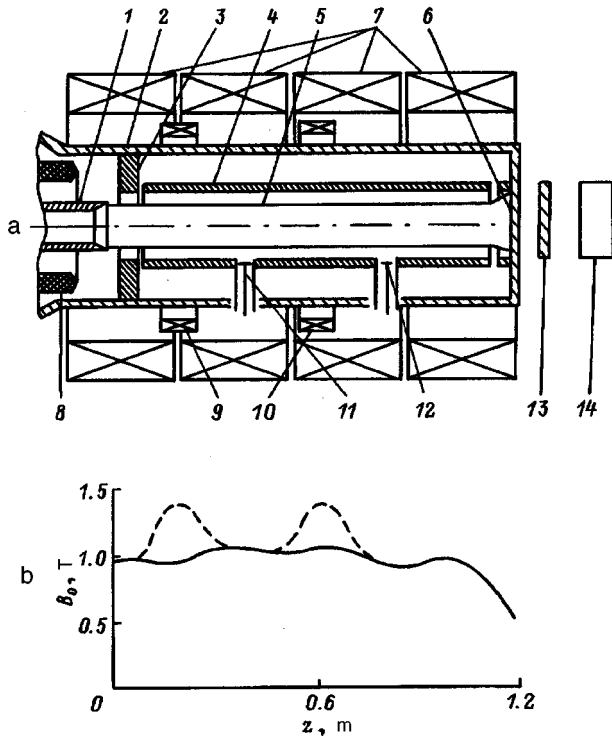The creation of local nonuniformities of the mirror-configuration magnetic field led to appreciable changes in the oscillation intensity. Figure 2 shows typical time behavior of the collector current I_{col} (Fig. 2a) and the amplitude A of the detected signals from probes 11 and 12 in the absence of local magnetic field nonuniformities (Fig. 2b), as well as the behavior when local isolated magnetic field nonuniformities, with an amplitude $\Delta B_m/B_0$ of around 30%, were created using coils 9 (Fig. 2c) and 10 (Fig. 2d).

If there are no magnetic field nonuniformities, the probe signals increase monotonically with time as far as the pulse peak of the collector current I_{col} (~ 1.2 μs) (Fig. 2b). In this case, the amplitude at probe 12 some distance from the cathode is considerably greater than that at probe 11 as a result of an increase of the space charge waves along the beam.[6,7] Decay of the current I_{col} after the pulse peak only leads to a decrease in the signal at the distant probe. The signal from probe 11 continues to increase, reaching a maximum 1.7–1.8 μs after the beginning of the pulse. This increase is clearly caused by an influx of cathode plasma to the probe 11. With this configuration of probes, the average plasma velocity should be $\sim 2 \times 10^7$ cm/s, which agrees with the measurements made by other authors.[3,10] Like probe 11, the probe 12 some distance from the cathode only records the oscillations from the near-field zone of the relativistic electron beam and does not ''feel'' the influence of the plasma over the time $t < 2.5$–3 μs.

The introduction of a magnetic field nonuniformity using coil 10 increases the growth rate of the oscillations of both probes. The largest signals from probe 12 (the region $0.5 \le t \le 1.6$ μs in Fig. 2d) are close to the maximum in the absence of local magnetic field nonuniformities (Fig. 2b).

The faster (compared with the case $\Delta B = 0$) oscillations of probe 11 in the time interval $t \le 1$ μs (Fig. 2d) can be explained if the magnetic mirror near coil 10 can cause re-
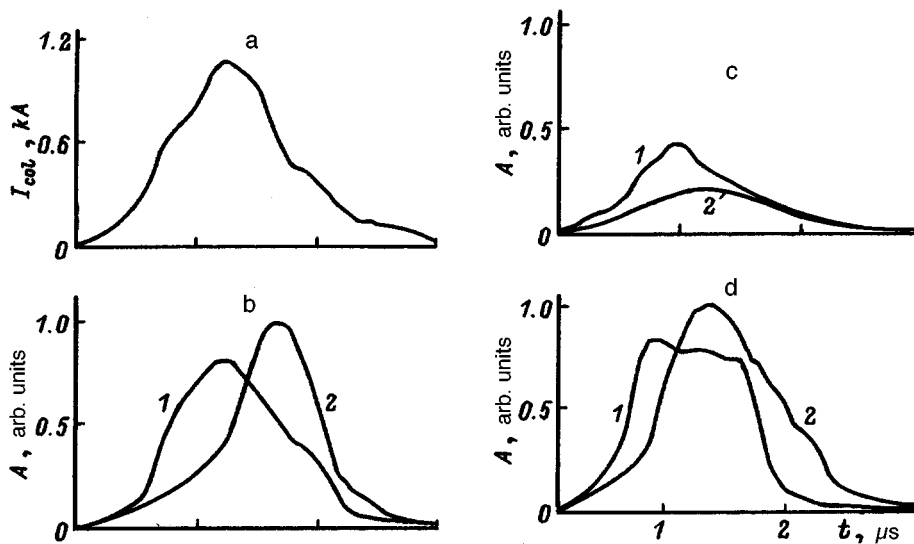

FIG. 1. a — Schematic of cross section through beam shaping and transport system: 1 — cathode, 2 — vacuum chamber, 3 — anode, 4 — transport channel, 5 — electron beam, 6 — collector, 7 — solenoids to generate magnetic field B_0, 8 — protective electrode, 9, 10 — auxiliary coils, 11, 12 — rf probes, 13 — pinhole camera, 14 — x-ray image converter; b — Magnetic induction versus z: solid curve — with no local nonuniformities, dashed curve — local magnetic field nonuniformities generated by coils 9 and 10. The end of the cathode (1) and the surface of the end collector (6) are positioned at $z = 0$ and $z = 1.2$ m, respectively.

spatial resolution of ~ 2–3 mm was achieved in the image converter measurements.

The apparatus was evacuated continuously during the experiments. The pressure did not exceed 10^{-5} Torr.

FIG. 2. Time behavior of the collector current and amplitude of the detected signals from probes 12 (1) and 11 (2). The values of A are given in arbitrary units which are the same for all the graphs.
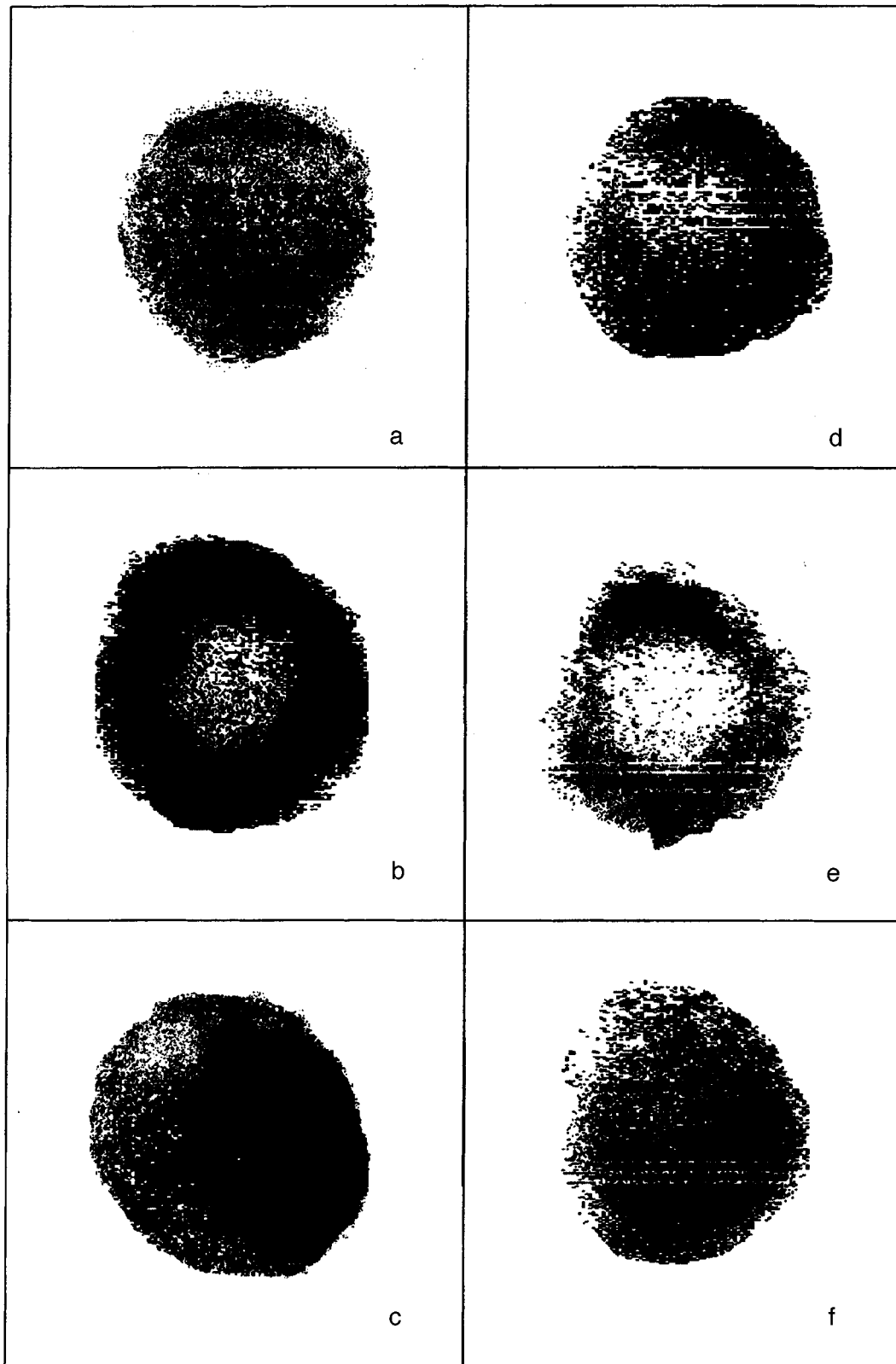
FIG. 3. Typical x-ray images of beam track on end collector: a, d — for a beam propagating in a uniform magnetic field, b, e — with a magnetic field nonuniformity created by coil *9*; c, f — with a magnetic field nonuniformity created by coil *10*; a–c — time-averaged images, d–f — instantaneous images at the center of the leading edge of the beam current pulse.

flection of beam electrons. Reflection leads to the development and rapid increase in the amplitude of the oscillations in the single trap between the cathode and the magnetic mirror, similar to that observed in helical electron beams in gy-rotron systems.[6,11] As the magnetic field B varies adiabatically, the mirrors reflect those electrons for which the transverse $(V_{\perp 0})$ and longitudinal $(V_{\parallel 0})$ components of the velocity ahead of the mirror satisfy
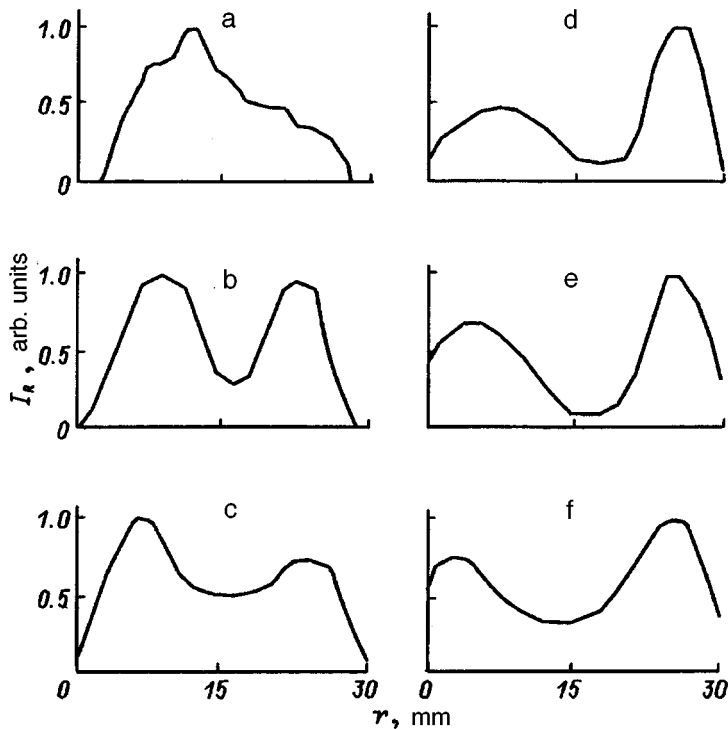
FIG. 4. Radial distributions of the luminescence intensity I_R at the image converter screen recorded in a uniform magnetic field (a–c), and with a magnetic field nonuniformity created by coil 9 (d–f): a, d — time-averaged distributions; b, e — instantaneous distributions at the center of the leading edge of the beam current pulse; c, f — instantaneous distributions at the peak of the beam current pulse. The values of I_R are in arbitrary units. The characteristics $I_R(r)$ are normalized so that the maximum of I_R on each graph is 1. The radial coordinate r is taken from the edge of the image.

$$\frac{V_{\perp 0}}{V_{\|0}} \geqslant (B_0 / \Delta B_\mathrm{m})^{1/2}.$$

Thus, the relativistic electron beam should contain electrons having fairly high transverse velocities $V_{\perp 0} \geqslant 1.7 \cdot V_{\|0}$ near the coil 10, since reflection is observed for $B_0 / \Delta B_\mathrm{m} \simeq 3$. Electrons with this velocity ratio may occur mainly as a result of emission of cathode plasma from the side surface in the transport channel, where there is a strong electric field transverse to the magnetic field. An increase in the ratio $V_{\perp 0} / V_{\|0}$ may also be promoted by the action of alternating space charge fields.

The accelerated increase in the amplitude of the signal from the probe 12 when the coil 10 is switched on, may be caused by the oscillations in the trap influencing the evolution of the space charge waves, or it may be the result of additional electron bunching in the beam as it passes through the magnetic field nonuniformity. This enhanced bunching effect is observed (see Refs. 12 and 13, for example) in the presence of space charge oscillations in sections of decreasing electron velocity (drift velocity jump).

The nonuniform field generated by the coil 9 has a different effect (Fig. 2c). An abrupt decrease in the amplitude of the signal from the probe 11 in the time interval $1.0 < t < 2.5 \ \mu$s is clearly caused by reflection of the cathode plasma flux from the magnetic mirror. The decrease in the signals from probe 11 for $t < 1 \ \mu$s and from probe 12 over the entire current pulse is clearly attributable to mechanisms which did not play an important part when coil 10 was switched on.

In order to understand why the same magnetic field nonuniformities in different sections of the system along the axis have a different influence on the collective processes in the relativistic electron beam, we note that these nonuniformities

are created at different distances from the cathode, in regions of the beam with alternating fields of widely differing amplitude. An increase in the space charge waves with increasing distance from the cathode can appreciably alter the electron transverse velocities and the current distribution over the beam cross section. These changes may be responsible for the differences in the action of the nonuniform fields generated by coils 10 and 9. The drift velocity jump may also depend on the amplitude of the alternating fields in the beam.

This explanation taking into account the space charge oscillations is also supported by studies of the beam structure near the collector. These studies indicate that the recorded oscillations influence the electron confinement and current distribution over the beam cross section. The results are illustrated in Figs. 3–5, which show typical x-ray images of the beam track on the end collector (Fig. 3) and radial distributions of the luminescence intensity $I_R(r)$ on the screen of the image converter (Figs. 4 and 5). Assuming that the magnetic field at the end collector (B_col) is reduced compared with that at the cathode (B_c) and assuming a corresponding expansion of the beam, the linear scale on the abscissa in Figs 4 and 5 is reduced $(B_c / B_\mathrm{col})^{1/2}$ times. This makes it more convenient to determine the changes in the beam shape at the collector compared with the cathode.

In the absence of any local magnetic field nonuniformities, the intensity drop over the beam cross section is comparatively small, not exceeding a factor of ~ 3 (Figs. 3a, 3d, and 4a, 4c). The instantaneous distributions $I_R(r)$ vary substantially during the pulses (Figs. 4b and 4c) and no annular beam structure can be identified on the time-averaged distribution (Figs. 3a and 4a). A comparison between the instantaneous distributions at different times indicates that the beam diameter increases at a rate of 2–4×10^5 cm/s at the
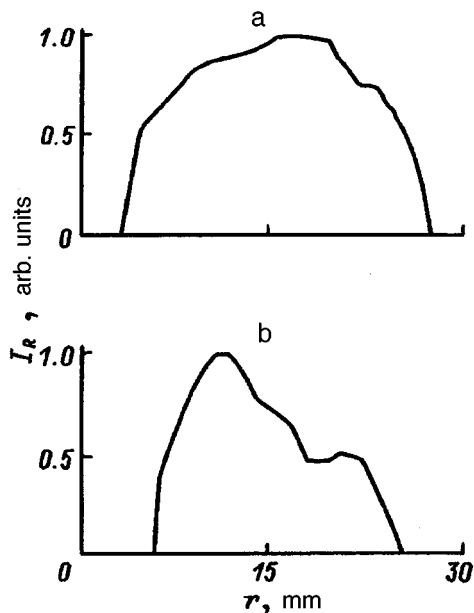
FIG. 5. Radial distributions of the luminescence intensity I_R at the image converter screen, obtained with a magnetic field nonuniformity created by coil *10*: a — time-averaged distribution and b — instantaneous distribution at the center of the leading edge of the beam current pulse.

MAIN RESULTS AND CONCLUSIONS

The space–time characteristics of a long-pulse relativistic electron beam have been determined and the influence of mirror-configuration magnetic field nonuniformities on these characteristics has been established. The mechanism for the action of these nonuniformities on the oscillations and structure of the beam has been explained assuming that the plasma and electron flux are reflected by the magnetic mirror and also allowing for a drift velocity jump. The possibility of suppressing space charge oscillations and improving beam confinement by using a magnetic field nonuniformity near the cathode has been identified.

pulse leading edge, which agrees with the existing published data[3,10] and is caused by expansion of the cathode plasma. The ''blurredness'' of the beam under these conditions is probably caused by the action of the alternating fields and the beam oscillations.

The introduction of a magnetic field nonuniformity using coil *9* enhances the contrast of the image and stabilizes it (Figs. 3b and 3e; Figs. 4b, 4e, and 4f). As a result, the annular structure of the beam can be identified clearly on the time-averaged distribution. The drop in the intensity I_R over the beam cross section reaches a factor of ~ 10. The change in the beam distribution over the cross section is clearly caused by the suppression of the space charge oscillations in the beam by the mirror-configuration magnetic field nonuniformity near the cathode. The establishment of this nonuniformity improves the confinement and transport of the beam.

The introduction of a magnetic field nonuniformity using the coil *10* some distance from the cathode causes a deterioration in the beam confinement, even compared with the case where both coils are switched off. This is observed as an increase in the luminescence intensity I_R of the central part of the cross section (compare Figs. 3d and 3f, and also Figs. 4b and 5b) and is caused by the increased oscillations (Fig. 2d).

[1] *Relativistic rf Electronics*, edited by A. V. Gaponov-Grekhov [in Russian], Institute of Applied Physics, Gorki (1979).

[2] *Relativistic rf Electronics*, edited by A. V. Gaponov-Grekhov [in Russian], Institute of Applied Physics, Gorki (1983).

[3] S. P. Bugaev, V. I. Kanavets, V. I. Koshelev, and V. A. Cherepenin, *Relativistic Multiwave Microwave Oscillators* [in Russian], Nauka, Novosibirsk (1991).

[4] L. Yu. Bogdanov, S. V. Voskresenskiĭ, and G. G. Sominskiĭ, *Problems of Physical Electronics* [in Russian], Leningrad Nuclear Physics Institute, Leningrad (1991), pp. 102–119.

[5] A. V. Arkhunov, L. Yu. Bogdanov, S. V. Voskresenskiĭ *et al.*, Izv. Vyssh. Uchebn. Zaved. Prikl. Nelin. Dinam. **3**, No. 4 (1995).

[6] A. V. Arkhunov, L. Yu. Bogdanov, S. V. Voskresenskiĭ *et al.*, Izv. Vyssh. Uchebn. Zaved. Prikl. Nelin. Dinam. **3**, No. 5 (1995).

[7] L. Yu. Bogdanov and G. G. Sominskiĭ, Zh. Tekh. Fiz. **65**(12), 77 (1995) [Tech. Phys. **40**, 1245 (1995)].

[8] A. V. Arkhipov, L. Yu. Bogdanov, S. V. Voskresenskiĭ *et al.*, in *Problems of Physical Electronics* [in Russian], Leningrad Nuclear Physics Institute, Leningrad (1989), pp. 28–47.

[9] Yu. L. Bakshaev, A. B. Basmanov, P. I. Blinov *et al.*, Fiz. Plazmy **15**, 992 (1989) [Sov. J. Plasma Phys. **15**, 575 (1989)].

[10] V. V. Gorev, G. I. Dolgachev, L. P. Zakatov *et al.*, Fiz. Plazmy **11**, 782 (1985) [Sov. J. Plasma Phys. **11**, 456 (1985)].

[11] O. I. Luksha and O. Yu. Tsybin, Lectures on Microwave Electronics and Radiophysics (Ninth Winter School/Seminar for Engineers), Saratov [in Russian], Saratov State University Press, Saratov (1993).

[12] G. G. Sominskiĭ and O. Yu. Tsybin, *Abstracts of Papers presented at Ninth All-Union Conference on Microwave Electronics*, Saratov, Vol. 1, [in Russian], Saratov State University Press, Saratov (1979), p. 115.

[13] É. A. Gel'vich and G. G. Sominskiĭ, Elektron. Promysh. No. 7–8, 20 (1981).

Translated by R. M. Durham

Two-period cylindrical energy analyzer with end electrodes

L. P. Ovsyannikova and T. Ya. Fishkova

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

Analytical expressions are obtained for the main parameters of a cylindrical analyzer with flat end electrodes held at the potential of the inner cylinder. Calculations are carried out for its one- and two-period operating regimes with first-order focusing with respect to angle for an object lying outside the analyzer. It is found that the greater the distance from the object to the front end of the analyzer, the smaller its focal power. The exit beam intensity is calculated for the situation in which the surface of the sample is scanned by the primary beam, and also for the case of an object of finite dimensions. Empirical formulas are found for the falloff of the intensity with distance from the center of the object. A comparison of numerical calculations using a program for two-dimensional charged-particle optics and calculations from our analytical formulas shows that their difference does not exceed 10%. © *1997 American Institute of Physics.* [S1063-7842(97)01508-0]

In Ref. 1 we proposed a cylindrical mirror analyzer with flat end electrodes and found its optimal operating regime, which provides energy dispersion and focal power as good as those of the classical cylindrical mirror. Here the out-distance of the object (its distance from the front end of the analyzer) is equal to one-half the radius of the inner cylinder, which is insufficient for the solution of a number of physical problems.

In the present paper we investigate the parameters of an analyzer for which the out-distance of the object is significantly greater. In this case the course of a trajectory is not mirror-symmetric relative to its turning point. For this reason, in what follows we call such a device a cylindrical analyzer with end electrodes (CAE).

We have found (in analogy with a cylindrical deflector with side electrodes[2]) an analytical expression for the potential distribution of a CAE for which the distance between the end electrodes is at least four times the distance between the cylindrical electrodes. It has the form (we place $z=0$ at the position of the front end)

$$\phi(r,z) = V_1 + (2/\pi)(V_2 - V_1)\sqrt{r_2/r}\left[\arctan\frac{\sin(\pi z/l)}{\sinh \pi(r_2-r)/l} - \frac{2\sin(\pi z/l)\sinh \pi(r_2-r)/l}{\sinh 2\pi(r_2-r_1)/l}\right], \quad (1)$$

where V_1, V_2 and r_1, r_2 are the potentials and radii of the inner and outer cylindrical electrodes, respectively, and l is the distance between the end electrodes at potential V_1.

It is useful to divide the potential distribution inside the CAE into two regions: a near-end region and a region in which the potential varies logarithmically. If the length of the near-end region is of the order of half the distance between the cylindrical electrodes, then the potential distribution in it is found by expanding expression (1) in a series. Solving the equation of motion to first order, we find that in the near-end region the charged particles move equally slowly. In this case, the entrance to the analyzer can be both through the end and through the inner electrode. At the entrance to the field region with logarithmically varying potential the slope angle of the trajectory remains equal to its value at the entrance to the CAE, but the energy of the particles decreases to a value corresponding to the equipotential at the entrance to the region.

At the exit of the charged-particle beam through the inner electrode, first-order focusing with respect to angle in the axis-to-axis regime is possible if the following relation between the focal power of the analyzer and the focusing angle (the initial entrance angle of the central trajectory) is fulfilled:

$$\sqrt{\pi}(r_j/r_1)\exp(g^2)[\mathrm{erf}(g)+\mathrm{erf}[\sqrt{g^2+\ln(r_j/r_1)}]]$$
$$= P - 2(r_j/r_1)g/(1+2g^2),$$

where
$$g = \sin\theta_0[\ln(r_2/r_1)/(-F)-\ln(r_j/r_1)]^{1/2},$$
$$r_j = (l_0+\Delta l)\tan\theta_0,$$

$$P = \frac{2g\tan\theta_0}{\tan\theta_i(\tan^2\theta_0 - 2g^2)}\left[1 + \frac{\ln(r_2/r_1)\tan^4\theta_0}{2Fg^4\tan^2\theta_i}\right],$$

$$\tan\theta_i = \tan\theta_0\sqrt{g^2+\ln(r_j/r_1)}/g,$$

$$F = -e(V_2-V_1)/\varepsilon_0. \quad (2)$$

Here l_0 is the distance from the front end of the CAE to the object, Δl is the length of the near-end region, θ_0 is the entrance angle (to the CAE) of the central trajectory, r_j is the coordinate of the trajectory at which it exits the near-end region, and θ_i is the exit angle of this trajectory from the CAE. F and ε_0 denote the focal power of the analyzer and its tuning energy, and e is the charge of the particle. The base of the analyzer and its energy dispersion in the longitudinal direction have the form

$$L/r_1 = \cot\theta_i + [Pg+(r_j/r_1)/(1+2g^2)]\cot\theta_0,$$

$$D_z/r_1 = \sin^2\theta_0\ln(r_2/r_1)g^{-2}(-F)^{-1}[(1+2g^2)L/r_1 - (r_j/r_1)\cot\theta_0 - \tan^2\theta_0\cot^3\theta_i]. \quad (3)$$

TABLE I.

l_0/r_1	$-F$	l_k/r_1	d_k/r_1	θ_0^0	α^0	D_z/r_1	D_r/r_1	I_0
0.5	0.662	0.54	0.06	42.5	9	7.4	6.7	1.0
1.0	0.655	0.21	0.06	39	6	6.9	6.6	0.52
1.5	0.643	0.08	0.06	35	4.5	6.5	6.5	0.31
2.0	0.623	0.07	0.06	32	4	6.3	6.5	0.23
2.5	0.602	0.06	0.06	29	4	6.2	6.3	0.19

FIG. 1. Two-period cylindrical energy analyzer with flat end electrodes and path of the charged-particle trajectories for an outlying object.

If the particles enter and depart through the inner electrode at a distance from the ends greater than half the distance between the electrodes, then the base of the analyzer and the longitudinal dispersion have the form

$$L/r_1 = -2F \cot^3\theta_0 [2 \ln(r_2/r_1)\cos^2\theta_0 + F]^{-1},$$

$$D_z = L/(2 \cos^2\theta_0). \qquad (4)$$

In formula (4) the relation between the focusing angle and the focal power of the analyzer is the same as for a classical cylindrical mirror.[3]

We have calculated the main parameters as functions of the out-distance of the object for a CAE with geometrical dimensions $r_2/r_1 = 2.3$ and $l = 5.2r_1$. The results of a numerical calculation of the CAE in the single-period operating regime (with single intersection of the system axis) are listed in Table I.

In Table I, l_k is the distance from the rear end of the analyzer to the beam crossover region, located on the system axis; d_k is the diameter of the crossover region; α is the beam capture angle; D_r is the energy dispersion coefficient at crossover under the condition of uniform initial intensity distribution in angle (the maximum intensity corresponding to the optimal operating regime of the analyzer at $l = 0.5r_1$ is taken to be equal to unity). The energy resolution of the CAE in the indicated regimes is 0.8–0.9%. Results calculated using formulas (2)–(4) coincide with the numerically calculated results listed in Table I to within 5–10%.

In Ref. 4 we determined the parameters of a single-period cylindrical mirror analyzer with end electrodes in the regime of surface scanning of an object with a thin primary beam, and also for an object of finite dimensions in the optimal operating regime. Here the exit beam intensity falls quite rapidly although not as rapidly as for a classical mirror. To increase the energy dispersion and also the beam intensity, including the scanning case, it is necessary to use a two-period operating regime (with two intersections of the system axis; Fig. 1). In what follows we will call such an energy analyzer a two-period cylindrical analyzer with end electrodes (TCAE). We kept the transverse geometrical dimensions of the TCAE the same as for the single-period analyzer. The longitudinal dimension was found from the condition of focusing a beam exiting from a point object at a distance from the front end of the system varying within wide limits $0.5 \leqslant l_0/r_1 \leqslant 2.5$. Numerical calculations show that the distance between the flat end electrodes should be equal to $11r_1$. In this case, beam focusing is realized on the system axis at a crossover with diameter $d_k = 0.1r_1$, and its position remains unchanged and equal to $l_k = 0.6r_1$. The remaining parameters for the TCAE are listed in Table II.
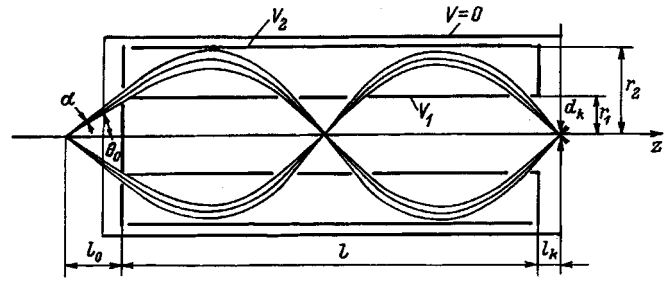
It can be seen from Table II that for a TCAE the energy dispersion is roughly two times higher than for a single-period analyzer, but the beam capture angle is somewhat greater for a small out-distance of the object ($l_0 \leqslant r_1$). For $l_0 \geqslant 1.5r_1$ the values of this angle for single-period and two-period analyzers are identical.

Figure 2 plots the exit beam intensity I scaled by the beam intensity I_0 at the tuning energy of the TCAE, as a function of the particle energy (the instrument function). It can be seen from the figure that for a position of the object $l_0 = 0.5r_1$ the energy resolution at half-maximum $\Delta\varepsilon/\epsilon_0$ is equal to 0.75%, and for $l_0 = 2.5r_1$ $\Delta\varepsilon/\epsilon_0 = 0.52\%$. Calculations of the instrument function at intermediate positions of the object showed that as l_0 increases between these two limits, the energy resolution of the TCAE remains within the limits 0.8–0.5%. Note that all the calculations, including those of the instrument function, assume the presence of a ring diaphragm at the entrance to the analyzer. Its dimensions correspond to the transmission angles of the analyzer in the various operating regimes (Tables I and II). To avoid field distortions, all openings in the electrodes should be covered with netting.

When scanning the surface of a sample with a thin primary beam in the direction transverse to the system axis, the intensity of the beam falling on the exit diaphragm falls considerably more slowly than for a single-period analyzer. Figure 3 plots the dependence of this intensity, scaled by the maximum intensity in the optimal operating regime, on the magnitude of the displacement of the primary beam from the center of the object. The fall in intensity when scanning can be approximated by a parabola $I = I_0[1 - (s/h)^2]$ for $l_0 \leqslant 0.5r_1$ and by a straight line $I = I_0(1 - s/h)$ for $l_0/r_1 \geqslant 1$. Here h is the distance of the primary beam from the center of the object at which $I = 0$. Its magnitude can be found from the empirical formula $h/r_1 = 0.13 - 0.03/(l_0/r_1)$. Note that the intensity of the beam from the center of the object I_0

TABLE II.

l_0/r_1	$-F$	θ_0^0	α^0	D_z/r_1	D_r/r_1	I_0
0.5	0.662	41	11.3	17	15	1.0
1.0	0.647	39	11	17	15	0.79
1.5	0.633	35	4.5	18	16	0.31
2.0	0.625	32	4	19	18	0.23
2.5	0.613	29	4	19.5	18.5	0.19

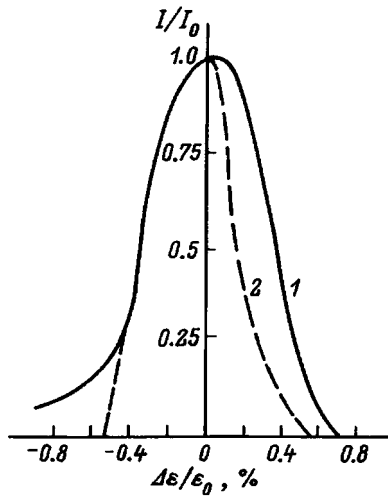FIG. 2. Instrument function of the TCAE for different positions of the object: $l_0 = 0.5r_1$ (*1*), $2.5r_1$ (*2*).
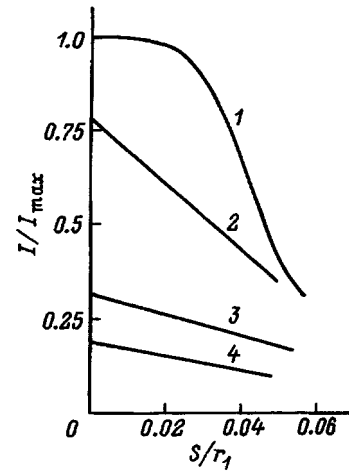

FIG. 3. Focal power of the TCAE when scanning an object in the transverse direction for different values of its location outside the analyzer; $l_0 = 0.5r_1$(*1*), $1.0r_1$ (*2*), $1.5r_1$ (*3*), $2.5r_1$ (*4*).

(Table II) falls noticeably as the object is distanced from the TCAE and, although the falloff of the intensity here takes place more slowly (Fig. 3), the focal power of the TCAE is substantially decreased.

The results obtained when scanning the object can also be used in the case of an object of finite dimensions. Under the condition of a uniform distribution of the initial beam intensity on an object of radius r_0, the intensity at crossover is equal to

$$I_k = I_0 \pi r_0^2 [1 - 0.5(r_0/h)^2] \quad \text{for} \quad l_0/r_1 \leqslant q 0.5, r_0^2$$
$$\times (1 - 2r_0/3h) \quad \text{for} \quad 1 \leqslant q l_0 / r_1 \leqslant q 2.5. \quad (5)$$

To summarize, analytical expressions have been obtained for the main parameters of a cylindrical energy ana-

lyzer with end electrodes under the condition of first-order focusing with respect to angle. We have found the operating regimes for positions of the outlying object varying within wide limits. We have calculated the instrument function of the analyzer, the energy resolution, and the focal power, including the case of an object with finite dimensions.

[1] L. P. Ovsyannikova and T. Ya. Fishkova, Zh. Tekh. Fiz. **64**, 174 (1994) [Tech. Phys. **39**, 1072 (1994)].

[2] L. P. Ovsyannikova and T. Ya. Fishkova, Nucl. Instrum. Methods A **363**, 494 (1995).

[3] V. P. Afanas'ev and S. Ya. Yavor, *Electrostatic Energy Analyzers for Charged Particle Beams* [in Russian], Nauka, Moscow, (1978), p. 108.

[4] L. P. Ovsyannikova, T. Ya. Fishkova, and E. V. Shpak, Pis'ma Zh. Tekh. Fiz. **21**(21), 19 (1995) [Tech. Phys. Lett. **21**, 870 (1995)].

Translated by Paul F. Schippnick

Study of aberration correction in two types of electrostatic lenses

L. A. Baranova and S. Ya. Yavor

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

A detailed comparative study is carried out of chromatic and spherical aberration in a crossed lens and in a lens formed by four cylindrical electrodes, the two inner ones of which are cut into four symmetrical longitudinal sections. The study was performed by numerical modeling. The possibility of simultaneous correction of both types of aberration in a linear image is demonstrated. A comparison is made with an equivalent axisymmetric lens. © *1997 American Institute of Physics.* [S1063-7842(97)01608-5]

The main aberrations governing the resolving power of most electron-optical devices are the spherical and the chromatic. They cause distortions of the image of point sources located both off and on the optical axis. To correct for spherical aberration, one usually uses multi-electrode lenses—octupoles and sextupoles.[1] It has also been shown that under certain conditions a single three-electrode crossed lens creates a linear image with zero or negative spherical aberration.[2] Correction of chromatic aberration has so far been realized mainly with the help of superimposed electric and magnetic quadrupoles.[3] However, the use of a magnetic field has a number of disadvantages significantly complicating practical work with the system. In Ref. 4 it was shown that chromatic aberration of a linear image can also be corrected in a purely electrostatic lens formed by a set of sequentially placed cylinders, some of which are cut into four symmetrical longitudinal sections (Fig. 1). The potentials are applied in such a way as to create a superposition of an axisymmetric lens and a doublet of quadrupole lenses. In Ref. 5 it was shown that in addition to chromatic aberration, spherical aberration of a linear image can also be corrected in such a lens. Reference 6 investigated correction of chromatic and spherical aberration of a five-electrode crossed lens. The lens is formed by flat electrodes with rectangular openings, turned in neighboring electrodes by 90° relative to the next one (Fig. 2). The potentials are applied to the electrodes in

such a way that the system is analogous in its optical properties to a doublet of quadrupole lenses with a superimposed axisymmetric lens. Thus, in purely electrostatic lenses formed by a superposition of axisymmetric and quadrupole field components, we have unveiled the possibility of correction of chromatic and spherical aberrations (apart from their dependence on the shape of the electrodes creating the field). At the same time it is well known that sequentially arranged axisymmetric and quadrupole lenses do not possess this property.[1]

In the present paper we present a detailed comparative study of chromatic and spherical aberrations in the two types of electrostatic lenses described above (Figs. 1 and 2). This work was carried out by numerical modeling using a three-dimensional program for calculating the field and the trajectories.[7] The geometrical parameters of the investigated lenses were as follows. In the first lens we took the diameter D of the cylinders as our unit of length, and in the second we took the length a of the short side of the rectangular opening as our unit of length. The lengths of the two inner cylinders of the first lens were taken to be equal to half the diameter (0.5). The spacing between the cylinders was taken to be equal to 0.02, and the width of the longitudinal cuts was
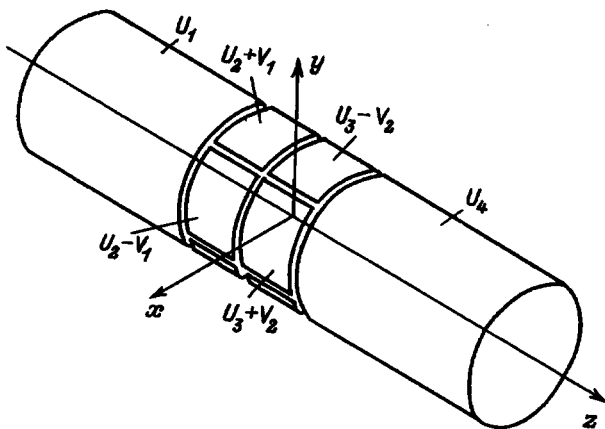

FIG. 1. Lens formed by four cylindrical electrodes, where the inner electrodes are cut into four equal sections.
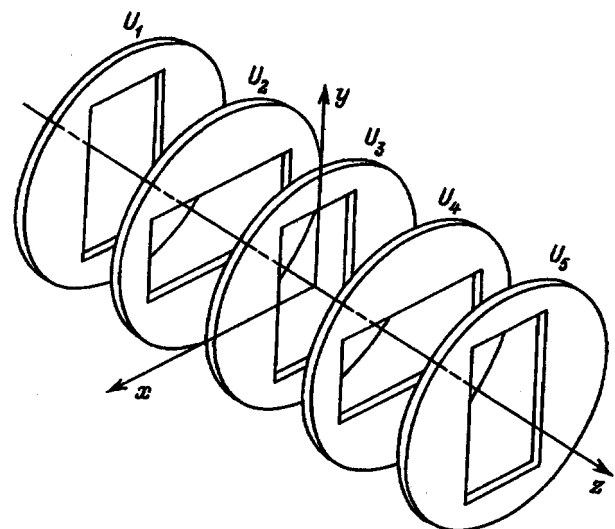

FIG. 2. Five-electrode crossed lens.

taken to be equal to 0.028. In the calculations the center of the coordinate system was located at the geometrical center of the system, and the two outer cylinders were closed off by flat disks at $z = \pm 6$. The distances between neighboring electrodes of the second lens were taken to be equal to $s = 0.25$, and the ratio of the sides of the rectangular opening $b/a = 2$. The center of the coordinate system, as before, was located at the geometrical center of the lens. The system was considered to be enclosed within a rectangular box with corner coordinates $x = \pm 1.25$, $y = \pm 1.25$, $z = \pm 6$. The thickness of the electrodes was assumed to be negligibly small. In object and image space the potentials on the bounding box were constant and equal to the potentials of the electrodes in contact with the lens. Between the electrodes the potentials on the box varied linearly.

The potentials applied to the lenses are shown in Figs. 1 and 2. The potentials U_1, U_2, U_3, U_4 forming the axisymmetric component of the field are applied to the four cylindrical electrodes in sequence as shown, and the potentials $\pm V_1$ and $\pm V_2$, applied to the segments of the two inner cylinders, form a quadrupole doublet, the first lens of which disperses charged particles in the xz plane and the second lens collects them. The potentials $U_1 - U_5$ are applied to the electrodes of the crossed lens. The first potential difference $U_2 - U_1$ retards the charged particles, the following two potential differences $U_3 - U_2$ and $U_4 - U_3$ accelerate them and the last difference $U_5 - U_4$ again retards them. For the arrangement of electrodes shown in Fig. 2, such a potential feed leads to the result that the first two potential differences $U_2 - U_1$ and $U_3 - U_2$ disperse the charged particles in the xz plane and the last two potential differences collect them. Thus, in both lenses, the xz plane is the DC plane of the quadrupole doublet (dispersing–collecting), and the yz plane is the CD plane (collecting–dispersing).

A detailed study of chromatic aberration of both types of lenses shows that its correction is achieved comparatively easily with astigmatic focusing. An entire series of regimes was obtained, characterized by zero and negative chromatic aberration of one of the linear images. We present results for a parallel beam at the lens entrance. Figure 3 plots the dependence of the position $z(F_{ix})$ of the linear image in the xz plane for four regimes on the charged particle energy E, where E is given in units of eU_1. Curves 1–3 correspond to the lens formed by cylindrical electrodes. In all cases $U_1 = 1$ and $U_4 = 4.0$, for curve 1 $U_2 = 0.4$, $U_3 = 5.0$, $V_1 = 1.068$, $V_2 = 1.40$, for curve 2 $U_2 = 0.5$, $U_3 = 4.0$, $V_1 = 0.933$, $V_2 = 1.20$, and for curve 3 $U_2 = 1$, $U_3 = 4.0$, $V_1 = 0.971$, $V_2 = 1.20$ (here and below all potentials are given in units of U_1). Curve 4 corresponds to the crossed lens with potentials $U_1 = 1.0$, $U_2 = -2.8$, $U_3 = 3.8$, $U_3 = 10.0$, $U_5 = 4.0$.

As can be seen from Fig. 3, the dependence of $z(F_{ix})$ on the particle energy has the same character for both lenses. Note that for the same acceleration and the same focal lengths, the potentials on the intermediate electrodes of the crossed lens are substantially higher in absolute value than on the electrodes of the cylindrical lens, i.e., the former is weaker than the latter. For the particle energy less than 0.6, the position of the image approaches the lens as the energy is increased, i.e., the chromatic aberration in this region is
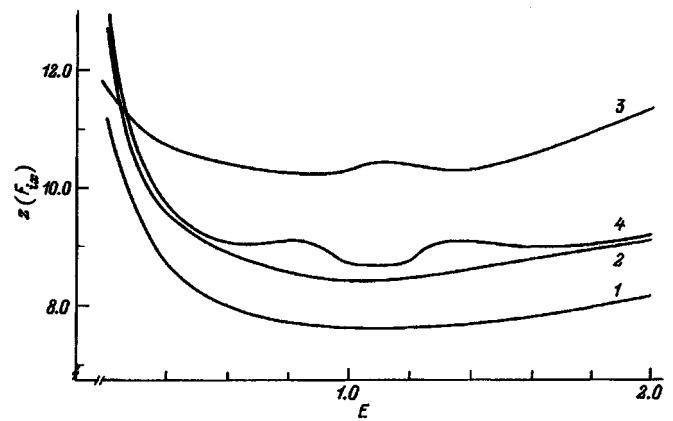

FIG. 3. Position of linear focus in the xz plane plotted versus particle energy.

negative. For large E ($E > 2$) the image moves away from the lens as the energy is increased, which corresponds to positive chromatic aberration. In the intermediate region $0.6 < E < 2$ chromatic aberration is small and at some points vanishes. Curves 3 and 4 have a wavy shape in this region, where the amplitude of oscillation exceeds the error of calculation.

We were able to obtain negative and zero chromatic aberration of a linear image only in accelerating lenses in the plane where the first quadrupole lens disperses. This image is located much farther from the lens than the second image perpendicular to it and therefore has significant length. We give by way of example the data for curve 4 (the crossed lens) at the point $E = 1.0$. For this energy the position of the focus in the xz plane is equal to $z(F_{iy}) = 0.213$, and the corresponding focal lengths are equal to $f_{ix} = 4.52$ and $f_{iy} = 0.723$. The corresponding linear focal length l scaled by x_0, where x_0 is the distance of a trajectory to the axis in object space, is equal to $l/x_0 = 11.7$.

It is of interest to consider chromatic aberration of shorter linear images, which corresponds to decreasing the astigmatism of the lenses. The latter may be achieved, for example, by increasing the potential U_2 in the case described above. If we increase U_2 from -2.8 to -1.8, then the focal lengths become equal to $f_{ix} = 2.65$ and $f_{iy} = 1.08$, and the positions of the foci take on the values $z(F_{ix}) = 3.85$ and $z(F_{iy}) = 0.390$. In this case the length of the linear image decreases to $l/x_0 = 3.20$. The coefficient of chromatic aberration becomes positive, but its magnitude remains small: $C_{cx} = 0.29$ for $E = 1.0$.

Further decrease of astigmatism can also be achieved by varying the potential U_5. If the potentials applied to the five electrodes of the crossed lens are equal to 1.0, -1.0, 3.8, 10.0, and 3.0, then the first-order parameters are $f_{ix} = 1.56$, $f_{iy} = 1.75$, $z(F_{ix}) = 2.13$, and $z(F_{iy}) = 0.730$. It is interesting to note that although the focal length in the yz plane is now greater than in the xz plane, the focal line F_{iy} as before is located closer to the lens than is the focal line F_{ix}. This is because the principal plane in the xz section is noticeably shifted toward image space while the principal plane in the yz section is shifted toward object space. In the given ex-

ample the length of the image is significantly decreased and is now equal to $l/x_0 = 0.8$. The chromatic aberration coefficient C_{cx} is small as before and equal to $C_{cx} = 0.22$ for $E = 1.0$.

The above results may be compared with the analogous results for an axisymmetric lens formed by three flat electrodes with opening diameter equal to 1 and inter-electrode distance equal to 0.5. For similar focal power, its chromatic aberration coefficient is roughly an order of magnitude greater than the corresponding coefficient for astigmatic lenses.

We were unable to find negative or zero chromatic aberration in single or retarding lenses by numerical modeling with any choice of parameters. Spurred by this negative result we undertook an effort to analyze the situation on the basis of an approximate analytical calculation. We considered a system consisting of two thin quadrupole lenses in the spatial interval between which the energy varies discontinuously. We showed that in this approximation the coefficient of chromatic aberration in the DC plane can vanish or change sign for intermediate acceleration, and in the CD plane, for intermediate retardation. The regions of regimes with corrected aberration found in the approximate calculation were then investigated by modeling lenses with cylindrical electrodes. Good agreement of the results of the approximate calculation and those of the more accurate numerical calculation was observed for intermediate acceleration between the quadrupole lenses. For intermediate retardation no such agreement was observed, i.e., neither negative nor zero chromatic aberration was observed. This is apparently explained by the excessive positive chromatic aberration inherent to retarding axisymmetric lenses, which was not taken into account in the approximate calculation.

Since in some regimes the lenses under study possess unusual chromatic aberration for electrostatic fields, it is of interest to also examine other properties in these regimes, first of all spherical aberration. We calculated the spherical aberration in both types of lenses in regimes with zero and negative chromatic aberration. The results obtained for linear focus in the xz plane are plotted in Fig. 4. The regimes represented by the curves are the same as in Fig. 3, and the particle energy is $E = 1.0$. The variable plotted along the abscissa, $x_i\prime$, is the slope angle of the trajectory in image space, and the variable plotted along the ordinate, Δz, is the longitudinal spherical aberration, i.e., the difference in the coordinates of intersection with the axis of the paraxial and non-paraxial trajectories. The difference Δz is assumed to be positive for positive spherical aberration.

It can be seen from Fig. 4 (notation the same as in Fig. 3) that spherical aberration is negative for all regimes. An analysis of the region of small $x_i\prime$ values shows that the third-order aberration coefficients C_{3x} are very small, and for the crossed lens and for regime 3 of the cylindrical lens $C_{3x} = 0$. With growth of $x_i\prime$ spherical aberration of the cylindrical lenses begins to grow abruptly in absolute value while for the crossed lens it remains small. Thus, it follows from a consideration of Figs. 3 and 4 that in certain regimes the investigated lenses can form a linear image with corrected spherical and chromatic aberration.
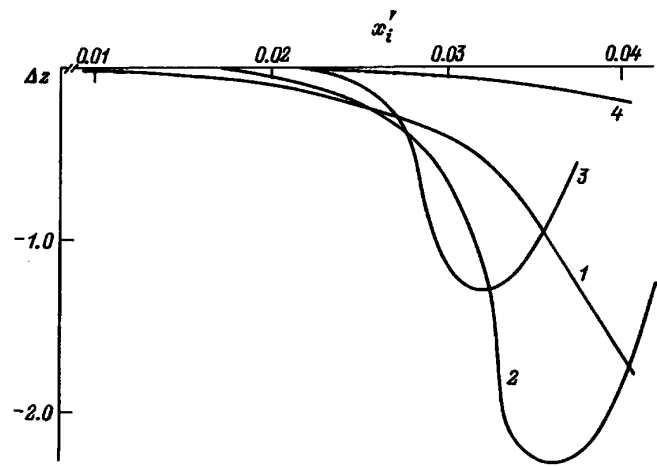

FIG. 4. Longitudinal spherical aberration of linear focus in the xz plane plotted versus slope angle of the trajectory in image space.

We also calculated spherical aberration of shorter linear foci, corresponding to decreased astigmatism of the lenses. The two regimes of this kind described above, are characterized, as was indicated, by small but positive chromatic aberration. The spherical aberration in these regimes is also small; however, the coefficient C_{3x} can be either positive or negative.

We also examined stigmatic regimes of lenses forming a point image of a point object. Of special interest here is the region of short focal lengths in which probe systems creating an image with large reduction operate. In axisymmetric lenses it is possible to achieve large reductions only with a retarding potential on the middle electrode, which leads to a high level of both spherical and chromatic aberration. The investigated lenses allow a significant reduction of the level of aberrations although in the stigmatic regimes we were not able to obtain negative or zero values.

By way of an example, we present results obtained for single lenses. In lenses with cylindrical electrodes we investigated in detail the regime characterized by the following electrode potentials: $U_1 = U_4 = 1.0$, $U_2 = 0.8$, $U_3 = 2.39$, $V_1 = 0.686$, $V_2 = 1.36$. Here the focal lengths in the xz and yz planes are equal, respectively, to $f_{ix} = 0.667$ and $f_{iy} = 3.53$ while the positions of the foci in these planes coincide: $z(F_{ix}) = z(F_{iy}) = 1.44$. The crossed lens with the same focal length in the xz plane ($f_{ix} = 0.667$) has the following electrode potentials: $U_1 = U_5 = 1.0$, $U_2 = -0.86$, $U_3 = 1.8$, $U_4 = 7.21$. Here the focal length in the yz plane is equal to $f_{iy} = 3.46$ while the positions of the foci are equal to $z(F_{ix}) = z(F_{iy}) = 1.45$. In both lenses the principal planes in the xz and yz planes are found at a greater distance from one another, with both shifted in opposite directions away from the center: e.g., for the crossed lens $z(H_{ix}) = 0.79$ and $z(H_{iy}) = -2.0$.

The chromatic aberration coefficients for the cylindrical lens are equal to $C_{cx} = 0.60$ and $C_{cy} = 8.4$, and for the crossed lens are equal to $C_{cx} = 0.37$ and $C_{cy} = 10.3$. The coefficients of spherical aberration are equal to $C_{3x} = 1.3$ and $C_{3y} = 60$ for the cylindrical lens and $C_{3x} = 0.56$ and $C_{3y} = 40$ for the crossed lens.

Let us compare these parameters with the analogous parameters of the single three-electrode axisymmetric lens whose geometry was described above. For $U_1 = U_3 = 1.0$ and the potential on the middle electrode $U_2 = -0.82$ the focal length of this lens is equal to the focal length of the investigated lenses in the xz plane ($f = 0.667$). The position of the focus in this case is equal to $z(F_i) = 0.522$, i.e., it almost coincides in position with the last electrode of the lens. Consequently, astigmatic lenses provide a significantly greater working interval.

The chromatic aberration coefficient of the axisymmetric lens is equal to $C_c = 4.8$, which is roughly an order of magnitude greater than the coefficients of the investigated lenses in the xz plane. In the yz plane the coefficients of chromatic aberration of astigmatic lenses are somewhat greater. However, as a consequence of the large focal length in this plane the exit slope angle of the trajectories is roughly five times smaller than in the axisymmetric lens, wherefore image blurring caused by chromatic aberration in the y direction remains smaller than in the axisymmetric lens by at least a factor of two.

The spherical aberration coefficient of the axisymmetric lens is equal to 4.3, which is several times greater than the coefficient of the investigated lenses in the xz plane. The spherical aberration coefficient of the astigmatic lenses in the yz plane is an order of magnitude greater, however, as a consequence of the fact that the cube of the slope angle of the trajectory in this plane is two orders of magnitude smaller, and image blurring in the y direction is again smaller than in the axisymmetric lens.

Figure 5 displays aberration figures due to spherical aberration for a parallel beam at the lens entrance, the outer trajectories of which are described by the equation $(x_0^2 + y_0^2)^{1/2} = 0.2$. Curve *1* corresponds to the cylindrical lens, and curve *2*, to the crossed lens. Each aberration figure was calculated as the set of points of the real intersections of the trajectories with the focal plane. Consequently, for a comparatively large entrance beam radius (0.2), they correspond to total spherical aberration and not just third-order aberration. It follows from Fig. 5 that the ratio of maximal dimensions of the aberration figure in the yz and xz planes is 3.5 for the cylindrical lens and 4.0 for the crossed lens. At the same time, the ratio of third-order aberration coefficients in the first case is equal to 46, and in the second, to 71. This fact is mainly explained by the large difference in exit slope angles of the trajectories in the xz and yz planes, as was indicated above. The radius of aberrational focal blurring in the axisymmetric lens for the same diameter of the entering parallel beam is equal to 0.3, which is at least five times
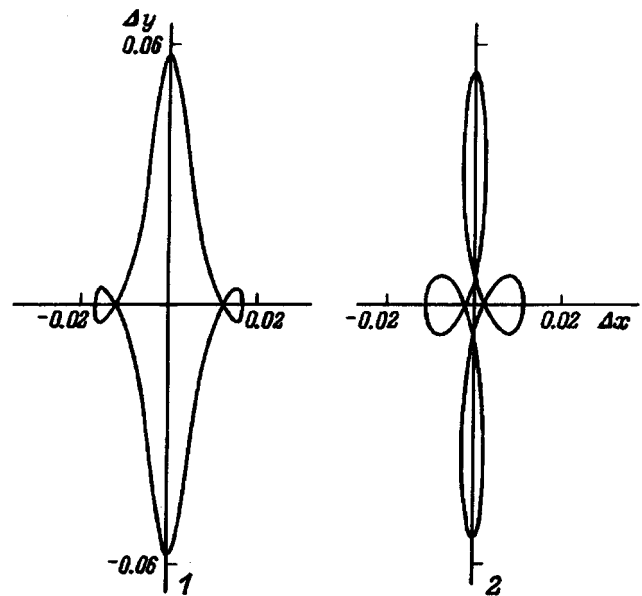

FIG. 5. Aberration figures in the focal plane for stigmatic regimes.

greater than the maximum spot size in astigmatic lenses.

In summary, a comparison of two types of lenses differing significantly in electrode design has shown that their aberrational properties are similar. In certain regimes both lenses allow simultaneous correction of spherical and chromatic aberration of a linear image. In stigmatic regimes a significant reduction of these aberrations is possible in comparison with an equivalent axisymmetric lens in the region of short focal lengths. In addition, astigmatic lenses provide a larger working interval, thanks to the shift of the principal planes, than does an axisymmetric lens. These properties of the investigated lenses make them promising candidates for application in probe systems.

[1] P. Hawkes and E. Kasper, *Principles of Electron Optics* (Academic Press, New York, 1989).
[2] L. A. Baranova and S. Ya. Yavor, *Electrostatic Electron Lenses* [in Russian], Nauka, Moscow (1986).
[3] V. M. Kel'man and S. Ya. Yavor, Zh. Tekh. Fiz. **31** (12), 1439 (1961) [Tech. Phys. **6**, 1052 (1962)].
[4] L. A. Baranova, N. S. Ul'yanova, and S. Ya. Yavor, Zh. Tekh. Fiz. **61**, 157 (1991) [Tech. Phys. **36**, 813 (1991)].
[5] L. A. Baranova, S. Ya. Yavor, and E. Munro, in *Proceedings of the SPIE Conference on Electron-Beam Sources and Charged Particle Optics, 1995*, Vol. 2522, p. 260.
[6] L. A. Baranova, S. Ya. Yavor, and F. H. Read, Rev. Sci. Instrum. **67**, (2), 1 (1996).
[7] Programme CPO3D, RB Consultants, Ltd., Department of Physics, University of Manchester, M13 9PL, United Kingdom.

Translated by Paul F. Schippnick

On the transformation of the radial profile of an intense ion beam by a plasma lens

A. A. Goncharov and A. N. Dobrovol'skiĭ

Institute of Cybernetics, National Academy of Sciences of the Ukraine, 252022 Kiev, Ukraine

V. F. Zadorozhnyĭ

Electrostatic plasma lenses (PL), proposed in Ref. 1 as a development of the idea of the Gabor lens with electronic space charge,[2] have over the course of many years demonstrated convincing advantages in comparison with conventional ion-optical systems. Their study continues to transition at a quickening pace from the sphere of fundamental research to applications.

Strong-current plasma lenses, i.e., lenses in which the potential of the transmitted beam substantially exceeds the maximum value of the external potential applied to the fixed electrodes of the lens, possess particularly interesting advantages. As experiments have shown,[3] such lenses possess many degrees of freedom, which allows one to vary the radial profile of the electric potential within wide limits. This makes it possible to eliminate spherical aberrations, thereby ensuring good focusing and defocusing of the beam (such a lens possesses the unique property of also working in the dispersing regime[4]), and also to use spherical aberrations to vary the radial profile of the beam on the target. This effect was discovered in Ref. 5, where it was shown that by regulating the spherical aberrations it is possible to regulate the radial profile of the beam transmitted through the plasma lens, in particular, to make it uniform. There, on the basis of the method of calculating the inner structure of nonlaminar axisymmetric charged particle beams with zero phase volume which was proposed and developed on the basis of a consistent kinetic approach in Ref. 6, a differential equation was obtained relating the initial beam profile at the entrance to the thin lens with the nature of the aberrations for which the beam profile at a given distance z becomes uniform.

The aim of the present paper is to find the self-consistent electric profile of a plasma lens transforming an inhomogeneous radial beam profile at the lens entrance into a uniform one in a given cross section beyond the lens.

For clarity it would be opportune to repeat the main points leading to the desired differential equation. We represent the radial profile of the beam current density $j(r)$ as a sum of elementary beams

$$j(r) = \sum_{n=1}^{N} i_n(r, V_n, z), \quad n \in [1, \ldots, N],$$

where V_n is the radial velocity acquired by the nth beam as it passes through the lens, and r and z are the radial and longitudinal coordinates, respectively.

This expression is valid at any point of the phase trajectory. By virtue of the equation of continuity in Lagrange form the following equation is valid for each of the elementary beams:

$$i_n = i_{0n}(r_{0n}(r, V_n, z)) \frac{r_{0n}(r, V_n, z)}{r} \left. \frac{\partial r_{0n}}{\partial r} \right|_{V=V_n}.$$

Representing the condition of uniformity of the beam $j(r)$ in an arbitrary cross section z in the form

$$\frac{\partial j(r)}{\partial r} = \sum_{n=1}^{N} \frac{\partial i_n}{\partial r} = 0$$

and taking account of the specifics of the passage of the beam through the thin lens, which affect only the radial component of the beam velocity, we obtain

$$\frac{\partial \ln i_{0n}}{\partial \ln r_{0n}} - \frac{z(V'_{0n} - V_{0n}/r_{0n})/V_b}{1 + V_{0n}z/V_{b_1 0n}} - \frac{zV''_{0n}r_{0n}/V_b}{1 + V'_{0n}z/V_b} = 0, \quad (1)$$

where V_b is the drift velocity along the z axis; V_{0n} is the radial velocity at the point $(r_0, 0)$, which generates the velocity $V_n(r)$, $V'_{0n} = \partial V_{0n}/\partial r_{0n}$, $V''_{0n} = \partial^2 V_{0n}/\partial^2 r_{0n}$.

This relation, carried back to the cross section $z=0$, where the thin lens is located, defines the conditions under which a beam with a given initial radial profile at the entrance to this lens can become uniform in an arbitrary prescribed cross section z. In particular, representing the radial profile of the entering beam as a Gaussian distribution $i_{0n} = i_0 \exp\{-r_{0n}^2/\theta^2\}$ and neglecting the quantities $V_{0n}z/V_b r_{0n}$ and $V'_{0n}z/V_b$, which is valid if $z \ll F$, where F is the focal length, reduces Eq. (1) to the form

$$d\left(-\frac{r_{0n}^2}{\theta^2} \right) = \frac{z}{V_b} \left(V''_{0n} + \frac{1}{r}V'_{0n} - \frac{1}{r^2}V_{0n} \right). \quad (2)$$

The solution of Eq. (2), which is bounded on the axis, has the form

$$V_{0n}(r_{0n}) = -\frac{1}{2}C_0 r_{0n} - \frac{1}{4}\frac{V_b r_{0n}^3}{z\theta^2}. \quad (3)$$

Allowing for conservation of energy in the thin plasma lens in the form $V_{0n}^2 = 2e\varphi(r_{0n})/M$ and the boundary condition $\varphi(r)|_R = \varphi_l$, we obtain the desired potential distribution
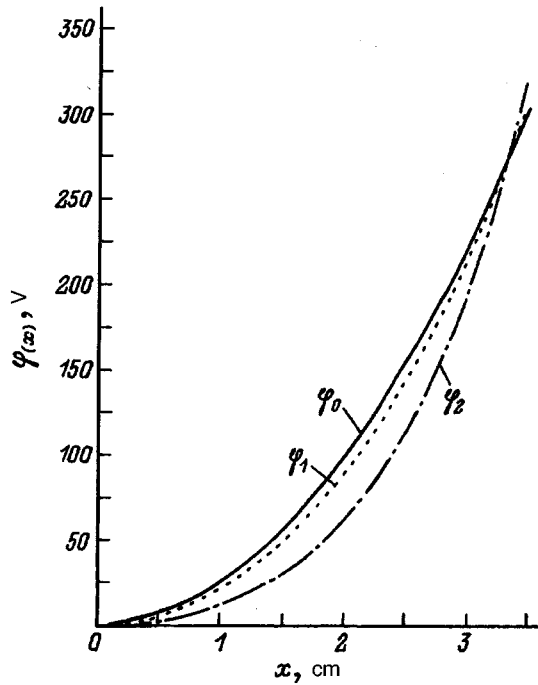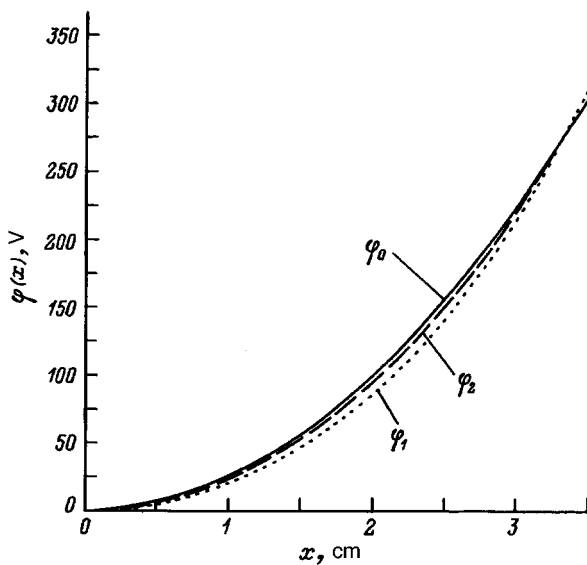
FIG. 1. Radial distribution of the potential $\varphi(x)$ for different values of the parameter θ: φ_0 is the curve without the term in x^3; φ_1 is the curve with the term in x^3 taken into account for $\theta=2$; φ_2 is the curve with the term in x^3 taken into account for $\theta=1$; $\varphi_l=300$V, beam energy 20 kV, $z=10$ cm.

$$\varphi(x)=\left[\left(\sqrt{\varphi}-\frac{V_b}{4z}\frac{R^3}{\theta^2}\right)x+\frac{V_b}{4z}\frac{R^3}{\theta^2}x^3\right]^2, \quad x=r/R. \quad (4)$$

Figure 1 plots the radial distribution of the potential for different values of θ. It can be seen that for the smaller value of θ the shape of the potential curve differs substantially from parabolic. Figure 2 plots the dependence $\varphi(x)$ for two

FIG. 2. Radial distribution of the potential $\varphi(x)$ for different values of the parameter z: φ_0 is the curve without the term in x^3; φ_1 is the curve with the term in x^3 taken into account for $z=10$ cm; φ_2 is the curve with the term in x^3 taken into account for $z=20$ cm; $\varphi_l=300$V, beam energy 20 kV, $\theta=2$.

different cross sections z.

In the solution of Eq. (1) we neglected the second term in the denominators of the second and third terms. In this case, the fundamental solution r_{0n}^{-1}, which we found for the simplified system, satisfies the complete homogeneous equation

$$\frac{(y'-r_{0n}^{-1}y)}{1+\varepsilon r_{0n}^{-1}y}+\frac{y^n r_{0n}}{1-\varepsilon y'}=0, \quad (5)$$

of which it is easy to convince oneself by direct substitution. And since the general solution of Eq. (1) is equal to the sum of the general solution of the homogeneous equation (5) and the particular solution of the corresponding inhomogeneous equation, the general solution of the complete equation is given by formula (3). Here note the following: iterative processes, as a rule, lead to infinite series. Equation (5), however, possesses a remarkable property: the second and higher approximations $o(\varepsilon^2)$, $o(\varepsilon^3)$, etc. are equal to zero. This finiteness of these series is extraordinarily important for finding the exact solution.

We will make use of this result to find the exact solution of Eq. (1) for initial distributions of parabolic form $i_{0n}=i_{00}(1-r_{0n}^2/a^2)$, where a is some constant. Then

$$\partial\ln\left[i_{00}\left(1-\frac{r_{0n}^2}{a^2}\right)\right]=y''+\frac{1}{r_{0n}}y'-\frac{1}{r_{0n}^2}y,$$

and consequently

$$y(r_{0n})=\frac{\ln i_{00}r_{0n}}{2}+\frac{C_1 r_{0n}}{2}-\frac{a^2}{2}\frac{1}{r_{0n}}$$
$$\times\left[\left(1-\frac{r_{0n}^2}{a^2}\right)\ln\left(1-\frac{r_{0n}^2}{a^2}\right)-\left(1-\frac{r_{0n}^2}{a^2}\right)\right]+C_0.$$

If we make the same assumptions about the integration constants and take the first term of the Taylor series of $\ln(1-r_{0n}^2/a^2)$, then for $a=\theta$ we obtain formula (3) for V_{0n}.

Now we can consider the second case. Let the field φ in the lens be so large that at the point r_0 there are N particles all with different velocities. Then the results of the first case do not obtain. We will solve the problem of obtaining a uniform beam in the z cross section in the following way.

We average the sum $\Sigma_1^N i_{0n}r_{0n}\partial r_{0n}/r\partial r$ over r_{0n} and obtain $j_0\partial r_0^2/2r\partial r$, where $r_0^2=\Sigma_1^N r_{0n}^2$, $j_0=(r_0^2/2r)^{-1}\Sigma_1^N j_{0n}r_{0n}\partial r_{0n}/r\partial r$. Now $r=r_0+V_0 z/V_b$, where V_0 is the mean velocity at the point r_0, which we must determine from the condition

$$\frac{\partial}{\partial r}\left[j_0\frac{1}{2r}\frac{\partial}{\partial r}r_0^2\right]=0.$$

Noting that $\partial r/\partial z=V_0(r_0)/V_b$, $V_0 z/V_b=\varepsilon x$, $V_0'z/V_b=\varepsilon x'$, and $x'=\partial x/\partial r$, we again arrive at an equation of the form (1)

$$\frac{\partial\ln i_0}{\partial\ln r_0}=\left(V_0'-\frac{V_0}{r_0}\right)-V_0^n r_0. \quad (6)$$

However, here the variables i_0 and V_0 at the point (r_0, V_0) have a different meaning. They characterize the motion at this point as a mixture of flows with different velocities.

Remark 1. In differentiating the velocity V_0 with respect to the radius r we assumed that this operation can always be performed. However, this assumption can fail if we assume generalized flows. Nonlaminar and moreover turbulent flows, as is well known, satisfy only a weak condition of differentiability. As Young[7] showed, weak limits are indeed a good mathematical model of turbulent motion.

Remark 2. In the present work we have postulated the distribution function I_{0n}; however, a different one may be entertained. As Morozov noted, the problem of beam formation in a consistent formulation should be a variational problem (Ref. 8, p. 300). Moreover, this should be a problem of optimal control if we take heed of Remark 1.

If we denote $\ln i_{0n}$ as u, then the above problem can be formulated thus. Find the control u such that

$$y' + \frac{1}{r_{0n}} y = u, \quad |u| \leq 1$$

and some functional $F(y, u)$ characterizing the quality of the process (energy consumption, beam diameter, current density at the cross section z, etc.) attains its minimum (maximum).

To summarize, in this work we have demonstrated the possibility of finding by analytical methods the radial profile of the electric potential of a thin plasma lens transforming a beam with nonuniform transverse profile into a uniform beam on a target located a given distance from the center plane of the lens.

A few words are in order here about the practical applicability of our results. The fundamental principle of equipotentialization of magnetic lines of force ensures in static plasma optics a one-to-one correspondence between the electric potential in the volume of the system and the magnetic flux Ψ. This correspondence is unique and simple: $\varphi = k\Psi(r, z)$, where k is a constant and the lines $\Psi(r, z) = \text{const}$ are the lines of force of the H field. Using only this (bijective) functional relation allows us quite accurately to predict the distribution $\varphi(z, r)$ over the fixed electrodes needed to eliminate spherical aberrations, i.e., to obtain the dependence $\varphi(0, r) = A r^2$ in the volume. At the same time, plasma optics in its present form does not give a direct recipe for obtaining any prescribed distribution $\varphi(0, r)$. At the same time, however, as experiments have convincingly demonstrated,[5,9] varying the parameters of the system (the transmitted beam current, configuration of the lines of force, distribution of the external potential) allows us to choose empirically the necessary distribution $\varphi(0, r)$.

[1] S. V. Lebedev and A. I. Morozov, Zh. Tekh. Fiz. **36**, 960 (1966) [Sov. Phys. Tech. Phys. **11**, 707 (1966)].

[2] D. Gabor, Nature **160**, 89 (1947).

[3] A. A. Goncharov and I. M. Protsenko, Ukr. Fiz. Zh. **36**, 1659 (1991).

[4] A. A. Goncharov, A. V. Zatyagin, and I. M. Protsenko, Pis'ma Zh. Tekh. Fiz. **15**(6), 1 (1989) [Tech. Phys. Lett. **15**, 205 (1989)].

[5] A. A. Goncharov, A. V. Zatuagan, and I. M. Protsenko, IEEE Trans. Plasma Sci. **21**, 578 (1993).

[6] V. S. Kuznetsov, V. I. Bogdanova, and O. L. Komarov, Radiotekh. Elektron. **16**, 1476 (1971).

[7] L. G. Young, *Lectures on the Calculus of Variations and Optimal Control Theory* (Saunders, Philadelphia, 1969).

[8] M. A. Leontovich (ed.), *Reviews of Plasma Physics* (Consultants Bureau, New York, 1980) [Russian original, Atomizdat, Moscow, 1974].

[9] A. A. Goncharov, A. N. Dobrovolsky, A. V. Zatuagan, and I. M. Protsenko, IEEE Trans. Plasma Sci. **21**, 573 (1993).

Translated by Paul F. Schippnick

Properties of diamondlike films obtained in a barrier discharge at atmospheric pressure

S. P. Bugaev, A. D. Korotaev, K. V. Oskomov, and N. S. Sochugov

Institute of High-Current Electronics, Siberian Branch of the Russian Academy of Sciences, 634055 Tomsk, Russia

Diamondlike films are synthesized from gaseous hydrocarbons in a barrier discharge at atmospheric pressure. The films were investigated using transmission electron microscopy, electron diffraction, and infrared spectroscopy. A technique for determining the quantitative characteristics of the films (hydrogen content, ratio of different types of carbon–carbon bonds and hydrocarbon groups) using standard samples is described. The highest-quality films were obtained from methane (ratio of hydrogen to carbon atoms H/C$=1.04$, fraction of diamondlike to graphitelike bonds $sp^3 : sp^2 = 100\% : 0\%$) and from a mixture of acetylene and hydrogen in the ratio 1:19 (H/C $= 0.73$, $sp^3 : sp^2 = 68\% : 32\%$). © *1997 American Institute of Physics.* [S1063-7842(97)01808-4]

INTRODUCTION

Diamondlike films (DLF) have attracted significant attention thanks to their unique properties, such as extreme hardness (up to 100 GPa), low wear and low friction coefficient, high electrical resistivity, chemical inertness and resistance to corrosion, high thermal conductivity, and biocompatibility. Today there is a wide range of techniques for obtaining diamondlike films.[1]

Diamondlike films were first obtained by deposition from carbon ion beams.[2] The films consisted almost completely of carbon atoms (a-C films) with diamond hybridization of the valence electrons (sp^3-hybridization), but the dimensions of the crystallographically ordered regions did not exceed a few tens of nanometers. It proved to be possible to enhance the properties of a-C films by using argon ion beams to sputter deposit a graphite target (with subsequent deposition of carbon atoms on the substrate) and to act on the film at the time of growth.[3] Diamondlike films obtained by chemical deposition from gaseous hydrocarbons activated with the help of radio-frequency,[4] microwave,[5] and other types of plasmas, contain a significant quantity of hydrogen (as high as 50%) and consist of several phases (finely dispersed diamond, graphite, and polymer phases). Nevertheless, such films are similar to diamond in their properties and are called amorphous diamondlike hydrogenated films, or a-C:H films.

The use of vacuum plasma methods to obtain diamondlike films informs the productivity and value of these technologies. In light of the above said, it is justified to attempt to produce diamondlike films with the help of electric discharges at atmospheric pressure. In the literature one can find reports on producing carbon films[6] in a radio-frequency plasma,[7] in an acetylene-torch flame,[8] and in a barrier discharge[9] at atmospheric pressure. Besides the simplicity that a barrier discharge presents, another attractive feature that it offers is the possibility of depositing diamondlike films on low-melting-point substrates. However, at present there are now data on the structure and properties of carbon films produced in a barrier discharge at atmospheric pressure. The acquisition of this information is the goal of the present work.

EXPERIMENTAL TECHNIQUE

Deposition of carbon films in a barrier discharge was carried out in a reaction chamber (Fig. 1), which consisted of two plane-parallel electrodes *1* and *2* and, located between them, the dielectric barrier *3* (a glass plate of thickness 1.5 mm) and vacuum-packed (*6*) gas-filled space *4* of thickness about 1 mm, through which with the help of an inflow–outflow system *5* the reaction gas (CH_4, C_2H_2, $C_2H_2 + H_2$, $C_2H_2 + Ar$) was pumped through the chamber. The pressure in the chamber was kept equal to atmospheric. Upon feeding unipolar microsecond voltage pulses of $100-200$ μm, amplitude U from -17 to $+17$ kV and pulse repetition frequency $f = 0.2 - 3$ kHz to the high-voltage electrode, a barrier discharge is created in the spacer gas. Under these conditions, the reaction gas dissociates, and carbon films are deposited on the dielectric barrier and the grounded electrode from the radicals forming as a result.

We found that higher-quality films are formed on the cathode than on the anode, and that their quality, as in Ref. 9, improves with growth of the breakdown voltage of the spacer gas. This is apparently due to the spatial inhomogeneity of the barrier discharge.[10] On this basis, all subsequent experiments on deposition of films were carried out at $U = -17$ kV with $f = 1$ kHz. The gas flow rate was $3 - 4$ liter/h. Further increase of the pulse repetition frequency f led to a decrease of the breakdown voltage of the spacer gas as a consequence of residual ionization of the gas, while lowering f did not lead to any changes besides a decrease of the rate of growth of the films. Varying the gas flow rate within the limits $1 - 30$ liter/h also had no effect on the quality of the films.

The films obtained as described above were examined by transmission electron microscopy, electron diffraction, and infrared absorption spectroscopy. Since diamondlike carbon films are usual amorphous, the most complete information about their structure and chemical composition is obtained spectroscopically: by Raman light scattering (Raman
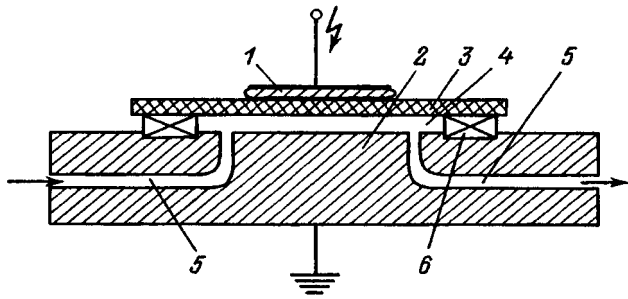
FIG. 1. Reaction chamber.

spectroscopy),[11] Auger spectroscopy,[12] electron energy loss spectroscopy,[13] etc. Infrared absorption spectroscopy is one of the most informative methods and at the same time it is a quite simple and accessible method for studying a-C:H films.

The region of valence vibrations of the C–H bonds is usually used for a quantitative and qualitative analysis of the structure and composition of a-C:H films. It is well known that the frequency and intensity of absorption of an individual C–H bond depend on the elemental and phase composition of the film as a whole, but only on the state of carbon atoms directly linked with a hydrogen atom (the type of electronic hybridization sp^3, sp^2, or sp and the type of hydrocarbon group CH_3, CH_2, or CH).[14] The concentration of any of the above-enumerated types of C–H bonds is proportional to the absorption integral calculated in the region of valence vibrations of the given bond type,[15]

$$n_i = A_i \int [(\alpha(k)/k)dk]_i, \tag{1}$$

where $\alpha(k)$ is the absorption coefficient, k is the wave number, and A_i is a proportionality factor.

On this basis, the authors of Ref. 16 proposed a formula for the ratio of hydrogen atoms to carbon atoms

$$H/C = 1/96 \, (\text{cm}^{-1}) \int (\alpha(k)/k)dk. \tag{2}$$

However, according to Ref. 15, which presents a detailed theoretical and experimental analysis of the behavior of various Si–H groups in a-Si:H films under conditions of infrared spectroscopy, the absorption intensities of the Si–H bonds in SiH, SiH_2, and SiH_3 differ substantially. This suggests that the proportionality factors A_i for a-C:H films are also different, and consequently formula (2) in general does not apply. Subsequent experiments have confirmed this.

Since the dependence of the absorption coefficient on the wave number for each type of C–H bond is approximately described by a Lorentz curve centered at the characteristic absorption frequency,[15] formula (1) may be rewritten in the form

$$n_i = A_i S_i/k_i, \tag{3}$$

where S_i is the area under the curve, k_i is the wave number corresponding to the characteristic absorption frequency of the given type of C–H bond.

Now the proportionality factors A_i are easily found from relations (3) by breaking down spectrograms of a material of known chemical composition and density into Lorentz curves (using tabulated values of k_i), computing the corresponding areas S_i, and calculating the concentrations n_i of the corresponding types of C–H bonds from the tabulated data.

We performed this procedure for thin films of polypropylene and polystyrene. We obtained the dependences $\alpha(k)$ from spectrograms taken with an IKS-29 infrared spectrometer. The spectrograms were decomposed into their constituent Lorentz curves with the help of a software package called ORIGIN, and the values of k_i were taken from Ref. 14. In this way we succeeded in obtaining the proportionality factors A_i for the hydrocarbon groups $sp^3 \, CH_3$ (a) (2960), $sp^3 \, CH_2$ (a) (2925), $sp^3 \, CH$ (2900), $sp^3 \, CH_3$ (s) (2870), $sp^3 \, CH_2$ (s) (2850), $sp^2 \, CH$ (3030) (sp^3, sp^2, and sp are the types of hybridization of the carbon atom; the numbers in parentheses are the wave number in cm^{-1} corresponding to the characteristic absorption frequency of the bond; and the notation (a) and (s) denote antisymmetric and symmetric vibrations, respectively. Because of the absence of standard thin films containing $sp^2 \, CH_2$ (3080) and $sp \, CH$ (3300) groups, a direct experiment to determine their proportionality factors was not carried out. Their values, according to Ref. 14, coincide roughly with the proportionality factors for $sp^2 \, CH$ (3030) and $sp^3 \, CH_2$ (2925), respectively, and we used these latter values in our calculations. It should be noted that the antisymmetric and symmetric vibrations belong to the same type of hydrocarbon group, and in the calculations it was necessary to use one type of vibration.

Now, having obtained from the spectrogram of an unknown hydrocarbon film the dependence $\alpha(k)$ and having decomposed it according to the above list of frequencies, it is possible to find n_i from formula (3) since the proportionality factors A_i for all possible C–H bonds are already known. Next it is easy to calculate the ratios of different types of hydrocarbon groups (CH_3:CH_2:CH), carbon–carbon bonds of different hybridization ($sp^3 : sp^2 : sp$) and the ratio of hydrogen atoms to carbon atoms

$$H/C = 12/(\rho/(\Sigma n_i \cdot 1_u) - 1), \tag{4}$$

where ρ is the film density and 1_μ is the atomic unit of mass.

This approach gives, in our opinion, more accurate quantitative characteristics of hydrocarbon films in general.

RESULTS AND DISCUSSION

Using methane as the reaction gas, the deposition rate of the hydrocarbon film was 2 μm/h. The density of the film, measured by submersion after separating it from the substrate, was 1.3 g/cm^2. Transmission electron microscopy studies showed that the films have a finely dispersed, amorphous structure (Fig. 2). It is clear that the film is inhomogeneous in density and most probably consists of several phases. The diffraction pattern obtained from a segment of the film consists of two diffuse rings whose centers correspond to interplanar distances of 2.08 and 1.18 Å. In Ref. 17 it was shown theoretically and experimentally that the given
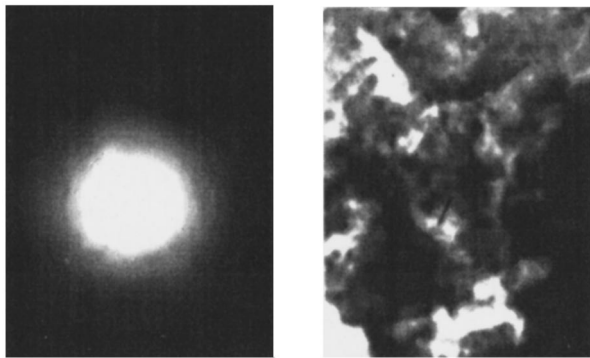
FIG. 2. Segment of film deposited from methane, and the corresponding diffraction pattern, both obtained by transmission electron microscopy.

pattern is characteristic of diamondlike films in which the dimensions of the regions of coherent electron scattering do not exceed $5-10$ Å.

The dependence of the absorption coefficient on the wave number in the region of valence vibrations of the C–H bonds for a film obtained in a barrier discharge from methane is shown in Fig. 3. The decomposition of $\alpha(k)$ into its component Lorentzians is indicated by dashed lines. The significance of the labeling of the curves was given above.

Calculations using formula (4) show that the given film contains a significant quantity of hydrogen (H/C$=1.04$). From formula (3) we obtained the ratios of hydrocarbon groups and C–C bonds: CH$_3$: CH$_2$: CH$=26\%$: 12% : 62%, and the bonds were all diamondlike. This confirms the conjecture made in Ref. 18 that a high hydrogen content in the film and in the gas phase (H/C$=4$) favors the formation of preferentially diamondlike bonds. A positive role here is also played by the initial sp^3-hybridization of carbon in the methane molecule. The relatively high content of CH$_3$ groups in the film points to a preferentially methyl mechanism of its
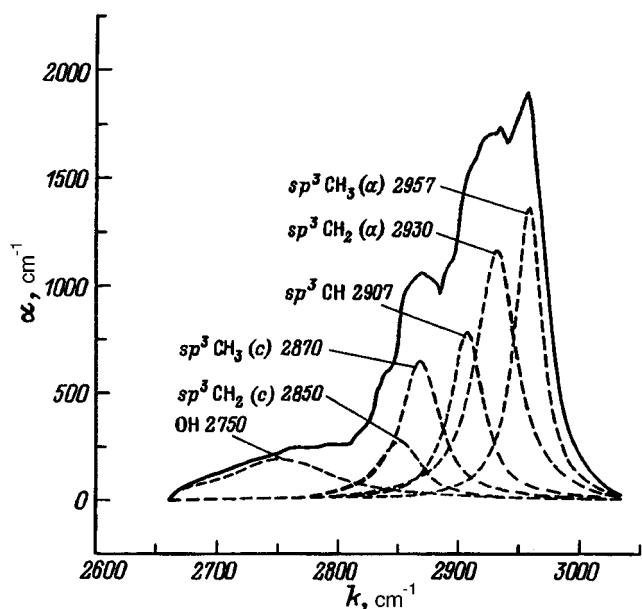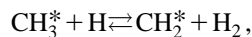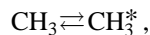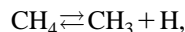

FIG. 3. Absorption coefficient versus wave number for a film obtained from methane, broken down into its components.

formation.[19] Methyl radicals, taking part in the film formation in a barrier discharge, as in any nonequilibrium plasma, appear as a consequence of dissociation of methane molecules by electron impact. The presence of CH$_2$ and CH groups is explained by the active removal of bound hydrogen from the film surface during the growth process (* indicates radicals included in the film)[20]

$$CH_4 \rightleftharpoons CH_3 + H,$$

$$CH_3 \rightleftharpoons CH_3^*,$$

$$CH_3^* + H \rightleftharpoons CH_2^* + H_2,$$

$$CH_2^* + H \rightleftharpoons CH^* + H_2. \tag{5}$$

The wide absorption band near 2750 cm^{-1} in Fig. 3 indicates the presence in the film of oxygen, whose inclusion is explained by insufficient purity of the methane. Chromatographic analysis revealed the presence in the methane of around 1 vol % oxygen. If we take its high chemical activity into account, then this amount is entirely sufficient to degrade the properties of the film, especially its adhesion. All the oxygen atoms in the considered film have sp^3-hybridization whereas for typical a-C:H films the sp^3/sp^2 ratio lies within the limits of 1 to 4 (Ref. 21). The film density (1.3 g/cm^3) was higher than in most polymers (0.95 g/cm^3) but lower than in solid diamondlike films (>1.7 g/cm^3).[22] Therefore, it may be surmised that the film consists of a mixture of at least two phases: a polymer phase, characterized by a high hydrogen content, and a finely dispersed diamond phase. The hydrogen fraction in the film exceeds the mean value for solid diamondlike films (H/C$=0.3-0.5$) but is near the upper limit of the interval of allowable H/C values since there are reports of a-C:H films with a hardness of 30–40 GPa, containing 50–60 at. % hydrogen.[22,23] Thus, the carbon films we have synthesized from methane in a barrier discharge can be assigned to the class of diamondlike a-C:H films.

The abrupt increase in the film growth rate (up to 60 μm/h) attendant to substituting acetylene for methane agrees with the experimental results of other authors[24] and is explained by the fact that the activation energy for film growth from C$_2$H$_2$ is lower than for CH$_4$ (Ref. 25). The wavenumber dependence of the absorption coefficient of film produced from acetylene (Fig. 4) differs from the analogous dependence for the film produced from methane (Fig. 3) by the presence of a significant quantity of graphite phase (absorption bands near 3020 and 3070 cm^{-1}) and carbyne (linear carbon) phase (3300 cm^{-1}). The ratios of types of C–C bonds and hydrocarbon groups were $sp^3 : sp^2 : sp$ $=34\%$: 55% : 11% and CH$_3$: CH$_2$: CH $=5\%$: 13% : 82%, respectively. The film density was around 1.2 g/cm^3. Despite their quite low hydrogen content (H/C $=0.66$), these films cannot be called diamondlike since the fraction of graphite and carbyne phases in them is large.

Even strongly diluting acetylene with argon (7% C$_2$H$_2$ + 93% Ar) does not lead to any positive changes in the structure of the film. Against the background of an insignificant increase of the fraction of diamondlike bonds due to a decrease in the fraction of carbyne (linear-carbon) bonds
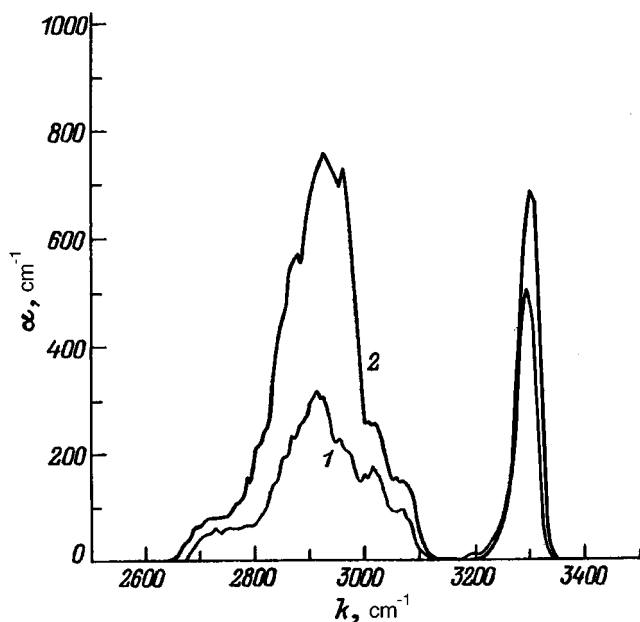
FIG. 4. Absorption coefficient versus wave number for films deposited in a barrier discharge: 1 — C_2H_2, 2 — C_2H_2 + 93% Ar.

FIG. 5. Absorption coefficients versus wave number for films deposited in a barrier discharge: 1 — 5% C_2H_2 + 95% H_2, 2 — 2.5% C_2H_2 + 97.5% H_2.

$(sp^3 : sp^2 : sp = 39\% : 54\% : 7\%)$ and an associated decrease in the percent content of the CH groups $(CH_3 : CH_2 : CH = 6\% : 19\% : 75\%)$ (Fig. 4) the relative hydrogen content of the film grows abruptly ($H/C = 1.30$). The reason for this is possibly an increase in the mean activation energy of the electrons in the discharge. A similar effect—an increase in the degree of electron dissociation of the molecules of the hydrocarbon gas attendant to dilution with argon—was observed (and modeled) in a plasma arc reactor.[20]

Diluting acetylene with hydrogen also does not lead to any noticeable changes all the way to 5% C_2H_2 in H_2. Figure 5 plots the dependence of the absorption coefficient on the wave number for a film obtained in a barrier discharge from a 5% C_2H_2 + 95% H_2 mixture. The rate of film growth dropped to 20 μm/h, and the density of the resulting film was 1.3 g/cm^3. In comparison with the two previous samples, no carbyne phase was observed in the film and the amount of graphitelike carbon was also significantly decreased $(sp^3 : sp^2 : sp = 68\% : 32\% : 0\%)$. An increase in the fraction of CH_3 and CH_2 groups $(CH_3 : CH_2 : CH = 22\% : 12\% : 66\%)$ and also an increase in the number of diamondlike bonds indicates the occurrence of a process of saturation by atomic hydrogen of the unsaturated carbon–carbon bonds ($C\equiv C$, $C=C$). Here the hydrogen content in the film remains low ($H/C = 0.73$) since bound hydrogen is removed from the film as it reacts with atomic hydrogen present in the discharge. Thus, ignoring the presence of an unknown amount of admixed oxygen (the absorption band near 2750 cm^{-1}), the film has parameters typical of diamondlike a-C:H films.

Further increase of the hydrogen content in the reaction mixture (2.5% C_2H_2 + 97.5% H_2) together with an increase in the fraction of diamondlike bonds $(sp^3 : sp^2 : sp = 89\% : 11\% : 0\%)$ leads to a growth of the H/C ratio in the film to $H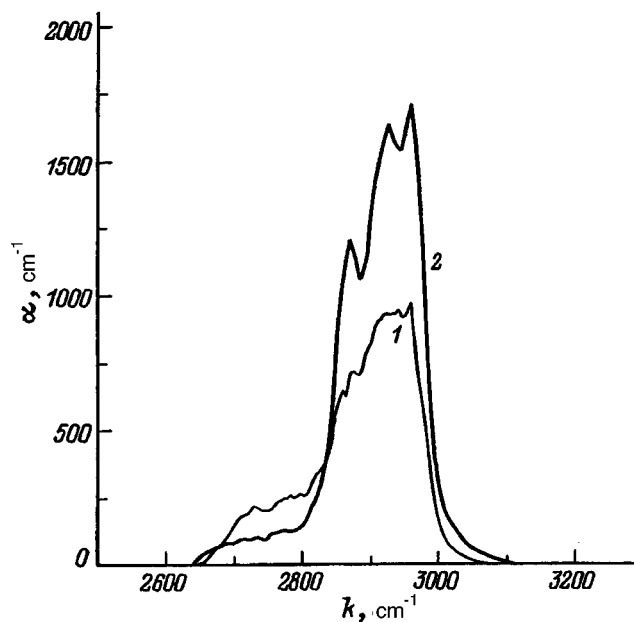/C = 1.45$ (Fig. 5). The hydrocarbon groups now stand in the ratio $CH_3 : CH_2 : CH = 18\% : 11\% : 71\%$. These numbers allow us to classify this film as a polymerlike a-C:H film.

CONCLUSIONS

Amorphous diamondlike hydrogenated films (a-C:H films) have been synthesized with high growth rates (tens of microns per hour) in a barrier discharge at atmospheric pressure from hydrocarbon gases. The highest-quality films were obtained from methane and from a 5% C_2H_2 + 95% H_2 mixture, which is the optimal acetylene–hydrogen mixture for a barrier discharge.

Infrared spectroscopic data were used to determine the chemical composition and other quantitative characteristics of the films. We employed an improved technique based on the use of standard hydrocarbon films. For films obtained from methane $H/C = 1.04$, $sp^3 = 100\%$, and for films obtained from the 5% C_2H_2 + 95% H_2 mixture $H/C = 0.73$ and $sp^3 : sp^2 = 68\% : 32\%$. On the basis of these numbers, these films may be called diamondlike. This conclusion is confirmed by transmission electron microscopy studies.

[1] S. Aisenberg, J. Vac. Sci. Technol. A **8**, 2150 (1990).
[2] S. Aisenberg and R. Chabot, J. Appl. Phys. **42**, 2953 (1971).
[3] E. G. Spencer, P. H. Schmidt, D. C. Joy, and F. J. Sansalone, Appl. Phys. Lett. **29**, 118 (1976).
[4] Z. Zhiao, Vacuum **40**(6), 505 (1990).
[5] W. Moller, in *Proceedings of the Third TATF, Strasbourg, France, 1991*, p. 273.
[6] T. G. Owano and C. H. Kruger, Plasma Chem. Plasma Process. **13**(3), 433 (1993).
[7] S. K. Baldwin and T. G. Owano, Appl. Phys. Lett. **67**, 194 (1995).
[8] D. B. Oakes and J. E. Butler, J. Appl. Phys. **69**, 2602 (1991).
[9] R. Schwarz and J. Salge, in *Proceedings of ISPC 11, Loughborough, England, 1993*, p. 1071.
[10] V. Gibalov and D. Braun, in *Proceedings of ISPC 10, Bochum, Germany, 1991*, Vol. 3.2, pp. 7.1–7.6.
[11] S. Kumar, Appl. Phys. Lett. **58**, 1836 (1991).

[12] V. V. Sleptsov, V. M. Elinson, and G. F. Ivanovskiĭ, Vak. Tekh. Tekhnol. **2**, No. 4, 69 (1992).

[13] J. J. Cuomo and J. P. Doyle, Appl. Phys. Lett. **58**, 466 (1991).

[14] L. J. Bellamy, *The Infra-Red Spectra of Complex Molecules*, 2nd ed. (Barnes and Noble, New York, 1958).

[15] M. Cardona, Phys. Status Solidi B **118**, 463 (1983).

[16] J. W. Zou, K. Schmidt, K. Reichelt, and B. Dischler, J. Appl. Phys. **68**, 1237 (1990).

[17] É. F. Chaĭkovskiĭ, V. M. Puzikov, and A. V. Semenov, Arkh. Nauki Mater. **7**(2), 187 (1986).

[18] B. Dischler and E. Bayer, J. Appl. Phys. **68**, 1237 (1990).

[19] D. G. Goodwin, J. Appl. Phys. **74**, 6888 (1993).

[20] Yu. A. Mankelevich, A. T. Rakhimov, and N. V. Suetin, Fiz. Plazmy **21**, 921 (1995) [Plasma Phys. Rep. **21**, 872 (1995)].

[21] X. Jiang and K. Reichelt, J. Appl. Phys. **68**, 1018 (1990).

[22] P. Catania and Y. Sampeur, in *Proceedings of CIP 93, Antibes, France, 1993*, p. 116.

[23] M. Ham and K. Lou, J. Vac. Sci. Technol. A **8**, 2149 (1990).

[24] A. Buuron and D. Schram, in *Proceedings of CIP 93, Antibes, France, 1993*, p. 89.

[25] B. V. Deryagin and D. V. Fedoseev, *Growth of Diamond and Graphite from the Gas Phase* [in Russian], Nauka, Moscow (1977).

Translated by Paul F. Schippnick

Theoretical model for describing degradation of thin hydrogen-containing films

G. V. Gadiyak

Institute of Computational Technologies, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

V. G. Gadiyak

Novosibirsk State University, 630090 Novosibirsk, Russia

M. L. Kosinova and E. G. Sal'man

Institute for Inorganic Chemistry, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

A new theoretical model is proposed to describe the behavior of films of composite hydrogen-containing compounds during heat treatment. The model is based on the thermal generation of hydrogen atoms and atoms of the composite compounds with their subsequent diffusion to the boundaries of the film and percolation through the surface into the atmosphere. Calculations are performed for the heat treatment of films of silicon nitride. A comparison with the experimental data in the literature demonstrates the high efficiency of the proposed model. © *1997 American Institute of Physics.* [S1063-7842(97)01908-9]

INTRODUCTION

Thin semiconductor and insulating films, obtained both by the traditional method of gas-phase deposition and by new methods using non-thermal activation (photo- and plasma-activation), contain a large amount of hydrogen.[1–3] Its presence is the main factor governing the physico-chemical stability and operating characteristics of the material. In particular, in MNOS structures (metal–silicon nitride–silicon oxide–semiconductor) using silicon nitride to record a charge, the concentration of traps for electrons and holes is one of the governing parameters affecting the reliability of devices.[4] A simple model of dehydrogenation of silicon nitride layers was proposed in Ref. 5 which explains quite well the experimental data obtained by the tensometric method of annealing the N–H bonds. The model allows for the following processes: 1) breaking of the N–H bonds with formation of atomic hydrogen, 2) regeneration of the N–H bonds, 3) diffusion of atomic hydrogen to the external interface, 4) formation of molecular hydrogen and its desorption (''passage of hydrogen through the surface''[5]) from the surface of the film, and 5) adsorption of molecular hydrogen to the surface of the film.

The authors of Ref. 5 also assumed that $C_{NH}^0 \gg C_{NH}^\infty$, which corresponds to complete removal of hydrogen from the film as $t \to \infty$. The system of equations describing the indicated processes was solved in the limiting cases in which the dehydrogenation process was limited by the breaking of the N–H bonds, by hydrogen transport through the film, and by its passage through the surface. In the first case the decomposition rate does not depend on the film thickness d_0, in the second the process rate $dC/dt \sim d_0^{-2}$, and in the third $dC/dt \sim d_0^{-1}$. However, the model proposed in Refs. 5 and 6 does not completely describe transformations taking place in the film during annealing. For example, it was shown earlier in Ref. 7 that during annealing of silicon nitride films, along with hydrogen, nitrogen is also released, which is not reflected in the kinetic equations of Ref. 5. In addition, the model is applicable only within a limited temperature range, above the rupture temperature of the Si–H bonds. At the same time, the electrical properties of silicon nitride films, in particular the concentration of charged defects, depend on the content of unsaturated (dangling) silicon and nitrogen bonds in configurations (in a simple nitrogen vacancy N≡Si, a divacancy Si≡N, a silicon vacancy Si–N and divacancies Si–N ... N–Si), which are traps for charge carriers. These traps determine various charge characteristics of metal–insulator–semiconductor devices,[4,8–10] specifically the fixed charge, density of surface states, position of the centroid of a strong-field-trapped charge, etc. The aim of the present paper is to develop a theoretical model of physico-chemical transformations in thin hydrogen-containing films linking the annealing conditions during their preparation with their charge characteristics.

PHYSICAL MODEL

It is assumed that in a thin film of MNx:H (where M = Si, B) the following bonds exist: M–N, M–H, and N–H. Processes taking place in the film can be described by the system of quasichemical reactions

$$M-H \underset{k_2}{\overset{k_1}{\rightleftarrows}} M_- + H, \tag{1}$$

$$N-H \underset{k_4}{\overset{k_3}{\rightleftarrows}} N_- + H, \tag{2}$$

$$M-H + H \overset{k_5}{\Rightarrow} M_- + H_2, \tag{3}$$

$$N-H + H \overset{k_6}{\Rightarrow} N_- + H_2, \tag{4}$$

$$M-N_- \underset{k_8}{\overset{k_7}{\rightleftarrows}} M_- + N, \tag{5}$$

$$M_- + N \overset{k_8}{\Rightarrow} M-N. \tag{6}$$

Equation (1) describes the rupture of an M–H bond with reaction rate constant k_1 and the formation of an M–H bond (the reverse reaction) with rate constant k_2. Equation (2) describes analogous processes for the N–H bond with forward and reverse rate constants k_3 and k_4. Equations (3) and (4) characterize the process of formation of molecular hydrogen in the volume of the film with participation of the M–N and N–H bonds, respectively, with rate constants k_5 and k_6 (Ref. 11). Equation (5) allows for the possibility of the formation of free atomic nitrogen and its entrapment by a free (dangling) M bond with rate constant of rupture of the M–N_ bond k_7 and rate constant k_8 of entrapment of nitrogen by the ruptured bond. It is assumed that the nitrogen has a dangling bond M–N_. Finally, Eq. (6) describes the process of saturation of the dangling bonds M_. As it is written, Eq. (5) assumes that the reaction of M–N bond rupture does not take place and formation of free nitrogen takes place only as a result of rupture of M–N_ (nitrogen with a dangling bond) (5).

System of chemical kinetic equations (1)–(6) corresponds to the following diffusion–kinetic equations:

$$\frac{\partial C_H}{\partial t} = \frac{\partial}{\partial x}\left(D_H \frac{\partial C_H}{\partial x}\right) + k_1 C_{MH} + k_3 C_{NH} - k_2 C_H C_M$$
$$- k_4 C_H C_{N_-} - k C_H C_{MH} - k_6 C_H C_{NH}, \tag{7}$$

$$\frac{\partial C_{MH}}{\partial t} = -k_1 C_{MH} + k_2 C_{M_-} C_H - k_5 C_{MH} C_H, \tag{8}$$

$$\frac{\partial C_{M_-}}{\partial t} = k_1 C_{MH} - k_2 C_{M_-} C_H - k_5 C_{MH} C_H$$
$$+ k_7 C_{MN_-} - k_8 C_{M_-} C_N, \tag{9}$$

$$\frac{\partial C_{H_2}}{\partial t} = \frac{\partial}{\partial x}\left(D_{H_2} \frac{\partial C_{H_2}}{\partial x}\right) + k_5 C_{MH} C_H + k_6 C_{NH} C_H, \tag{10}$$

$$\frac{\partial C_{NH}}{\partial t} = -k_3 C_{NH} + k_4 C_{N_-} C_H - k_6 C_{NH} C_H, \tag{11}$$

$$\frac{\partial C_{N_-}}{\partial t} = -k_7 C_{MN_-} + k_3 C_{NH} - k_4 C_{N_-} C_H + k_6 C_{NH} C_H, \tag{12}$$

$$\frac{\partial C_N}{\partial t} = \frac{\partial}{\partial x}\left(D_N \frac{\partial C_N}{\partial x}\right) + k_7 C_{MN_-} - k_8 C_{M_-} C_N, \tag{13}$$

where the constants k_1, k_3, and k_7 characterize the rate of thermal generation of hydrogen atoms upon rupture of M–H and N–H bonds and of nitrogen atoms upon rupture of M–N_bonds, respectively; the constants k_2, k_4, and k_8 characterize the rate of formation of bonds of hydrogen with dangling bonds M_and N_and entrapment of nitrogen atoms by dangling bonds M_ respectively; the constants k_5 and k_6 characterize the rate of formation of molecular hydrogen in

the film; D_H, D_{H_2}, and D_N are the diffusion coefficients for atomic and molecular hydrogen and atomic nitrogen. Here C_{MH} and C_{NH} are the concentrations of hydrogen atoms on the M–H and N–H bonds and C_{M_-} and C_{N_-} are the concentrations of the dangling bonds; C_H, C_{H_2} and C_N, C_{N_2} are the concentrations of free-atomic and molecular hydrogen and nitrogen, respectively. For system of equations (7)–(13) we used the following boundary conditions: an inner boundary impervious to hydrogen and nitrogen:

$$\frac{\partial C_H}{\partial x}(x=L, t)=0; \quad \frac{\partial C_{H_2}}{\partial x}(x=L, t)=0;$$

$$\frac{\partial C_N}{\partial x}(x=L, t)=0,$$

and at the outer boundary we used equality of fluxes:

$$-D_H \frac{\partial C_H}{\partial x}(x=0, t) = K_{10} C_H - K_{11} C_H^{cham}. \tag{14}$$

$$-D_{H_2} \frac{\partial C_{H_2}}{\partial x}(x=0, t) = K_{20} C_{H_2} - K_{21} C_{H_2}^{cham}, \tag{15}$$

$$-D_N \frac{\partial C_N}{\partial x}(x=0, t) = K_{30} C_N - K_{31} C_N^{cham}. \tag{16}$$

Here it was assumed that evolution of hydrogen and nitrogen takes place through a stage of desorption of products from the surface. C_H^{cham}, $C_{H_2}^{cham}$, and C_N^{cham} are the concentrations of atomic and molecular hydrogen and atomic nitrogen in the vacuum chamber. It is assumed that rapid recombination of atomic hydrogen and atomic nitrogen takes place in the reaction chamber with formation of molecular forms, i.e., the recombination time τ_{rec} is much less than the characteristic diffusion times τ_{dif} and characteristic times of formation of nitrogen and hydrogen in the film τ_{reac}. Therefore $C_H^{cham}=0$ and $C_N^{cham}=0$. The constants K_{10}, K_{20}, and K_{30} characterize the permeability of the boundary of the film to reaction products in it. The rate of permeation of molecular hydrogen from the vacuum chamber into the film was assumed to be zero. In other words, the process of adsorption of molecular hydrogen was assumed to be very slow and to have a small sticking coefficient. We also allowed for the permeability of the quartz glass of the vacuum chamber to hydrogen. In this case we may write the following relation for the hydrogen flux from the chamber into the surrounding air:

$$J_{H_2} = \gamma_2 (C_{H_2} - C_{H_2}^{amb}) S_{cham}, \tag{17}$$

where γ_2 is the permeability coefficient of quartz glass of thickness l_0, S_{cham} is the surface area of the vacuum chamber (here we have introduced yet additional constants and parameters to the problem—γ_2, l_0, S_{cham}).

The chamber walls were assumed to be impermeable to nitrogen as experiment indicates.[7]

The initial conditions for the concentrations were assigned in the form of uniform profiles along the thickness of the film

$$C_{MH}(x,t=0)=C_{MH}^0, \quad C_{M_-}(x, t=0)=C_{M_-}^0,$$

$$C_{NH}(x, t=0) = C_{NH}^0,$$

$$C_{N\_}(x, t=0) = C_{N\_}^0 \quad C_{H_2}(x, t=0) = 0,$$

$$C_H(x, t=0) = 0, \quad C_N(x, t=0) = 0.$$

For the initial concentrations of dangling bonds we used experimental data for a test case with silicon nitride,[7,10] which indicate that the concentration of the N–H bonds is equal to $C_{NH}^0 = 2.4 \times 10^{21}$ cm^{-3}. The concentrations of the Si–H bonds and of the dangling bonds Si_ and N_were assumed to be insignificant, i.e., we have $C_{SiH}^0 < 10^{16}$ cm^{-3}, $C_{Si\_}^0 < 10^{16}$ cm^{-3}, and $C_{N\_}^0 < 10^{16}$ cm^{-3}. To solve the system of equations (7)–(13) numerically, we used an implicit scheme. The time derivatives were approximated to first order. The time step was chosen to grow with time with increment $\delta t^n / \delta t^{n-1} = 1.1$.

To approximate the spatial derivatives, we used a second-order conservative scheme on a nonuniform grid. The nonlinear system of difference equations was solved by successive scalar fitting using the solution from the previous iteration. The accuracy of calculation was not worse than 0.1%.

To estimate the parameters k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, D_H, D_{H_2}, D_N, K_{10}, K_{20}, K_{30}, we invoked the following considerations. For the diffusion coefficients D_H, D_{H_2}, and D_N and generation rate constants we assumed the Arrhenius law

$$D_i = D_{i0} \exp(-\delta E_i / kT),$$

$$k_j = f_{j0} \exp(-\delta E_j / kT),$$

where the activation energy for diffusion of atomic hydrogen δE_H was taken to be equal to 0.45 eV (Ref. 12), and molecular hydrogen δE_{H_2}, 2.43 eV (Refs. 13 and 14).

The activation energies for bond rupture δE_j were taken to be as follows: $\delta E_1 = 1.5$ eV for rupture of the Si–H bond,[8] $\delta E_3 = 2.0$ eV for rupture of the N–H bond, $\delta E_7 = 2.2$ eV for rupture of the Si–N bond. The frequency factors f_{j0} are of the order of the characteristic frequency of the phonon vibrations of the atoms in the lattice $f_0 \sim 10^{11}$ s^{-1}. For the rate constants k_2, k_4, k_5, k_6, and k_8 we used the diffusion approximation, which allows us to write

$$k_2 = 4\pi D_H r_{MH} \xi_2, \quad k_4 = 4\pi D_H r_{NH} \xi_4,$$

$$k_5 = 4\pi D_H r_{MHH} \xi_5,$$

$$k_6 = 4\pi D_H r_{NHH} \xi_6, \quad k_8 = 4\pi D_H r_{MN} \xi_8,$$

where r_{MH} and r_{NH} are the interaction radii of atomic hydrogen with the dangling bonds of the M atoms and N atoms, respectively; r_{MHH} and r_{NHH} are the interaction radii of hydrogen atoms with hydrogen atoms on the M–H and N–H bonds, respectively; r_{MN} is the interaction radius of atomic nitrogen with dangling bonds of the M atoms; the constants ξ_2, ξ_4, ξ_5, ξ_6, ξ_8 are the probabilities, respectively, that an atom that has approached a defect will saturate the dangling bonds in reactions (1), (2), (5), and (6) or form molecular hydrogen in reactions (3) and (4). The values of all the constants used in the model are $K_{10} = 8 \times 10^{-8}$ cm/s,

$K_{20} = 5 \times 10^{-8}$ cm/s, $K_{30} = 2 \times 10^{-8}$ cm/s, $f_1 = 10^7$ cm^{-1}, $f_3 = 10^9$ cm^{-1}, $f_7 = 10^9$ cm^{-1}, $\delta E_1 = 1.5$ eV, $\delta E_3 = 2.0$ eV, $\delta E_7 = 2.2$ eV, $D_{H0} = 8 \times 10^{-4}$ cm^2/s, $\delta E_H = 0.45$ eV, $D_{H_20} = 5 \times 10^{-4}$ cm^2/s, $\delta E_{H_2} = 0.45$ eV, $D_{N0} = 4 \times 10^{-4}$ cm^2/s, $\delta E_N = 2.43$ eV, $r_{MH} = 10^{-7}$ cm, $r_{NH} = 10^{-7}$ cm, $r_{MHH} = 10^{-7}$ cm, $r_{NHH} = 10^{-7}$ cm, $r_{MN} = 10^{-7}$ cm, $\xi_2 = 10^{-5}$, $\xi_4 = 10^{-5}$, $\xi_5 = 10^{-7}$, $\xi_6 = 10^{-6}$, $\xi_8 = 10^{-6}$, $\gamma_1 = 10^{-4}$ cm/s, $\gamma_2 = 5 \times 10^{-5}$ cm/s.

RESULTS AND DISCUSSION

To test the proposed model we used experimental data on the time dependence of the pressure in a vacuum chamber during thermal processing. Since in the course of solving system of equations (7)–(13) we know the diffusive fluxes of atomic and molecular hydrogen and nitrogen we can find the nitrogen and hydrogen concentrations as functions of time; for nitrogen we have

$$C_{N_2}^{cham}(t) = -1/2V_0 \int_0^t S_k D_N \frac{\partial C_N}{\partial x}(x=0, t)dt, \tag{18}$$

and for hydrogen we have

$$C_{H_2}^{cham}(t) = -1/V_0$$
$$\times \left(\int_0^t S_k \left(D_{H_2} \frac{\partial C_{H_2}}{\partial x}(x=0, t) \right. \right.$$
$$\left. + D_H/2 \frac{\partial C_H}{\partial x}(x=0, t) \right) dt$$
$$\left. - \int_0^t S_{cham} \gamma_2 (C_{H_2}^{cham} - C_{H_2}^{cham}(t)) dt \right), \tag{19}$$

where V_0 and S_{cham} are the volume and surface area of the vacuum chamber and γ_2 is the permeability coefficient of quartz for molecular hydrogen.

In the model of an ideal gas it is possible to calculate the partial pressures of hydrogen and nitrogen

$$P_{N_2}(t) = C_{N_2}^{cham}(t) \cdot k_B T, \tag{20}$$

$$P_{H_2}(t) = C_{H_2}^{cham}(t) \cdot k_B T_0. \tag{21}$$

The latter quantities were measured experimentally in Ref. 7. In the solution of system (7)–(13) we also considered boundary conditions of another type. The approximation (7)–(13), (14)–(16) assumes that hydrogen and nitrogen migrate to the surface and are then desorbed from it with characteristic time $\tau_d \ll \tau_{des}$. In addition, it is also assumed that the reverse process—adsorption of nitrogen and hydrogen followed by their penetration into the film—is blocked. Another formulation of the boundary condition on the outer surface of the film is possible.[15] In this formulation the particle flux from the surface is proportional to the concentration difference between the film and the chamber. For example, for nitrogen we have

$$J_N \simeq h_N (C_N - C_N^{cham}),$$

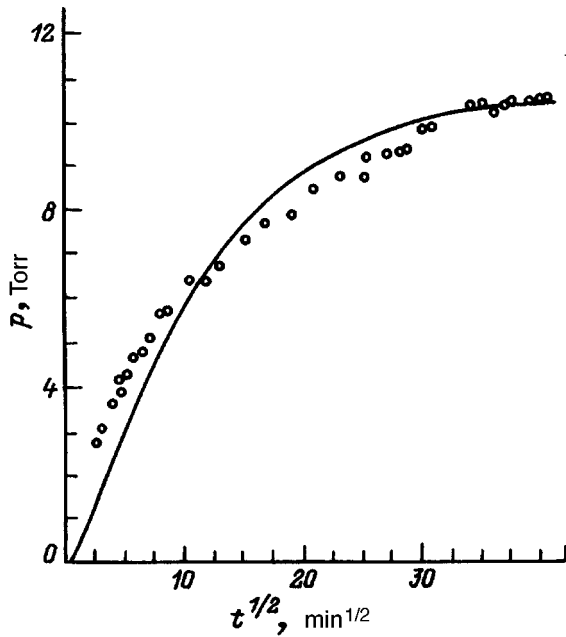where h_N is the permeability of the surface for nitrogen.

FIG. 1. Time dependence of the nitrogen pressure in the chamber for a Si_3N_4 film. Circles are experimental values from Ref. 6, the solid curve plots the calculated dependence.

FIG. 2. Time dependence of the total pressure for a Si_3N_4 film; notation—the same as in Fig. 1.

A study of the latter boundary condition showed that in this case the experimental values of the nitrogen pressure in the chamber are not achieved. The calculated value of the pressure was always two orders of magnitude lower than the experimental value for variation of the parameters over a wide range of values.

Figure 1 plots the experimental and calculated values of the nitrogen pressure in the chamber. The parameters governing the reproduction of the experimental values are the initial concentration of the N–H bonds C_{NH}^0, the permeability of the film surface K_{30} and the diffusion coefficient D_N. The initial concentration C_{NH}^0 and the permeability K_{30} influence the maximum value of the nitrogen pressure while the permeability and the diffusion coefficient influence the rate of growth of the pressure (they assign the initial value of the time derivative of the pressure and determine whether the profile is concave or convex, i.e., the sign of the second derivative).

Figure 2 plots experimental and calculated values of the total pressure versus time. Note that at early times the total pressure is controlled by the hydrogen pressure; then, around 10^2 min the hydrogen pressure reaches its maximum value and begins to decrease due to the permeability of the quartz walls. For this reason, starting around 10^2 min, when the values of the nitrogen and hydrogen pressures have equalized, the nitrogen pressure governs the behavior of the total pressure curve. The parameters governing the position and magnitude of the maximum pressure are the permeability of the film surface to molecular hydrogen K_{20} and the permeability of the chamber walls to atomic and molecular hydrogen, γ_2. Figure 3 shows the concentration distribution of atomic nitrogen C_N over the width of the film at different times. As follows from the figure, the nitrogen profile is
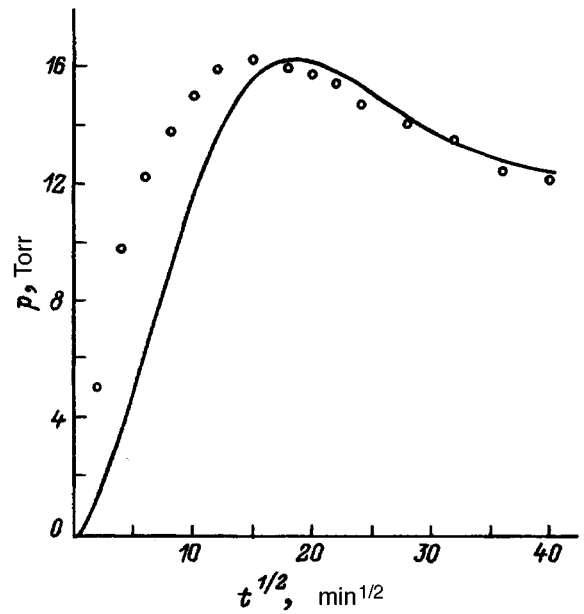
uniform over the width of the film; only at late times does the nitrogen concentration near the surface noticeably drop as a result of the fact that generation of nitrogen atoms decreases but the permeability of the surface remains unchanged. The profiles of the remaining concentrations C_{NH}, C_{N_-}, C_H, C_{H_2}, C_{SiH}, and C_{Si_-} remain uniform functions of the thickness coordinate of the film and vary only with time (Figs. 4–7). Thus, Fig. 4 plots the time dependence of the atomic hydrogen concentration. At first, there is a rapid growth up to $t \sim 10^{-4}$ min thanks to thermal rupture of the N–H bonds; then, after the concentration has reached a value
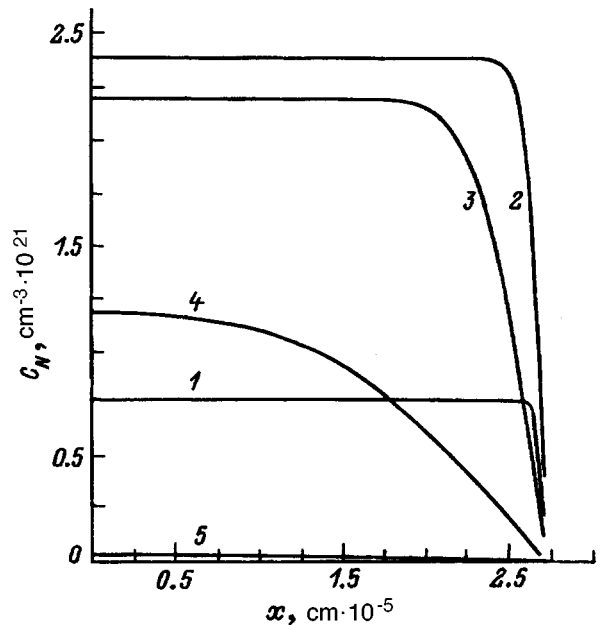

FIG. 3. Distribution of free nitrogen over the thickness of the structure: t (s): $1 — 10^1$, $2 — 10^2$, $3 — 10^3$, $4 — 10^4$, $5 — 10^5$.
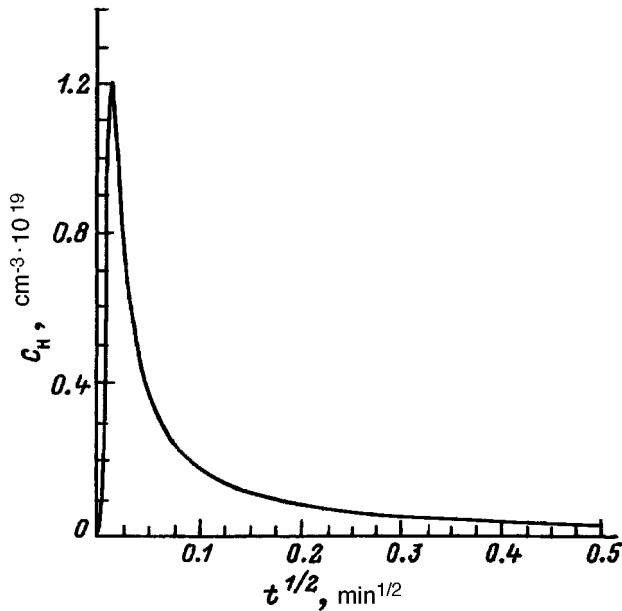
FIG. 4. Time dependence of the concentration of atomic hydrogen.

FIG. 6. Time dependence of the concentration of molecular hydrogen (*1*), the broken silicon bonds (*2*), and atomic nitrogen (*3*).

of $C_H \sim 10^{19}$ cm^{-3} it falls due to depletion of the hydrogen source (Fig. 5), i.e., the initial complement of N–H bonds. The temporal behavior of the nitrogen concentration (curve *3*, Fig. 6), is analogous to that of hydrogen. The characteristic time to reach maximum concentration is ~ 2 min, which is much larger than the time required by the mobile hydrogen to reach its peak concentration. These times stand in approximately the same ratio as the diffusion coefficients: $t_H/t_N \sim D_H/D_N$. Thanks to the small value of the diffusion coefficient (in comparison with hydrogen) and the smaller value of the permeability, the maximum value of the concen-
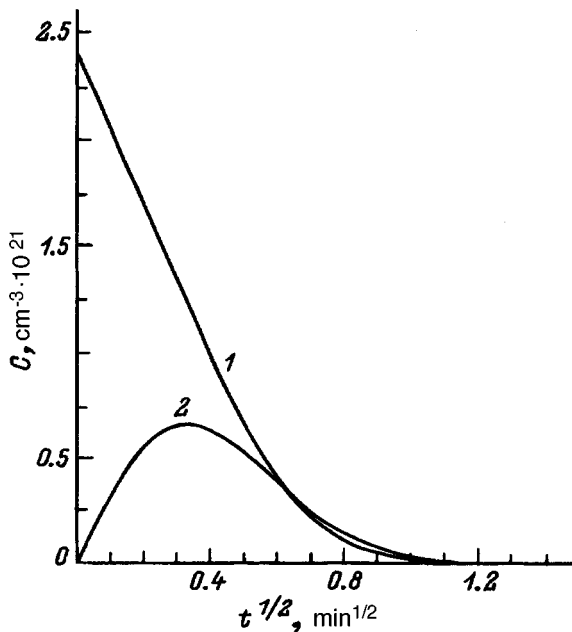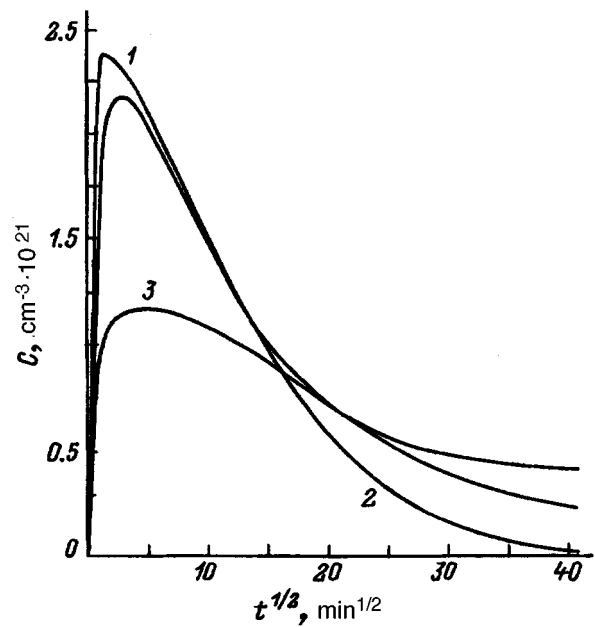
tration of mobile nitrogen reaches the value of the initial concentration of N–H bonds: $C_N \sim C_{NH} \sim 10^{21}$ cm^{-3}. The maximum concentrations of molecular hydrogen C_{H_2} and the dangling silicon bonds C_{Si^-} are of the same order of magnitude. The time to reach maximum is the same as before. Note that at late times the concentration of dangling silicon bonds C_{Si^-} rolls out to a plateau with the value $\sim 10^{21}$ cm^{-3}.

As our calculations showed, the parameters ξ_i for the reverse reaction constants turned out during fitting to have very small values ($\ll 1$) (see the parameter values listed above). This indicates that the obtained defects Si_and N_are unstable and an energy relaxation of the centers to a different geometric defect configuration takes place. In particular, in-

FIG. 5. Time dependence of the concentration of the N–H bond (*1*) and of the broken nitrogen bonds (*2*).
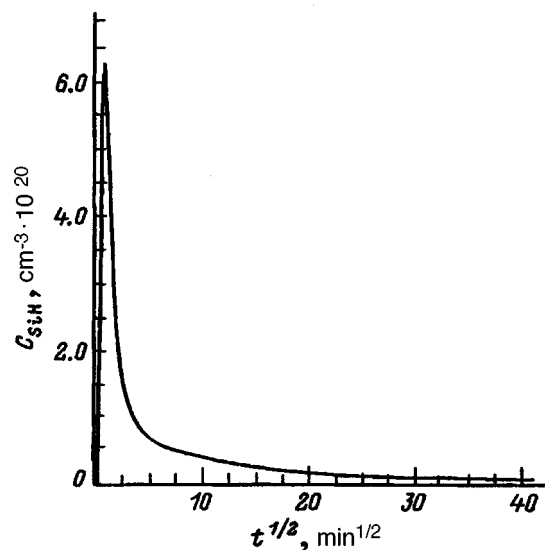

FIG. 7. Time dependence of the concentration of the Si–H bonds.

stead of formation of a nitrogen vacancy the dangling silicon bonds close up with each other and the reconfigured defect center is no longer paramagnetic.[8,16]

As follows from these calculations, during thermal processing not only is hydrogen removed from the film, but the concentration of dangling bonds also varies nonmonotonically. These dangling bonds are traps for carriers, and the latter result may therefore lead to electrical degradation of the film and, consequently, to degradation of the device structure. The nonmonotonic variation of the dangling bond concentration with time necessitates close attention to the process of annealing of the structure. In particular, at early times during the annealing it is possible to increase the defect concentration instead of decreasing it.

In conclusion, we would like to thank V. A. Gritsenko for numerous discussions of this work and I. I. Gadiyak for assistance with the production of the manuscript.

[1] Z. L. Akkerman, N. I. Fainer, and M. L. Kosinova *et al.*, in *Proceedings of SPIE, the International Society of Optical Engineering, 1992*, Vol. 1783, 530.

[2] Z. L. Akkerman, M. L. Kosinova, and N. I. Fainer *et al.*, Thin Solid Films **260**, 156 (1995).

[3] O. V. Polyakov and M. V. Badalian, Phys. Status Solidi B **185**, K1 (1994).

[4] A. V. Rzhanov, *Silicon Nitride in Electronics* [in Russian], Nauka, Novosibirsk (1982).

[5] Z. L. Akkerman, L. V. Khramova, and T. P. Smirnova *et al.*, Neorg. Mater. **26**, 988 (1990).

[6] Z. L. Akkerman, L. V. Khramova, and T. P. Smirnova *et al.*, Neorg. Mater. **26**, 993 (1990).

[7] L. V. Chramova, T. P. Chusova, and G. A. Kokovin, Thin Solid Films **147**, 267 (1987).

[8] V. A. Gritsenko, *Structure and Electronic Structure of Amorphous Insulators in Silicon MIS Structures* [in Russian], Nauka, Novosibirsk (1993).

[9] J. Robertson, Philos. Mag. B **69**, 307 (1994).

[10] W. L. Warren, J. Robertson, and J. Kanicki, Appl. Phys. Lett., **63**, 2685 (1993).

[11] G. V. Gadiyak, Fiz. Tekh. Poluprovodn. **31**, 257 (1997) [Semiconductors **31**, 207 (1997)].

[12] V. G. Litovchenko and A. P. Gorban', *Principles of the Physics of Metal–Insulator–Semiconductor Micro-Electronic Systems* [in Russian], Naukova Dumka, Kiev (1978).

[13] H. J. Stein, Mat. Res. Soc. Symp. Proc., Vol. 59, 523 (1986).

[14] G. A. Kachurin and I. E. Tyschenko, Mikroélektronika **23**(6), 3 (1994).

[15] B. E. Deal and A. S. Grove, J. Appl. Phys. **36**, 3770 (1965).

[16] L. V. Khramova, T. P. Chusova, and V. A. Gritsenko *et al.*, Elektron. Tekh. Ser. 6, Mater. **1**, 10 (1984).

Translated by Paul F. Schippnick

System for recording and analysis of reflection high-energy electron diffraction patterns

G. M. Gur'yanov, V. N. Demidov, N. P. Korneeva, V. N. Petrov, Yu. B. Samsonenko, and G. É. Tsyrlin

Institute of Analytical Instrument Making, Russian Academy of Sciences, 198103 St. Petersburg, Russia

An efficient and fast system for recording and analysis of reflection high-energy electron diffraction (RHEED) patterns is described. The software developed for this system includes three program packages: one for operating in the single-window mode, one for operating in the four-window mode, and one for the linear regime. Examples are given of the use of the system for monitoring and control of growth of III–V semiconductor compounds by molecular-beam epitaxy. Using this system, we discovered an effect wherein a periodic splitting of the RHEED peaks occurs during the growth of GaAs (100). © *1997 American Institute of Physics.* [S1063-7842(97)02008-4]

INTRODUCTION

Reflection high-energy electron diffraction (RHEED) is one of the most effective means for monitoring a solid surface and is widely used in semiconductor fabrication technologies, including molecular-beam epitaxy (MBE). Information in the RHEED method is contained in the image obtained on a fluorescent screen as a result of diffraction of electrons with energies in the range 5–50 keV incident on a solid surface at grazing angles.[1] An analysis of static diffraction images allows one to assess the crystal structure of the surface layer and also examine the microrelief of the surface, and also obtain quantitative information about the density of the monoatomic steps and the distribution of two-dimensional nuclei on the surface.[2,3] Observation of the dynamics of a RHEED pattern during epitaxial film growth makes it possible to evaluate with great accuracy the rate of growth, where the growth time of one monolayer corresponds (in general) to a period in the oscillating time dependence of the reflection intensity in the diffraction image.[4] For complete and effective exploitation of the RHEED method, it is necessary that recording and quantitative analysis of the diffraction patterns be performed on the image as a whole and in its parts. In this context, discrete measurements of the images should usually take significantly less time than growth of a single monolayer.

A photomultiplier is most frequently used for these purposes, recording the variation of the intensity of the RHEED signal on the fluorescent screen.[5] However, such a system is extremely limited and allows one to measure in real time the intensity of only one reflection, and at that an integrated intensity. At the same time, however, information about the RHEED intensity in different reflections and about the shape of the reflection is extremely important from the point of view of studying fundamental surface processes during MBE.

At present there exist a number of systems allowing one to record small segments of diffraction images in real time;[6,7] however, they either are not fast enough or are too cumbersome and costly and require the use of special instrumentation. In the present paper we describe an inexpensive and fast system for recording optical images, which allows one to computer analyze diffraction images and also measure real-time intensity variations of fragments of diffraction patterns (up to 2000 points on an IBM PC AT 286) with a discretization of 40 ms. Such a system also includes a software package designed for solving problems of molecular-beam epitaxy. Using this system allowed us to discover an effect of periodic splitting of the RHEED peaks during epitaxial growth of GaAs (100).

SYSTEM FOR RECORDING RHEED PATTERNS

A structural diagram of the system for recording RHEED patterns is shown in Fig. 1. The system consists of a television video camera (TV camera) with power supply, video monitor, frame grabbers, an IMB PC AT personal computer with monitor and specially designed software. Information from the fluorescent screen can also be written to a video recorder.

As the television camera we used a PTU-84 camera produced by the firm ''Volna'' in Novgorod and realized on a super silicon (diode array) camera tube, which provided high spectral sensitivity in the fluorescence region of the luminophore. It should be noted that any TV camera can be used in the given system as long as it puts out a standard television signal. The camera was mounted to the flange of the growth module of the MBE setup (ÉP1203) in such a way as to prevent direct light from hitting the objective and to provide a clear view of the required region of the image on the fluorescent screen. RHEED patterns, formed on the fluorescent screen, are recorded by the video camera, the video signal from which is fed in parallel to the frame-grabber card and video monitor, providing uninterrupted observation of the RHEED pattern during the course of the experiment. The frame grabber is realized on a printed circuit card fitting into a free slot of the computer. The computer can be any IBM-compatible personal computer.

The frame grabber is designed to convert an analog video signal into digital form (256 gray levels, 512×512 pixels) and to store the grabbed frame in a buffer with a 50-Hz refresh rate. It provides programmable read-out from
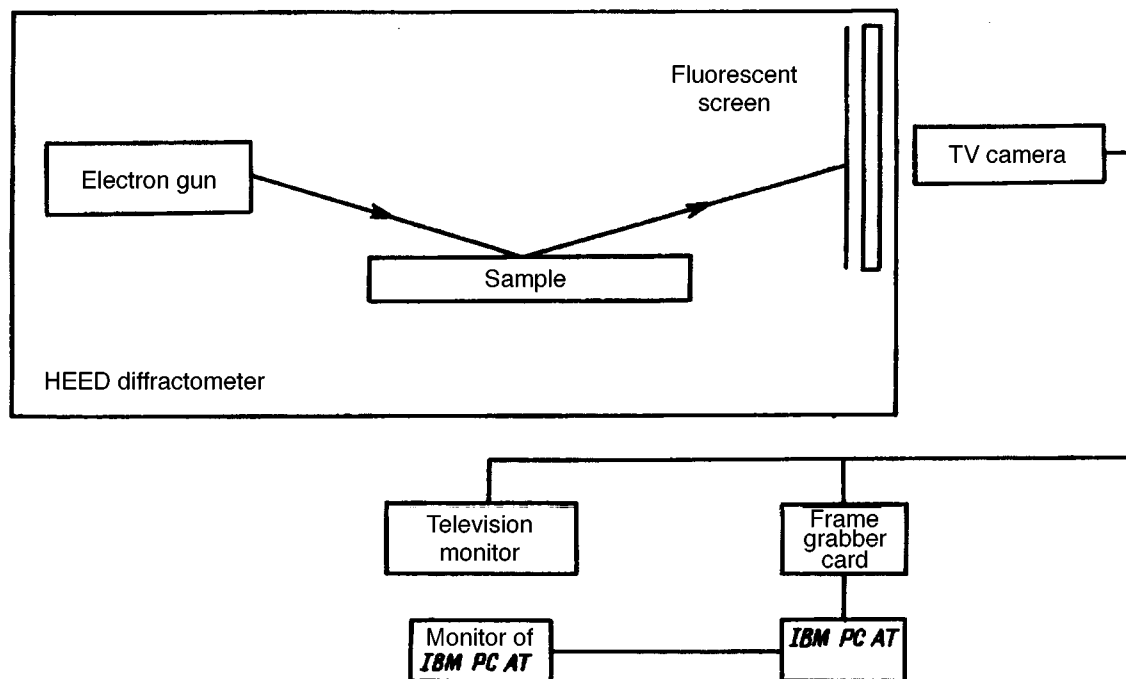
FIG. 1. Structural diagram of the system for recording diffraction images.

the buffer to the computer. A structural diagram of the frame grabber, elucidating its operation, is shown in Fig. 2. The device includes an analog-to-digital converter (ADC) with an input amplifier for amplitude encoding of the input video signal, a master clock providing temporal discretization of the video signal, CTX and CTY counters designed to address the buffer in the recording regime, RGX and RGY registers designed to address the buffer in the regime of programmed data transfer to the computer, bus formers (BF and DAT BF) ensuring signal matching. Line and frame clock pulses (LCP and FCP), tapped from the TV camera, ensure synchronization during temporal encoding of the television signal.

The main specifications of the system when using the type of TV camera indicated are the following: spectral sensitivity region 350–800 nm, maximum sensitivity of the TV camera 5×10^{-5} lm, number of pixels 512×512, read-in time of one pixel to the buffer 80 ns, number of gray levels 256.

SOFTWARE

The software for this system is written in C programming language, and is divided into three independent packages: a) the single-window regime, b) the four-window regime, and c) the linear regime. All three packages consist of programs for information acquisition and processing. All of the programs have a user-friendly menu with explanations.

a) *The single-window regime.* The main goal of this package is to record intensity variations within the chosen segment of the diffraction image (reflection) at the video refresh rate (50 Hz) and to determine the rate of growth of the epitaxial layer of the desired composition by secondary processing of the acquired information. When working in this regime, the overall pattern is first culled from the fluorescent screen and imaged on the monitor (512×512 pixels).

Then, an area of interest is selected on the monitor (128×128 pixels in size) and imaged on the monitor with magnification. After this, the size of the window is assigned and its position chosen at some arbitrary point of the magnified image. Next, a time is chosen over which information will be acquired from the window and on the operator's command the program commences readings. Here the information being recorded (signal intensity) is averaged over the area of the window and displayed on the monitor in the form of a dependence of the intensity on measurement time. When the set time is up or the program is paused by the operator before completion, the operator can, with the help of special markers, measure intervals between arbitrary points in time, calculate the rate of growth, write information to disk (including information about the diffraction image) or open up the obtained result. If this is necessary, then the user can subsequently analyze the acquired information with the help of the processing program. An example of the use of this program is given in Fig. 3. This example pertains to the growth of a GaAs/(Al,Ga)As/AlAs hetero semiconductor superlattice (the left, middle, and right parts of the oscillating dependence, respectively), grown at a substrate temperature of 600 °C and effective arsenic pressure in the growth zone of 4.2×10^{-6} Pa. Information was read off in the zero reflection in the $[0\bar{1}1]$ at an electron beam energy of 12.5 keV and $1°$ incidence angle. Note that each opened-up result can be provided with an explanatory caption (up to 150 characters).

b) *Four-window regime.* This package can be used when measuring RHEED intensities in several diffraction beams at once at a rate equal to the video refresh rate (50 Hz). The main difference between this case and the single-window regime is the possibility of choosing up to four windows of arbitrary size at arbitrary positions within the diffraction image; in every other respect all the options of the previous
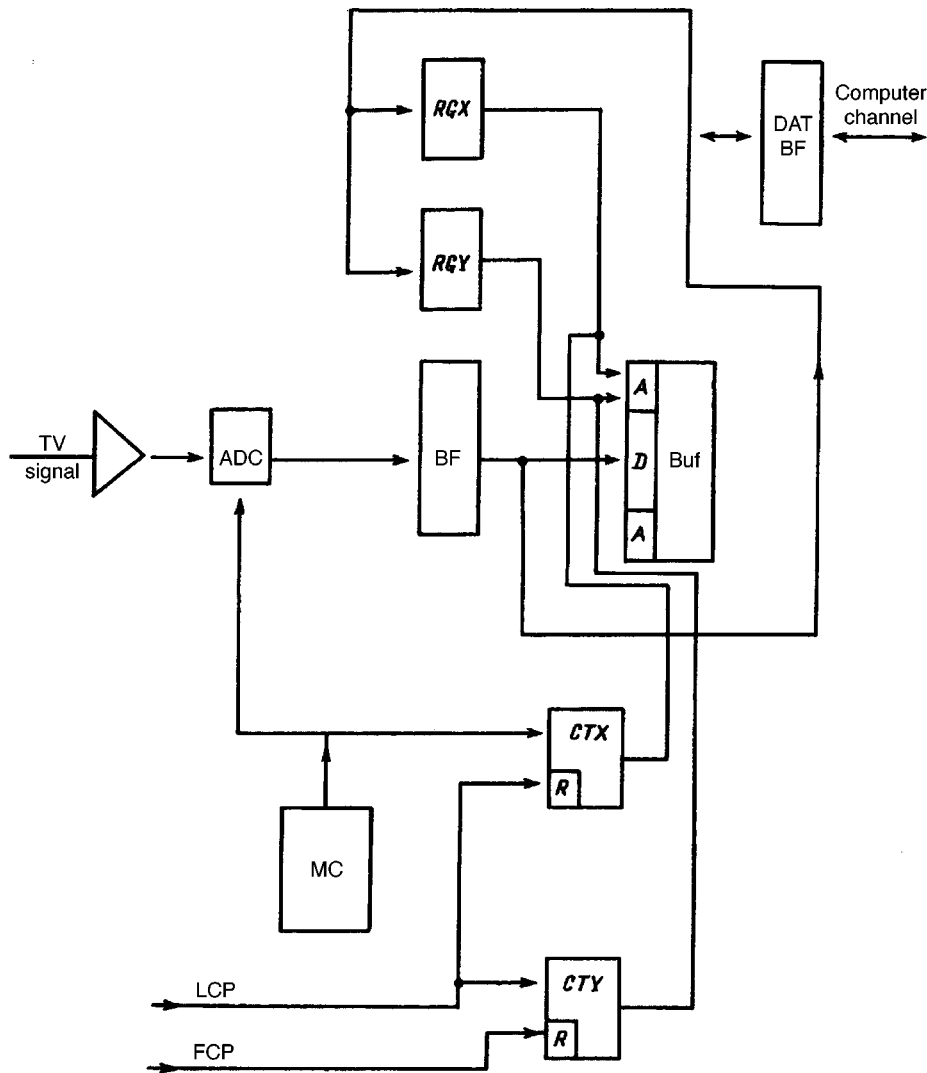
FIG. 2. Structural diagram of the frame grabber. ADC — 1107PV2 analog-to-digital converter, MC — master clock, LCP and FCP — line and frame clock pulses, Buf — buffer, RGX and RGY — registers for addressing the buffer when transferring information to the computer, CTX and CTY — binary counters for addressing the buffer, BF and DAT BF — bus formers.

package are retained. Figure 4 shows an example of the use of this software for a GaAs epitaxial layer grown under conditions of transition of the surface reconstruction from (2×4) to (4×2) (substrate temperature 550 °C, flux ratio $As_4/Ga = 1/2$). Time dependence curves of the RHEED intensity were read off in the $[011]$ direction at different reflections. From the figure the presence of a phase shift of the oscillations in different diffraction beams is evident, which demonstrates possibilities of the system when investigating peculiarities of the mechanism of MBE growth.

c) *The linear regime.* This package is intended for analysis of profiles of diffraction images with a frequency equal to the video refresh rate (50 Hz). The main difference from the previous regimes is the possibility of measuring RHEED in-
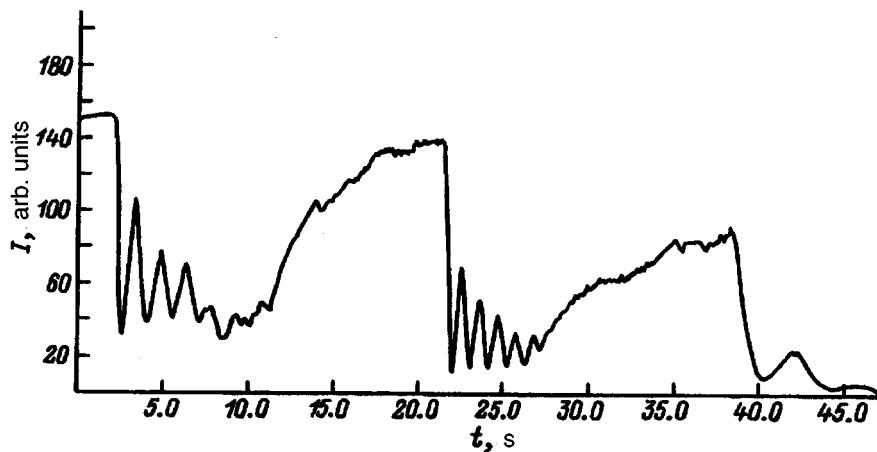

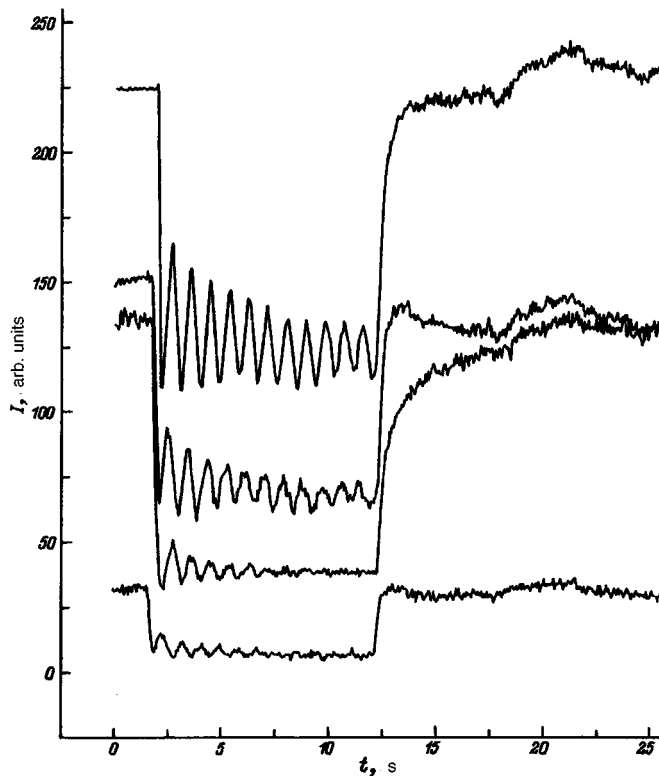FIG. 3. Time dependence of the RHEED intensity for the single-window mode.

FIG. 4. Time dependence of the intensity at different points of the diffraction image for the four-window mode.

FIG. 5. Profile of the (01) reflection for growth of GaAs (100), measured at different times (in numbers of monolayers): 1 — initial profile (before growth), 2 — 0.5, 3 — 1.0, 4 — 2.5, 5 — 3.0.

tensities in the quasilinear variant between two arbitrarily chosen points in the area of interest of the diffraction image. The line consists of a set of windows (up to 16) with arbitrary but fixed size. All possible operations when recording the image (monitoring the RHEED intensity in any window, writing to disk, opening up the results, the possibility of operational measurement of the rate of growth) remain the same as in the previous regimes. The information obtained in this regime is the most complete in comparison with the above two regimes and allows one to investigate complex physical phenomena associated with MBE growth.

The secondary processing program allows the user to graphically image information (in the form of time dependences of the signal intensity) recorded in any of the windows at different wavelengths in the unsmoothed or smoothed (by spline approximation) regime, where from 1 to 16 curves can be displayed on the screen simultaneously. Next, the rate of growth or various characteristic times can be calculated (e.g., the time of evanescence of the oxide layer, the rate of change of the surface reconstruction, the transition time from a two-dimensional to a four-dimensional mechanism of growth of the film, etc.).

EFFECT OF PERIODIC SPLITTING OF RHEED PEAKS DURING GROWTH OF GaAs(100)

The results for the linear regime demonstrate the highly informative nature of the RHEED method for studying MBE growth processes. In particular, the use of this regime allowed us to discover an effect of periodic splitting of the RHEED peaks during epitaxial growth of GaAs (100).

After the standard procedure of removing the oxide layer from the substrate surface[8] and growing a buffer layer under conditions that ensured 2×4 surface reconstruction, growth was interrupted and the surface was kept under a stream of arsenic to smooth it down. After this, the Ga door was opened and measurements were made of changes in the profile of the intensity distribution in the specular reflection in the $[0\bar{1}1]$ direction in the linear regime over the course of 30 s. Here the range of measured angles of the diffraction beam along the profile was $1-4°$ and the angle of incidence of the primary electron beam was equal to $1°$.

Figure 5 displays profiles of the (00) reflection in the MBE growth process at different times. The essence of the discovered effect reduces to a redistribution of the intensity of segments of the diffraction image, located along the dif-

fraction line near the specular reflection. As it passes through its intensity minimum, the oscillating specular reflection splits into two or three maxima, where the additional peaks are located at larger ($>1°$) reflection angles. Despite the fact that there have been many studies reporting the observation of oscillations and there has even been mention of a phase shift of the oscillations recorded at different points of the diffraction pattern, this is the first report of such a result.

We compared the distribution of the additional intensity maxima relative to the main peak of the specular reflection with the distance between the zero reflection and the fractional reflections of the zero Laue zone. We took this as a manifestation of surface reconstruction during the MBE growth process in the direction parallel to the plane of incidence of the ray allowed at the times of the appearance of nuclei of the new layer during layered growth, i.e., at the times when the intensities of the main reflections of the diffraction pattern are minimal. We hope to present a more detailed study of this effect in a separate paper.

CONCLUSION

We have described an inexpensive and effective system for recording and analysis of RHEED patterns and have presented concrete examples of its application to the growth of III–V semiconductor compounds by molecular-beam epitaxy. In our view, the system we have described makes it possible to observe and analyze all the main features revealed by the RHEED method during molecular-beam epitaxy. In addition, the system we have described allowed us to discover an effect of periodic splitting of the RHEED peaks during epitaxial growth of GaAs (100). I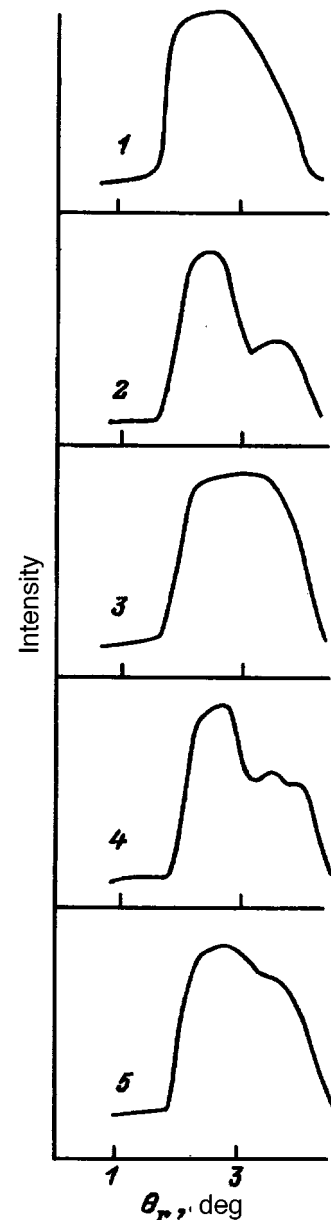n addition, it may be noted that this system can be incorporated in an automated growth complex and will enable the growth of quantum-size structures with an accuracy of film thickness control of fractions of a monolayer.

[1] M. A. Herman and H. Sitter, *Molecular Beam Epitaxy: Fundamentals and Current Status* (Springer-Verlag, Berlin, 1989), p. 376.

[2] P. J. Dobson, B. A. Joyce, J. H. Neave, and J. H. Zhang, J. Cryst. Growth **87**, 1 (1987).

[3] A. P. Senichkin, A. S. Bugaev, and R. A. Molchnovsky, *Abstracts of the First International Symposium ''Nanostructures: Physics and Technology''*, St. Petersburg, Russia (1993), p. 102.

[4] J. H. Neave, B. A. Joyce, P. J. Dobson, and N. Norton, Appl. Phys. A **31**, 1 (1983).

[5] C. Ca and H. H. Wieder, Rev. Sci. Instrum. **61**, 917 (1990).

[6] R. Bolger and P. K. Larson, Rev. Sci. Instrum. **57**, 1363 (1986).

[7] J. S. Resh, J. Stroizer, K. D. Jamison, and A. Ignatiev, Rev. Sci. Instrum. **61**, 771 (1990).

[8] G. M. Gur'yanov, N. N. Ledentsov, V. N. Petrov *et al.*, Pis'ma Zh. Tekh. Fiz. **19**(18), 64 (1993) [Tech. Phys. Lett. **19**, 591 (1993)].

Translated by Paul F. Schippnick

Effect of focusing of primary electrons on their reflection from a crystal and on the associated Auger emission

M. V. Gomoyunova and I. I. Pronin

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

The orientational dependence for different groups of secondary electrons — quasi-elastically scattered, inelastically reflected with excitation of a plasmon and with ionization of the core level $M_{4.5}$, and the Auger electrons $M_{4.5}VV$ — are measured in the primary electron energy range $0.6-1.5$ keV. The data are obtained for a Nb (100) single crystal by varying the azimuthal angle of incidence of the primary beam, with complete collection of secondaries. A relationship is established between the processes of focusing and defocusing of the electrons that have penetrated into the crystal in the $\langle 110 \rangle$ and $\langle 133 \rangle$ directions, which differ substantially in the atomic packing density. Specific details of the Auger orientation effect, due to the focusing-induced variation of the flux density of the reflected electrons, are identified and explained. The contributions, both of anisotropy of ionization of the core level and of variation of the backscattering intensity, to the angular dependence of Auger emission and reflection with ionization loss are estimated. The possibilities of using such orientational dependences for an element-sensitive analysis of the local atomic structure of surfaces are assessed. © *1997 American Institute of Physics.* [S1063-7842(97)02108-9]

INTRODUCTION

In recent years, in connection with the large successes achieved in the structural analysis of surfaces by moderate-energy photoelectron and Auger electron diffraction,[1–3] and also with the application toward these ends of quasi-elastically scattered electron diffraction,[4–7] interest has again arisen in orientation effects due to the primary beam.[3,8–12] The reason for this interest lies in the generality of the mechanism of formation of the effects used by these diffraction methods, which consists in focusing of electrons by atomic chains.

Effects associated with the primary beam, as was shown already in many early works,[13–20] are manifested in an amplification of the secondary electron emission as the primary electrons move along the densely stacked atomic planes and directions of the crystal. For a long time they were explained on the basis of interference of Bloch waves within the framework of the formalism of the dynamic theory of electron diffraction.[18,21,22] These ideas, however, are difficult to apply to the description of electron scattering processes in the thin near-surface layers of a crystal (with thickness of the order of a monolayer), where it is necessary to allow for the discrete nature of the construction of its atomic planes. In this situation, it is simpler to use models based on an analysis of electron scattering by the individual atoms of the crystal (forward focusing) and focusing of electrons by atomic chains.[8,9,23] To date, however, not all aspects of this problem have been sufficiently well examined and the literature continues to discuss questions pertaining to the maximum lengths of the focusing chains, their dependence on interatomic distances and electron energy, the mechanism of electron defocusing, etc.[11,24–30] What is needed is a clarification of the possibilities of using orientation effects due to the primary beam in the structural analysis of surfaces.

The aim of the present paper is, first of all, to obtain information about focusing and defocusing of primary electrons in the thin near-surface layer of a crystal by investigating the orientation effect for quasi-elastically scattered electrons and electrons that have undergone a one-time energy loss upon reflection. It is, second, to clarify the possibilities of applying the orientational dependence of Auger emission and reflection with ionization energy losses to an element-sensitive analysis of the local atomic structure of surfaces. As the object of study we have chosen a single crystal of niobium, whose surface structure has been well studied.

EXPERIMENTAL TECHNIQUE

The measurements were performed in a three-grid quasispherical braking-field energy analyzer with an energy resolution of 0.5% in the range 0.6–1.5 keV of primary electron energies E_p. Secondary electrons in different segments of the spectrum were recorded by modulational methods. Primary attention was given to electrons emitted by the thin near-surface layer of thickness on the order of 10 Å. Such electrons are quasi-elastically scattered electrons that have interacted with phonons upon reflection, electrons that have been reflected inelastically with one-time characteristic energy losses, and Auger electrons. The amplitude of the elastic reflection peak may serve as a measure of the intensity of quasi-elastic scattering since in the investigated energy range the above types of electrons make the overwhelming contribution.[7]

We investigated the effect of orientation of the primary beam on the intensity of secondary electron emission in the most direct way—by varying its azimuthal plane of incidence with the polar angle Θ held fixed at 45°. The azimuthal angle of incidence of the beam φ could be varied within the limits $0-360°$. The azimuthal dependences re-
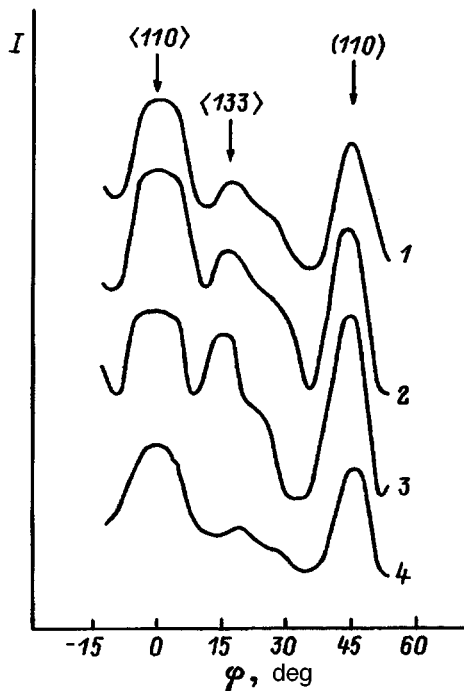
FIG. 1. Orientational dependence $I(\varphi)$ of the intensity of quasi-elastic electron scattering (*1*), reflection with single excitation of a bulk plasmon (*2*) and with ionization loss $NbM_{4.5}$ (*3*), and emission of Auger electrons $M_{4.5}VV$ (*4*), obtained by varying the azimuthal angle of incidence of the primary beam.

corded for the quasi-elastically scattered electrons possessed good symmetry. Taking account of the symmetry of the $\langle 100 \rangle$ face of a body-centered cubic crystal, it is possible when measuring low-intensity signals (of the electrons reflected with ionization energy losses and of the Auger electrons) to restrict the range of variation of the angle φ to an interval of 45°. In order to exclude the effect of diffraction of secondary electrons on the examined orientation effect, we collected all of the electrons emitted into the reflection (back) hemisphere.

The accuracy of exposure of the Nb (100) face was better than 1°. The sample surface was cleaned by taking it through several cycles of high-temperature heating in superhigh vacuum with subsequent heating in an oxygen atmosphere at a pressure of $(5-10) \times 10^{-4}$ Pa. Oxygen and carbon surface contamination did not exceed 0.1 monolayer, which was monitored by Auger spectroscopy. The measurements were carried out at room temperature in a vacuum not poorer than 5×10^{-8} Pa.

MEASUREMENT RESULTS

Experimental curves of the azimuthal dependence $I(\varphi)$ for Nb (100), measured at $E_p = 1$ keV are shown in Fig. 1 for four groups of electrons: quasi-elastically scattered electrons, electrons reflected with one-time energy loss due to excitation of plasmons ($\hbar\omega_p = 25$ eV) and due to ionization of the core level $M_{4.5}$ ($E_i = 206$ eV), and finally Auger electrons $M_{4.5}VV$. These dependences are in many ways similar. In all cases a significant amplification of emission for incidence of primary electrons in the crystallographic direction $\langle 110 \rangle$,
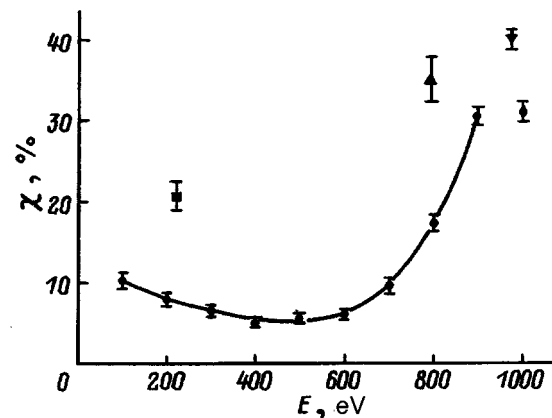

FIG. 2. Values of the parameter χ: ♦ — quasi-elastically scattered primaries, ▼ — electrons reflected with excitation of a plasmon, ▲ — electrons reflected with ionization energy loss, ■ — Auger electrons, and ● — electrons of the continuum.

near the $\langle 133 \rangle$ direction, and parallel to the densely stacked (110) planes is observed. Note that since the disorientation of the beam relative to the $\langle 133 \rangle$ direction did not exceed 2° and the full-widths at half-maximum of the experimental curves was equal to roughly 10°, this factor may to first order be neglected.

Quantitatively, the magnitude of the orientation effect (anisotropy of reflection or emission of Auger electrons) can be estimated with the help of the parameter χ characterizing the degree of focusing of the primary electrons on the ionic framework of the crystal: $\chi = (I_{max} - I_{min})/I_{max}$, where I_{max} and I_{min} are the signal intensity at the considered maximum of the azimuthal dependence and its deepest minimum. The dependence of χ on the energy of the emitted electrons for $E_p = 1$ keV is plotted in Fig. 2. The points, indicated by different symbols, belong to different groups of characteristic electrons. The curve (corresponding to the continuum electrons) reflects the anisotropy of the secondary electrons of the continuum. It is clear from the figure that the highest values of χ are observed for electrons reflected with one-time excitation of a plasmon. The anisotropy of reflection of electrons with ionization energy losses is somewhat lower, but exceeds that observed for quasi-elastic scattering and is also substantially higher than the anisotropy of the background of secondary electrons having the same energy. The lowest degree of focusing among the above groups of electrons is observed for the Auger electrons. In this case the orientation effect for the secondary electrons of the continuum is also much more weakly expressed.

The effect of the energy of the primary electrons on their focusing in the crystal is illustrated by the curves in Fig. 3, which correspond to electrons reflected with ionization energy losses. Similar azimuthal dependences were also observed for the remaining groups of electrons. These data are generalized in Fig. 4, where they are displayed in the form of energy dependences $\chi(E_p)$. It can be seen that the focusing effect is manifested differently for different orientations of the primary beam. The results for the Auger electrons also differ in their specifics. Thus, for the $\langle 110 \rangle$ direction the values of χ for all groups of reflected electrons increase
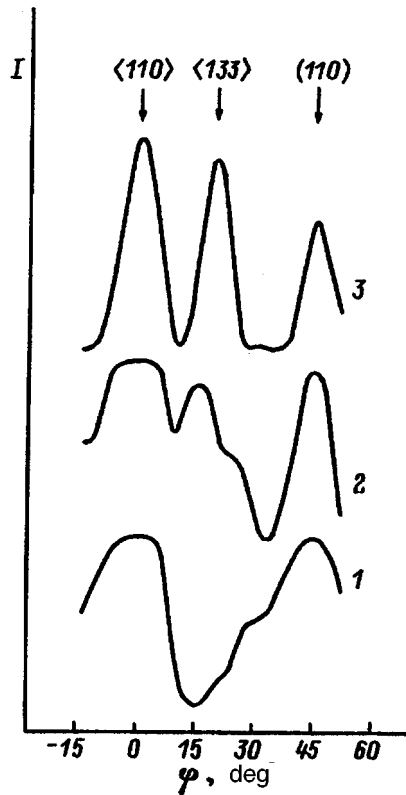
FIG. 3. Effect of primary electron energy on the form of the orientational dependence $I(\varphi)$ for reflection with ionization of the core level $M_{4,5}$ for $E_p = 0.6$ (1), 1 (2), 1.5 keV (3).

FIG. 4. Energy dependence for quasi-elastic scattering (1), reflection with excitation of a plasmon (2), reflection with ionization of the core level (3), and Auger emission (4), characterizing focusing of primary electrons along the $\langle 110 \rangle$ (a) and $\langle 133 \rangle$ (b) directions.

monotonically with E_p, but for the Auger electrons they vary hardly at all or even fall somewhat in the region of energies below 1 keV; only above 1 keV do they grow, and only insignificantly at that. For the electron beam oriented in the $\langle 133 \rangle$ direction focusing is generally not observed for $E_p = 0.6 - 0.7$ keV. It arises and is rapidly amplified only at higher energies. This allows us to speak of the existence of an energy threshold. Finally, for the primary electrons incident in the (110) planes the dependences $\chi(E_p)$ have a special form and exhibit maxima for the above groups of reflected electrons.

DISCUSSION

1) *Quasi-elastic electron scattering (QEES) and inelastic reflection with excitation of plasmons.* The simplest model of quasi-elastic scattering treats the phenomenon as consisting of three stages: 1) small-angle elastic scattering of electrons penetrating into the crystal, which causes a focusing effect; 2) quasi-elastic scattering of electrons at a large angle as a result of their interaction with phonons; 3) subsequent release of electrons from the crystal. The maxima of the orientational dependences are explained within the framework of the model by an increase in the primary electron flux density near the scattering centers of the crystal and growth of the probability of phonon scattering of the electrons as a consequence of the focusing effect arising as a result of the motion of the primary electrons along the densely packed rows of atoms. A departure from such orien-

tations is accompanied by a weakening of focusing and a decrease of the probability of quasi-elastic scattering.

Inelastic reflection of electrons with excitation of plasmons brings one more elementary act into the picture, which may take place both before and after quasi-elastic scattering into the back hemisphere. Since excitation of plasmons occurs as a result of the long-range Coulomb interaction with the electronic subsystem of the crystal, their probability of generation is essentially insensitive to electron focusing. Besides, since the electron has the greatest probability of exciting long-wavelength plasmons with small momentum, no noticeable change in the momentum of the primary electron takes place in this case. As a result, acts of plasmon generation disturb the process of electron focusing only slightly, and differences in the orientation effects for the given group of electrons as well as for the quasi-elastically scattered electrons arise mainly as a consequence of differences in their mean depths of emergence z. Since this magnitude is roughly twice as large for the electrons reflected with excitation of plasmons, the relation between the corresponding values of χ is determined by whether focusing of the primary electrons at depths greater than the mean depth of emergence of the quasi-elastically scattered electrons is amplified or attenuated.

Basing ourselves on the above model, let us estimate the lengths of the focusing chains for the $\langle 110 \rangle$ and $\langle 133 \rangle$ directions, the interatomic distances d for which are equal to 4.4 and 7.2 Å, respectively. The path traversed by a quasi-

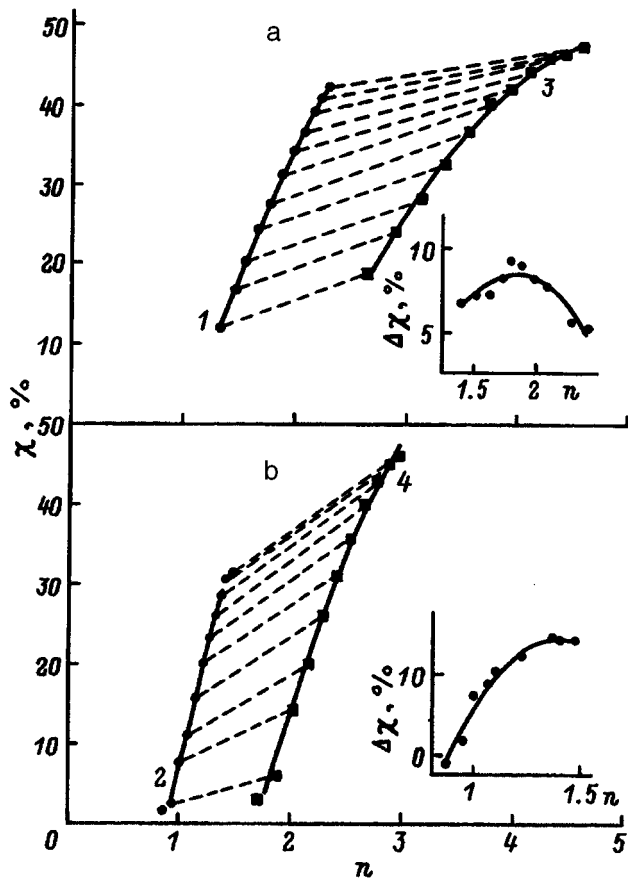FIG. 5. Dependence $\chi(n)$ of the degree of focusing of the primary electrons in the ion core of the crystal on the mean number of atoms (equal to $n+1$) with which the electrons interact as they move along the $\langle 110 \rangle$ (a) and $\langle 133 \rangle$ (b) directions. The graphs shown in the insets plot the dependence $\Delta\chi(n)$ demonstrating the variation of the absolute increment of the degree of focusing going from the quasi-elastically scattered electrons to the electrons reflected with excitation of a plasmon.

elastically scattered electron in the crystal is determined by the mean free path λ of the electron relative to the inelastic interaction and consists of two parts: the path l of the electron as it penetrates into the crystal, and the path L traversed by it as it moves back up toward the surface. The sum of these two pathlengths is equal to λ. In our case of complete collection of the reflected electrons for the incident polar angle equal to 45°, we may take $l \cong L$. Thus, knowing the values of d and λ (Ref. 31), we can express the lengths of the focusing chains l in terms of the number of interatomic distances traversed on average by the electron up to the act of quasi-elastic scattering: $n = \lambda/2d$. This allows us to map the curves $\chi(E_p)$ onto the curves $\chi(n)$ shown in Fig. 5 (curves 1 and 2). Figure 5 also shows the analogous curves for the case of reflection with excitation of a plasmon (curves 3 and 4). In the case of quasi-elastic scattering only a small number of atoms take part in the focusing process. For the $\langle 110 \rangle$ direction the values of n on average do not exceed three, and for the $\langle 133 \rangle$ direction $n<2$. Hence the reason for the appearance of the electron focusing threshold, observed for the given direction, becomes clear. Indeed, in order for focusing of an electron as it scatters from a surface atom to have an effect on the probability of quasi-elastic scattering

by an atom of the second layer, it is necessary that the observed electron should have at least been able to reach it, i.e., the relation $n>1$ must be fulfilled. As can be seen from Fig. 5, the critical point $n=1$ for the direction $\langle 133 \rangle$ lies in the considered energy range. In this case, for $n<1$ the values of χ are near zero while in the region $n=1-1.4$ an abrupt linear growth of χ to the value 0.33 is observed for $E_p=1.5$ keV. Note that the given ''crystallographic'' threshold differs in its nature from the well-known ''energy'' threshold due to the mechanism of scattering of the electrons by the atoms.[32]

For the $\langle 110 \rangle$ direction the value of n significantly exceeds unity for $E_p=0.6$ keV and the corresponding threshold is not observed. However, in this case the efficiency of focusing grows significantly with growth of n although not as rapidly as for the the $\langle 133 \rangle$ direction. In both cases this growth is due to forward elongation of the electron–atom scattering diagram with increasing E_p and overall amplification of forward scattering in comparison with backscattering, which diminishes the role of quasi-elastic scattering from the upper monolayer, which does not contribute to the orientation effect. In the region $n>2$ ($l>9.2$ Å) growth of focusing is weakened and $\chi(n)$ is observed to deviate from a linear dependence, which indicates the onset of electron defocusing, which, as is well known, is most strongly manifested for the most densely packed directions.

For the electrons reflected with excitation of a plasmon, the dependence $\chi(n)$ for the $\langle 133 \rangle$ direction also grows linearly with n at first (Fig. 5, curve 4). Then, at around $n \simeq 2.5$ ($l=18$ Å) its growth is observed to slow down. The reason for this slowing down must also be ascribed to the onset of defocusing of the electrons propagating in this direction. The given value of l is considerably larger than that obtained for QEES for beams oriented along the $\langle 110 \rangle$ direction. This indicates that in the case of atomic chains with low packing density defocusing processes come into play only at large distances. For the $\langle 110 \rangle$ direction the dependence $\chi(n)$ can be traced out only starting from $n=2.7$ (Fig. 5, curve 3). It has essentially no linear segment, which indicates the combined occurrence of processes of primary electron focusing and defocusing in the given range of scattering chain lengths, in agreement with the above-described results.

Let us now consider how the orientation effect is affected by the twofold increase in the depth of emergence of the electrons observed for a fixed energy in going from the quasi-elastically scattered electrons to the electrons reflected with excitation of a plasmon. For greater ease of visualization in such a comparison the corresponding points in Fig. 5 are connected by dashed lines. It can be seen that lengthening of the atomic chains is always accompanied by a growth of χ, demonstrating that in all cases focusing dominates over defocusing, in particular for the $\langle 110 \rangle$ direction, at least for $n=5$. Amplification of this effect is manifested in different ways for rows of atoms with different packing density. Thus, an absolute increase in the anisotropy for the $\langle 133 \rangle$ direction is observed up to values of E_p greater than for the $\langle 110 \rangle$ direction (Fig. 5, insets), so that the values of χ for $E_p=1.5$ keV for the atomic chains are equalized to a significant degree. This means that the focusing properties of two different

M. V. Gomoyunova and I. I. Pronin   964

atomic chains of the same length, but consisting of a different number of atoms, turn out to be very similar.

To conclude this section, let us dwell briefly on the salient features of quasi-elastic scattering of electrons penetrating into a crystal parallel to the densely stacked (110) planes. In this case, for $\Theta = 45°$ the electrons scattered forward by the atoms of the upper layer pass in the immediate vicinity of only one other atom of the chain and then move a significant distance between chains. The high values of χ observed in this case already for $E_p = 0.6$ keV are apparently due to quasi-elastic scattering by their nearest neighbors, of the primary electrons focused by the surface atoms. Amplification of the orientation effect for energies up to 1 keV can be explained by the already noted peculiarities of the electron–atom differential scattering cross sections, and the subsequent falloff—by an increase in the contribution from the region of electron propagation in the space between the chains.

2) Inelastic reflection of electrons with ionization energy loss, and Auger electron emission. Excitation of a core electron by a primary electron with transfer of the binding energy E_i, which is considerably smaller than E_p, leads to scattering of the primary electron through only a small angle. Reflection of electrons with ionization energy loss under such conditions may be treated in a manner similar to inelastic reflection with excitation of plasmons, where this excitation comprises a process of primary electron focusing, ionization of a core level, and scattering through a large angle. However, in contrast to generation of plasmons, the probability of excitation of a core electron depends on the degree of overlap of its wave function with the wave function of the primary electron and, consequently, should be, like quasi-elastic scattering through a large angle, sensitive to electron focusing. In this case, the influence of the anisotropy of phonon scattering of electrons is realized mainly upon reflection, when scattering through a large angle precedes ionization. The anisotropy of ionization of a core level is manifested mainly through the second component of reflection, when excitation of a core electron outstrips the changing of direction of motion of the primary electron. What form of anisotropy prevails in the orientation effect can be determined by comparing these results with those obtained for reflection with excitation of plasmons. It is clear from Fig. 4 that for $E_p = 0.6$ keV the values of χ for them are roughly equal. Noting that the mean depths of emergence of both groups of reflected electrons are similar, we may conclude that the anisotropy of ionization of the core level $M_{4.5}$ is roughly the same as for phonon scattering through a large angle, although with increasing energy the latter becomes somewhat higher. Thus, despite the fact that the loss itself is sensitive to the nature of the ionized atoms, the effect of primary electron focusing on the probability of quasi-elastic scattering (averaged over the entire zone of emergence of the electrons) hinders the use of the orientation effect of the considered group of electrons as a new method of structural analysis of surfaces possessing elemental sensitivity.

Focusing of primary electrons, which is the cause of the anisotropy of ionization of the core levels, is also manifested in the orientation effect of emission of Auger electrons generated during their excitation. As was noted above, the azimuthal dependences for the $M_{4.5}VV$ Auger electrons are qualitatively similar to those observed for the reflected electrons. Differences between them are revealed by comparing the energy dependences $\chi(E_p)$, especially for the $\langle 110 \rangle$ direction (Fig. 4a). To explain the specifics of the orientation effect of Auger emission, it is necessary to take into account that there exist two sources of excitation of Auger electrons: first, the primary electrons propagating into the depths of the crystal and, second, the inelastically reflected and fast true-secondary electrons, also passing through the emergence zone of the Auger electrons as they move toward the surface. It should be stressed that in all the applications of the Auger-emission orientation effect to surface structural analysis known to us the role of the backward flux has not been considered although this question has been analyzed in connection with the problem of quantitative Auger analysis of crystals.[33] At the same time, as is evident from Fig. 2, the intensity of the backward flux also depends on the orientation of the primary beam and this factor must be taken into account in an analysis of the angular dependence of the Auger electrons. Growth of their emission is due not only to the influence of primary electron focusing on ionization of the core levels, but also to an increase in the backward electron flux intensity due to it.

The variation with growth of the primary energy E_p of the contributions of the forward and backward fluxes to generation of the Auger signal leads to a redistribution of the roles of these sources in the formation of the Auger-emission anisotropy. As is well known, the greatest probability of ionization of the core electrons with binding energy E_i is realized for the primary energy $E_p = (2-3) \cdot E_i$. In our case, this condition is fulfilled for primary electrons with energy 0.6 keV. The efficiency of the backward flux is considerably lower, and since the total intensity of secondary electrons capable of exciting Auger electrons does not exceed 20% of the intensity of the primaries, the overwhelming fraction of the Auger electrons (in any case more than 80%) are excited by the forward flux. As for the anisotropy of excitation of the Auger electrons, for the backward flux it is also lower than for the primary electrons since as the energy losses grow the anisotropy decreases (Fig. 2). In sum, for $E_p = 0.6$ keV the orientation effect is determined almost entirely by the influence of primary electron focusing on the ionization of the core level $M_{4.5}$ and, consequently, can be the basis of a structural analysis of surfaces that is sensitive to the nature of their atoms.

With growth of the energy of the primary electrons the probability of excitation of Auger electrons falls. At the same time, the intensity of the backward flux of electrons capable of generating Auger electrons grows. The orientational dependence of the backward flux intensity also increases somewhat according to our data. In sum, the contribution of the forward flux to the anisotropy of Auger electron generation falls as E_p increases, while the contribution of the backward flux, on the contrary, increases, which leads to a compensation effect. This explains the above-noted weakness of the initial segment of the experimentally observed dependence $\chi(E_p)$ for the $\langle 110 \rangle$ direction (Fig.

4a). The thickness of the near-surface layer is no longer determined only by the depth of emergence of the Auger electrons and can be noticeably larger since it now also depends on the backward flux. Therefore the Auger orientation effect at larger E_p partly loses its sensitivity to the nature of the atoms emitting the Auger electrons and now depends on the total anisotropy of electron reflection by the investigated object, which complicates the structural analysis.

CONCLUSION

1. On the basis of the dependence of the intensity of quasi-elastic electron scattering and inelastic reflection with excitation of plasmons on the azimuthal angle of incidence of the primary beam, we have analyzed the focusing properties of atomic chains with widely different packing density. We have shown that for the primary beam oriented along the densely packed $\langle 110 \rangle$ direction, for which we observed the greatest reflection anisotropy, focusing processes dominate, at least for chains consisting of six atoms. The effect of defocusing becomes noticeable for chains constructed from three or more atoms. For the $\langle 133 \rangle$ direction, we have discovered a ''crystallographic'' threshold of focusing, determined by the relation between the interatomic distance on the chain and the length of the mean free path of an electron. Differences in the focusing properties of different chains decrease with growth of the electron energy (with increase of their length).

2. We have shown that the orientation effect of electron reflection with ionization energy losses $M_{4,5}$ is controlled both by anisotropy of ionization of the shell level itself and by anisotropy of quasi-elastic scattering of primary electrons through large angles. The latter circumstance hinders the use of the given orientation effect as a new element-sensitive method of analysis of the local atomic structure of surfaces. The azimuthal dependences of emission of Auger electrons excited during ionization of the same shell level are qualitatively similar to those observed for the reflected electrons. However, there are fundamental differences in the trend of their variation with energy for the densely packed $\langle 110 \rangle$ direction. The specifics of the Auger orientation effect are due to the presence of two sources of Auger electron generation: the forward and backward electron fluxes. Only at the lowest primary electron energies is the role of the backward flux unimportant, where the effect is governed only by the anisotropic nature of ionization of the shell levels, which allows it to be used for an element-sensitive structural analysis of the thin near-surface layer of crystals. For kilovolt primary electrons, their focusing has an effect on the total backscattering intensity, which complicates the interpretation of the experimental data.

[1] W. F. Egelhoff, Jr., Crit. Rev. Solid State Mater. Sci. **16**(3), 213 (1990).
[2] C. S. Fadley, *Synchrotron Radiation Research: Advances in Surface and Interface Science*, Vol. 1, *Techniques*, edited by R. Z. Bachrach (Plenum, New York, 1992), Ch. 9, pp. 421–518.
[3] S. A. Chambers, Surf. Sci. Rep. **16**(6), 261 (1992).
[4] N. Nakamura, K. Anno, and S. Kono, Surf. Sci. Lett. **262**(3), L101 (1992).
[5] J.-M. Pan, B. L. Machhoff, U. Diebold *et al.*, Surf. Sci. **291**(2), 381 (1993).
[6] M. Erbudak, M. Hochstrasser, T. Schulthess *et al.*, Philos. Mag. Lett. **68** (3), 179 (1993).
[7] M. V. Gomoyunova, I. I. Pronin, N. S. Faradzhev *et al.*, Fiz. Tverd. Tela **36**, 2295 (1994) [Phys. Solid State **36**, 1250 (1994)].
[8] Y. Gao and J. Cao, Phys. Rev. B **43**, 9692 (1991).
[9] Y. N. Idzerda and G. A. Prinz, Phys. Rev. B **43**, 11460 (1991).
[10] S. Mroz and M. Nowicki, Surf. Sci. **297**(1), 66 (1993).
[11] S. Valeri, A. di Bona, and G. C. Cazzadi, Phys. Rev. B **50**, 14617 (1994).
[12] M. V. Gomoyunova and I. I. Pronin, Poverkhnost', No. 12, 92 (1995).
[13] A. B. Laponsky and N. R. Whetten, Phys. Rev. Lett. **3**, 510 (1959).
[14] R. W. Soshea and A. J. Dekker, Phys. Rev. **121**, 1362 (1961).
[15] A. I. Abroyan and Timov, Fiz. Tverd. Tela **9**, 3628 (1967) [Sov. Phys. Solid State **9**, 2860 (1967)].
[16] R. M. Stern and H. Taub, Phys. Rev. Lett. **20**, 1340 (1968).
[17] A. R. Shchul'man, V. V. Korablev, and Yu. A. Morozov, Fiz. Tverd. Tela **10**, 1913 (1968) [Sov. Phys. Solid State **10**, 1512 (1968)].
[18] H. Taub, R. M. Stern, and V. F. Dvoryankin, Phys. Status Solidi **33** (2), 573 (1969).
[19] G. Allie, E. Blanc, and D. Dufayard, Surf. Sci. **43**(2), 545 (1974).
[20] M. V. Gomoyunova, S. L. Zaslavskiĭ, and I. I. Pronin, Fiz. Tverd. Tela **20**, 3645 (1978) [Sov. Phys. Solid State **20**, 2106 (1978)].
[21] S. K. Andersen and A. Howie, Surf. Sci. **50**(1), 197 (1975).
[22] M. V. Gomoyunova, O. V. Konstantinov, and I. A. Shmulevich, Izv. Akad. Nauk SSSR. Ser. Fiz. **46**, 2308 (1982).
[23] S. A. Chambers, I. M. Vitomirov, and J. H. Weaver, Phys. Rev. B **36**, 3007 (1987).
[24] S. Y. Tong, H. C. Poon, and D. R. Snider, Phys. Rev. B **32**, 2096 (1985).
[25] W. F. Engelhoff, Jr., Phys. Rev. Lett. **59**, 559 (1987).
[26] M. L. Xu, J. J. Barton, and M. A. Van Hove, Phys. Rev. B **39**, 8275 (1980).
[27] A. P. Kaduwela, D. J. Friedman, and C. S. Fadley, J. Electron Spectrosc. Relat. Phenom. **57**(2), 223 (1992).
[28] S. Hüfner, J. Osterwalder, T. Greber *et al.*, Phys. Rev. B **42**, 7350 (1990).
[29] V. Seelmann-Eggebert, R. Fasel, E. C. Larkins *et al.*, Phys. Rev. B **48**, 11838 (1993).
[30] M. Erbudak, T. Schulthess, and E. Wetli, Phys. Rev. B **49**, 6316 (1994).
[31] M. P. Seach, Surf. Interface Anal. **2**(1), 85 (1986).
[32] H. C. Poon and S. Y. Tong, Phys. Rev. B **30**, 6211 (1984).
[33] M. V. Gomoyunova, S. L. Dudarev, and I. I. Pronin, Surf. Sci. **235**(1), 156 (1990).

Translated by Paul F. Schippnick

Precise estimates of modulational reconstruction of the response function and its derivatives

N. D. Kuz'michev

N. P. Ogarev Mordovian State University, 430000 Saransk, Russia

Formulas are obtained which estimate the errors arising in the modulational reconstruction of the response function and its first and second derivatives in the hysteresis-free case. The algorithm used can yield formulas for estimating the errors in the presence of hysteresis. The results presented can be used to recover the magnetization and the $I-V$ characteristics of high-temperature superconductors from the amplitude spectrum of the harmonics of the response signal obtained by subjecting the sample to modulated magnetic field and current. © *1997 American Institute of Physics.* [S1063-7842(97)02208-3]

Recently, in studies of the magnetic properties and also the $I-V$ characteristics[1] of high-temperature superconductors (HTSC) attention has been given to the harmonic spectrum of the response signal obtained by subjecting the sample to a modulated magnetic field or current (see Refs. 1–6, etc.). The response signal here is a periodic function of time and by virtue of nonlinearity and hysteresis of the properties (e.g., magnetic) of the sample has a complicated (nonsinusoidal) shape. Even for small modulation amplitudes the response spectrum contains higher harmonics. A study of the nonlinearity of physical properties in many problems of experimental physics is of special interest. Thus, the problem arises of reconstructing initial dependences. This problem can be solved with difficulty if the nonlinear part of the dependence in question is manifested against a significant linear background. An additional difficulty arises as a result of hysteresis. The initial dependence (response function) reconstructed from the spectrum of amplitudes of the higher harmonics of the response signal is of greater value than the directly measured dependence. This is because the higher harmonics contain rich information about the analytic properties of the response function and its derivative.

In many cases the well-known modulation technique[7] allows one to get around the above problem, but it was developed for the case of small modulation amplitudes in the absence of hysteresis in the investigated quantity. In Refs. 8–10, I attempted to generalize the indicated technique to the case of derivatives of the modulation amplitudes with allowance for hysteresis with the help of Taylor and Fourier series and obtained formulas for reconstructing the initial dependences.

In the present paper I estimate the errors arising in a reconstruction of the response function using the modulation technique.

We denote the modulating influence (magnetic field, current, etc.) by $x = x_0 + a \cos(\omega t)$. Here x_0 is the static or slowly varying influence, a is the modulation amplitude, ω is the angular frequency, and t is time. We denote the investigated dependence (response function) as Y. The response function will be a periodic function of time. The Fourier series for the hysteresis response function Y will have the form

$$Y(x_0 + a \cos \omega t) = \frac{Y'_0}{2} + \sum_{n=1}^{\infty} [Y'_n \cos(n\omega t) + Y''_n \sin(n\omega t)]. \tag{1}$$

The Fourier coefficients $Y'_n(x_0, a)$ and $Y''_n(x_0, a)$ (amplitudes of the real (in-phase) and imaginary (quadrature) parts of the harmonics of the response function) are given by

$$Y'_n = \frac{1}{\pi} \int_0^{2\pi} \{[Y_-(x_0 + a \cos \omega t) + Y_+ \\ \times (x_0 + a \cos \omega t)]/2\} \cos(n\omega t) d(\omega t), \tag{2}$$

$$Y''_n = \frac{1}{\pi} \int_0^{\pi} [Y_-(x_0 + a \cos \omega t) - Y_+ \\ \times (x_0 + a \cos \omega t)] \sin(n\omega t) d(\omega t). \tag{3}$$

Here Y_+ is the branch of the function Y for growing x, and Y is the branch for falling x. The following formulas were obtained in Ref. 10 for reconstructing the response function and its first and second derivatives:

$$\overline{Y}(x_0) = [Y_-(x_0) + Y_+(x_0)]/2 = (Y_0/2) \\ + \sum_{n=1}^{\infty} (-1)^n Y'_{2n}(x_0, a), \tag{4}$$

$$\Delta Y(x_0) = Y_-(x_0) - Y_+(x_0) = \sum_{n=1}^{\infty} (-1)^n Y''_{2n+1}(x_0, a), \tag{5}$$

$$\frac{d}{dx} Y(\overline{x}_0) = \frac{1}{a} \sum_{n=1}^{\infty} (-1)^{n+1}(2n-1) Y'_{2n-1}(x_0, a), \tag{6}$$

$$\frac{d}{dx} \Delta Y(x_0) = \frac{2}{a} \sum_{n=1}^{\infty} (-1)^{n+1}(2n) Y''_{2n}(x_0, a), \tag{7}$$

$$\frac{d^2}{dx^2} Y(\overline{x}_0) = \frac{1}{a^2} \sum_{n=1}^{\infty} (-1)^{n+1}(2n)^2 Y'_{2n}(x_0, a), \tag{8}$$

$$\frac{d^2}{dx^2}\Delta Y(x_0) = \frac{2}{a^2}\sum_{n=1}^{\infty}(-1)^n(2n-1)^2 Y''_{2n-1}(x_0,a).$$

(9)

Formulas (4), (6), and (8) are valid in the absence of hysteresis in $Y(x)$ since in this case $Y=\bar{Y}$ and $\Delta Y=0$.

The indicated numerical technique for reconstructing the initial dependences requires measurements of the dependence of the amplitudes of the harmonics Y_n of the response function on the static influence x_0 for a give modulation amplitude, which is not always feasible. In some cases it is easier to measure the dependence of the harmonics Y_n on the modulation amplitude for $x_0=0$. For this case, in the absence of hysteresis the reconstruction formulas are

$$Y(a) = Y_0(a)/2 + \sum_{n=1}^{\infty}Y_n(a),$$

$$dY/da = (1/a)\sum_{n=1}^{\infty}n^2 Y_n(a).$$

(10)

In practice in the reconstruction of the response function with the help of formulas (4)–(10) it is necessary to use finite sums instead of series and to allow for experimental measurement errors, particular measurement noise defined by the signal-to-noise ratio. In this case, to estimate the accuracy of reconstruction of the initial dependence and its derivatives it is necessary to use the criterion of series convergence [Ref. 11, p. 426], in particular for the Fourier series (1), and in some cases it is not necessary to take into account a large number of terms of the series (strong convergence).

In many experimental situations the response function is a voltage U (emf) on a sensor which is directly proportional to the dependence in question Y or its first derivative dY/dx, for example, the $I-V$ characteristic or the differential magnetic susceptibility dM/dH (M is the magnetization).[1–6] Taking the above said into account, we may write

$$U(x) = CY(x).$$

(11)

Here C is the instrument constant. In the case of $I-V$ characteristics $C=1$ and $x\equiv I$ (where I is the current). Note that the amplitudes U_n of the harmonics of the response signal are directly proportional to Y_n.[1,2,4,8,9] We write this convergence criterion, by way of example, for series (6) in the absence of hysteresis in Y, i.e., for $Y=\bar{Y}$, as

$$\left|a\frac{dU}{dx} - \sum_{n=1}^{N}(-1)^n(2n-1)U_{2n-1}\right| < \delta U,$$

(12)

where a is the modulation amplitude, Y is the true initial dependence, δU is the voltage defining the accuracy of reconstruction, and N is the number of odd experimental points U_k.

The number N is bounded by U_{noi} or Δu_k, where U_{noi} is the rms noise voltage. The error Δu_k of the voltage measurements U_k is determined by U_{noi}. For all $k \leqslant N$ $U_k > U_{noi}$ or Δu_k.

The second term on the left-hand side of inequality (12) is

$$Ca\frac{dY^*}{dx} = \sum_{n=1}^{N}(-1)^n(2n-1)U_{2n-1}.$$

As our final estimate of the accuracy of reconstruction we have

$$\left|\frac{dY}{dx} - \frac{dY^*}{dx}\right| \lesssim \frac{\delta U}{|C|a},$$

(13)

i.e., the true value of the derivative is included within the limits

$$\frac{dY}{dx} = \frac{dY^*}{dx} \pm \frac{\delta U}{|C|a}.$$

Here dY^*/dx is the approximately reconstructed first derivative. Analogous estimates may be obtained for quantities (4), (5), and (7)–(10). Note that for prescribed values of U_{noi} or Δu_k the number N (the number of observed harmonics) is proportional to the modulation amplitude and the degree of nonlinearity of the initial dependence. The degree of nonlinearity is determined by the number of dominant terms of the Taylor series for Y. A high degree of nonlinearity and the presence of hysteresis result in slow convergence of series (1).

Now let us estimate δU. Assume that the measurement error has the same value Δu for all the harmonics. For the sum in inequality (12) the error is equal to [Ref. 12, p. 598]

$$\sqrt{\sum_{n=1}^{N}(2n-1)^2\Delta u^2} = \Delta u\sqrt{N(2N-1)(2N+1)/3}.$$

This means that the exact value of the sum in inequality (12) lies within the limits $Ca(dY^*/dx) \pm \Delta u\sqrt{N(2N-1)(2N+1)/3}$. Now only the residual term R_N in series (6) remains to be taken into account, which is smaller in absolute value than U_0/N, i.e.,

$$|R_N| < U_0/N.$$

(14)

Here U_0 is some positive value of the voltage which we will estimate below. In fact, since for absolute convergence of the residual term $\sum_{n=N+1}^{\infty}(-1)^n(2n-1)U^t_{2n-1}$ it is necessary that the true values U^t_{2n-1} (not to be confused with the experimental values U_{2n-1}) have the estimate (assuming the existence of the second derivative d^2U/dx^2, satisfying the Dirichlet conditions) $|U^t_{2n-1}| < U_0/(2n-1)^3$ [Ref. 13, p.505; Ref. 14, p.495], we obtain

$$|R_N| = \left|\sum_{n=N+1}^{\infty}(-1)^n(2n-1)U^t_{2n-1}\right| < U_0$$

$$\times \sum_{n=N+1}^{\infty}\frac{1}{(2n-1)^2} < U_0\sum_{n=N+1}^{\infty}\frac{1}{n^2} < \frac{U_0}{N}.$$

As U_0 we may take, for example, the maximum absolute value of the set of experimental amplitudes of the harmonics U_{2n-1}, i.e., $U_0 = \max|U_{2n-1}|$. Here $n=1,2,3,\ldots$. Then the accuracy of reconstruction is equal to

$$\delta U = \sqrt{[N(2N-1)(2N+1)/3]\Delta u^2 + (U_0/N)^2}. \quad (15)$$

The quantity $\delta U(N)$ has a minimum, and the value N_{opt}, at which the minimum of δU occurs is found by solving the fifth-order algebraic equation

$$(4N^2 - 1/3)\Delta u^2 = 2U_0^2/N^3$$

or, neglecting the 1/3 term on the left-hand side of the equation, we obtain

$$N_{opt} \approx 2^{-1/5}[U_0/(\Delta u)]^{2/5} \approx 0.87(U_0/\Delta u)^{2/5}. \quad (16)$$

For example, for $U_0/\Delta u = 10$ (10% accuracy) we obtain $N \approx 2.18$, i.e., $N_{opt} = 2$. This means it is necessary to take the amplitudes of the first and third harmonics.

The error of reconstruction of the first derivative of the response function is given by

$$\delta(dY/dx) = \frac{\delta U}{|C|a}$$
$$= \sqrt{[N(2N-1)(2N+1)/3]\Delta u^2 + (U_0/N)^2}. \quad (17)$$

In the reconstruction of the response function itself Y we obtain the following estimates:

$$CY^* = U_0/2 + \sum_{n=1}^{N} (-1)^n U_{2n},$$

$$\sqrt{\sum_{n=0}^{N} \Delta u^2} = \Delta u \sqrt{N+1},$$

$$\delta u = \sqrt{(N+1)\Delta u^2 + [U_0/(2N)]^2}, \quad (18)$$

$$N_{opt} = 2^{-1/3}(U_0/\Delta u)^{2/3} \approx 0.8(U_0/\Delta u)^{2/3}. \quad (19)$$

Here $U_0 = \max|U_{2n}|$, $n = 0, 1, 2, \dots$.

In the reconstruction of the second derivative of the response function d^2Y/dx^2 we obtain the following estimates:

$$a^2C(d^2Y^*/dx^2) = \sum_{n=1}^{N} (-1)^{n+1}(2n)^2 U_{2n},$$

$$\sqrt{\sum_{n=1}^{N} (2n)^4 \Delta u^2} = \frac{4\Delta u}{(30)^{1/2}}$$
$$\times \sqrt{N(N+1)(2N+1)(3N^2+3N-1)},$$

$$\delta U = \sqrt{(8/15)N(N+1)(2N+1)(3N^2+3N-1)\Delta u^2 + [U_0/(2N)]^2}. \quad (20)$$

The quantity N_{opt} is found by solving a seventh-order algebraic equation. N_{opt} may be approximately estimated by the formula

$$N_{opt} \approx (32)^{-1/7}\left(\frac{U_0}{\Delta u}\right)^{2/7} \approx 0.61(U_0/\Delta u)^{2/7}. \quad (21)$$

Note that in the reconstruction of the response function we assumed the existence of its first derivative, and in the reconstruction of the second derivative we assumed the existence of the third derivative, both satisfying respective Dirichlet conditions.

As can be seen, for example, from expression (17), the quantity $\delta(dY/dx)$ (we will denote it by $\delta_1 = \delta(dY/dx)$) also depends on the modulation amplitude inverse-proportionally and through U_0. As the modulation amplitude grows, the first term in the radicand decreases and the second one grows. The rate of growth of $(U_0/N)^2$ is determined by U_0. Thus, there is an optimal value of the modulation amplitude a_{opt}. The quantity a_{ext} (extreme value) can be found by solving the approximate equation

$$a \approx 8U_0/[3(dU_0/da)]. \quad (22)$$

In order to find the minimum, it is necessary in addition to follow δ_1 to its minimum. In the reconstruction of the response function Ya_{ext} is found by solving the equation

$$U_0(dU_0/dx) \approx 0. \quad (23)$$

Equations (22) and (23) can be used in the case when, for example, the amplitude of the kth harmonic of $U(0)$ remains maximal upon variation of the modulation amplitude and the analytic or numerical form of $U_0(a)$ has been determined.

If the indices of the maximum amplitudes change upon variation of the modulation amplitude, then Eqs. (22) and (23) remain valid, but the analytic form of $U_0(a)$ will be different in different regions of variation of the amplitude.

Under conditions in which a small number of harmonics are observed, e.g., N of the order of 3–5, the error of reconstruction is mainly determined by the first term of the radicand in Eq. (17), i.e., by the term containing Δu. The reason for this lies in the rapid convergence of series (6), whose residual term (14) is much smaller than U_0/N, i.e., $|R_N| \ll U_0/N$. In general, the residual term of the series is estimated by $|R_N| \propto U_0/N^m$. Here the exponent m may be of the order of 10 or greater. Analogous estimates can be also obtained for formulas (5), (7), (9), and (10) with the help of the given algorithm.

To summarize, it has been found that with the help of a modulational measurement and reconstruction technique it is possible to determine the numerical form of the response function and its derivatives (at least the first and second). In addition, it is possible to determine the error of reconstruction of the indicated function and its derivatives.

[1] The dependence of the voltage response V on the current I, i.e., $V(I)$.

---

[1] C. Jeffries, Q. Lam, and Y. Kim et al., Phys. Rev. B **37**, 9840 (1988).
[2] J. Sun, M. Scharen, and L. Bourne et al., Phys. Rev. B **44**, 5275 (1991).
[3] W. Xing, B. Heinrich, and J. Chrzanowski et al., Physica C **205**, 311 (1993).
[4] M. A. Vasyutin and N. D. Kuz'michev, Pis'ma Zh. Tekh. Fiz. **18**(23), 5 (1992) [Tech. Phys. Lett. **18**, 764 (1992)].
[5] A. I. Golovashkin, N. D. Kuz'michev, and I. S. Levchenko et al., Fiz. Tverd. Tela **31**, 233 (1989) [Sov. Phys. Solid State **31**, 679 (1989)].
[6] N. D. Kuz'michev and M. A. Vasyutin, Sverkhprovod. Fiz. Khim. Tekh. **7**(1), 93 (1994).
[7] L. Solymar, *Superconductive Tunnelling and Applications* (Chapman and Hall, London, 1972).

[8] N. D. Kuz'michev, Pis'ma Zh. Tekh. Fiz. **17**(7), 56 (1991) [Tech. Phys. Lett. **17**, 259 (1991)].

[9] N. D. Kuz'michev, Zh. Tekh. Fiz. **64**, 63 (1994) [Tech. Phys. **39**, 1236 (1994)].

[10] N. D. Kuz'michev, Pis'ma Zh. Tekh. Fiz. **20**(22), 39 (1994) [Tech. Phys. Lett. **20**, 906 (1994)].

[11] G. M. Fikhtengol'ts, *Course in Differential and Integral Calculus* [in Russian], Nauka, Moscow (1970), Vol. 2.

[12] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series* [in Russian], Nauka, Moscow (1981).

[13] G. M. Fikhtengol'ts, *Course in Differential and Integral Calculus* [in Russian], Nauka, Moscow (1970), Vol. 3.

[14] V. I. Smirnov, *Course in Higher Mathematics* [in Russian], Nauka, Moscow (1974), Vol. 2.

Translated by Paul F. Schippnick

BRIEF COMMUNICATIONS

Measurement of the thermal diffusivity of the surface layers of opaque solid objects

V. I. Turinov

A method is discussed for the layer-by-layer determination of the thermal diffusivity a and thickness d of the subsurface layers of opaque solid objects from measurements of the phase difference of signals of two centered ring-shaped $p-n$ photodiode junctions by successively varying the coefficient of angular magnification of the optical system of the device in which a radiation source with Gaussian cross-sectional beam power density creates a thermal object on the surface of the object in the form of harmonically varying concentric heat waves. Estimates are found for the frequency range and minimum angular magnification corresponding to the limits of validity of the relations used to calculate a and d. © *1997 American Institute of Physics.* [S1063-7842(97)02308-8]

In the development of the method for measuring the thermal diffusivity according to the scheme laid out in Ref. 1, let us consider the thermal problem arising in the irradiation of a sample by focused radiation varying according to the harmonic law $Q = Q_1(1/2)(1+m\cos\omega\tau)$, where Q_1 is the radiation power density (W/cm^2), m is the modulation depth coefficient, $\omega = 2\pi f$ is the radiation frequency, and τ is time. For simplicity we set $m=1$. As a result of the irradiation of the surface of the sample, a heat wave (thermal object) is formed, propagating symmetrically outward from the heated spot. Let the distribution Q_1 over the beam cross section be described by a Gaussian law $Q_1 = Q_0\exp(-\rho^2/2\rho_0^2)$ as is characteristic of laser beams, electron beams, plasma jets, and a number of other sources.

For a semi-infinite object (under the condition of cooling due to heat conduction) the problem of surface heat waves in polar coordinates has the form

$$\frac{\partial T}{\partial \tau} = a\frac{\partial^2 T}{\partial \rho^2} + \frac{a}{\rho}\frac{\partial T}{\partial \rho}, \quad 0 < \tau < \infty, \tag{1}$$

$$T - t_0 \to 0 \quad \text{for } \rho \to \infty,$$

$$\frac{\partial(T-t_0)}{\partial \rho} = \frac{Q_0}{\lambda}\exp(-\rho^2/2\rho_0^2)\cos\,\omega\tau, \tag{2}$$

where a is the thermal diffusivity, λ is the thermal conductivity, and t_0 is the surface temperature of the object in the limit $\rho \to \infty$.

Taking the Laplace transform in τ with the parameter s in Eqs. (1) and (2), we obtain[2]

$$T_L - \frac{t_0}{s} = \frac{Q_1}{\lambda}\frac{sI_0\left(\sqrt{\frac{s}{a}}\rho\right)}{(s^2+\omega^2)I_1\left(\sqrt{\frac{s}{a}}\rho_0\right)}, \tag{3}$$

where $I_\nu(\mu)$ is the modified Bessel function.

Applying the decomposition theory to Eq. (3) (which has simple roots $s = \pm i\omega$ and μ_n), we arrive at the solution for the original function

$$T - t_0 = \frac{Q_1\sqrt{a}}{2\lambda\omega^2}\left[\frac{i\omega\sqrt{i\omega}I_0\left(\sqrt{\frac{i\omega}{a}}\rho\right)e^{i\omega\tau}}{I_1\left(\sqrt{\frac{i\omega}{a}}\rho_0\right)}\right.$$
$$\left. + \frac{(-i\omega)\sqrt{-i\omega}I_0\left(\sqrt{\frac{-i\omega}{a}}\rho\right)e^{-i\omega\tau}}{I_1\left(\sqrt{\frac{-i\omega}{a}}\rho_0\right)}\right]. \tag{4}$$

Relation (4) is valid for a periodic steady-state distribution of the surface temperature of the object in the limit $\tau \to \infty$ (neglecting the sum of terms corresponding to the roots of the Bessel functions μ_n).

Let us analyze these approximation solutions, which are of practical interest for measurements of the thermophysical parameters of samples according to the scheme laid out in Ref. 1. Let, as in Ref. 1, the thermal object, the optical system (in the simplest case a lens), and the ring-shaped $p-n$ photodiode junctions be centered and the width $\rho_0 < r_1/K$, where K is the angular magnification coefficient of the optical system and r_1 is the small radius of the first ring of the $p-n$ junction.

1. In Eq. (4) we expand $I_\nu(\mu)$ in the asymptotic series

$$I_0(z) \approx (1/\sqrt{2\pi z})e^{-z}\left(1+\frac{1}{8z} = \dots\right),$$

$$I_1(x) \approx (1/\sqrt{2\pi x})e^{-x}\left(1-\frac{3}{8x} + \dots\right)$$

at large z and x and keep only the first term. The approximate solution for the original function (4) now takes the form

$$T - t_0 = \frac{Q_1}{\lambda 2^{3/2}} \sqrt{\frac{\rho_0}{\rho}} \sqrt{\frac{2a}{\omega}} \exp[-(\varphi - \varphi_0)]$$

$$\times \cos\left[\omega\tau - (\varphi - \varphi_0) + \frac{\pi}{4}\right],$$

$$\varphi_i = \rho_i \sqrt{\frac{\omega}{2a}}. \tag{5}$$

For the first and second $p-n$ junctions we set the coordinates of the centers of the rings of the thermal object which the photodiode ''sees'' equal to $\rho_i = (r_i + R_i)/2K$, where r_i and R_i are the small and large radii of the ring-shaped $p-n$ junctions and are such that their fields of view do not overlap. The signals of the $p-n$ photodiode junctions, e.g., in the spectral range $8-14$ μm for $T \leqslant 600$ K (the Rayleigh–Jeans range) are described by the functional dependence $U_i = A_i R_d^{(i)}(T - t_0)$, where $R_d^{(i)}$ is the differential resistivity of the $p-n$ junctions, and A_i is a proportionality factor.[1] The wavelengths λ_n of the periodic oscillations of the surface temperature of the sample, which are equal to integer fractions of the segment $\rho_2 - \rho_1$, are equal to

$$\rho_2 - \rho_1 = 2\pi n \sqrt{\frac{2a}{\omega_n}} = n\lambda_n, \tag{6}$$

where $n = 1, 2, 3, \ldots$.

For $n = 1$ the wave has its maximum length.

Substituting the corresponding expressions for ρ_2 and ρ_1 in relation (6), we obtain

$$a_n = \frac{\omega_n}{2}[(r_2 + R_2 - r_1 - R_1)/n 4\pi K]^2, \tag{7}$$

i.e., the thermal diffusivity is determined by the measured frequencies ω_n of the maxima of the harmonic signals on the first and second rings of the ring-shaped $p-n$ photodiode junctions with a 2π phase difference.

Assuming the second term of the asymptotic series of $I_1(x)$ to be e times less than the first, we estimate the range of variation $\omega'_{\min} \geqslant (3e/8)^2(a/2\rho_0^2)$ in which the approximate solution (5) is valid. This condition overlaps the more stringent condition $\mu^2 \geqslant 4(\nu + 1)$, which gives $\omega_{\min} \geqslant 8a/2\rho_0^2$. Since the thickness d of the sample layer from which information about the coefficient a is taken is proportional to the thermal wavelength, by putting $\omega'_{\min}$ into relation (6) we find $d_{\max} \leqslant 11\pi\rho_0/e$ (for $n = 1$) and $K_{\min} \geqslant (3e/8)^2[(r_2 + R_2) - (r_1 + R_1)]/\pi\rho_0$. We measure (for $K_{\min} = $ const) the sequence ω_n from $n = 1$ (for $\omega_1 \geqslant \omega'_{\min}$) to $n = K$ and calculate the corresponding sequences $d_n = d_{\max}/n$ and a_n from formulas (6) and (7). The quantities a_n are the values averaged over d_n. Invoking the rule of moments, we perform a correction and refine the values $a_n^{(0)} = [(a_n^{(1)} + \varepsilon_{n+1})(a_n^{(1)} - \varepsilon_{n+1})]^{1/2}$, $\varepsilon_{n+1} = |a_n - a_{n+1} \times (d_{\max}/n+1)|$, $a_n^{(1)} = [(a_n^{(2)} + \varepsilon_{n+2})(a_n^{(2)} - \varepsilon_{n+2})]^{1/2}$, etc. up to $n + m = K$, where $a_n^{(0)}$ is the refined value of the thermal diffusivity of the nth interior layer of thickness $d_{n,n+1} = d_{\max}/n(n+1)$. To structure with respect to a a lower layer of thickness $d_{\max}/2$, we vary the magnification coefficient K so as to measure the diffusivity a of a lowest layer of

specified thickness $d_{\max} - d_{\max}^{(1)}$, where $d_{\max}^{(1)}$ is calculated from formula (6) for $n = 1$ and $K = K_1$. We assign the coefficient a_1 to be equal to the average $a_1 = (a_1^{(0)} + a_1^{(1)})/2$. Measuring ω_n for $K_1 = $ const, we obtain the new sequences $d_n^{(1)}$ and $a_n^{(1)}$, which give a finer structuring (with respect to a) of the upper layers, analogous to the procedure outlined above. The required sequence K_p is chosen in accordance with the goals of the study. The data set $a_n^{(m)}$ and $d_n^{(m)}$ is correlation-processed with the aim of revealing in which layer an inhomogeneity in the coefficient a (a flaw) is located or how many layers it occupies, the ultimate representation being in the form of thermographic images of subsurface layers in a fixed-step scan.[4]

To extend the range of $K_{\min}$ (with the aim of increasing the depth of layer-by-layer scanning) we retain the first two terms in the asymptotic series for $I_0(z)$ and $I_1(x)$. We then write the approximate solution (4) as

$$T - t_0 = \frac{Q_1}{\lambda \sqrt{2}} \left(\frac{\rho_0}{\rho}\right)^{3/2} \sqrt{\frac{2a}{\omega}} \exp[-(\varphi - \varphi_0)]$$

$$\times \cos[\omega\tau - (\varphi - \varphi_0) + \Psi],$$

$$\Psi = \arctan\left\{\left[8(\varphi_0 + 8\rho_0\rho) - \frac{3}{2}\left(\frac{2a}{\omega}\right)\right] \middle/ \left[8(3\varphi_0 - 8\rho_0\rho) + \frac{3}{2}\left(\frac{2a}{\omega}\right)\right]\right\}, \tag{8}$$

$$\rho_2 - \rho_1 = 2\pi n \sqrt{\frac{2a}{\omega_n}} - (\Psi_2 - \Psi_1)\sqrt{\frac{2a}{\omega_n}}. \tag{6'}$$

Expanding $\arctan[(b - c)/(1 + bc)]$ in a power series and neglecting terms of the order of the small parameter ρ_0 reduces Eq. (6') to the form

$$(\rho_2 - \rho_1)\left[1 + \frac{1}{(\rho_2 + \rho_1)}\sqrt{\frac{2a}{\omega_n}}\right] = 2\pi n \sqrt{\frac{2a}{\omega_n}}, \tag{6''}$$

by way of which the coefficient a is determined from measured values of ω_n by iteration for the limiting value $d''_{\max}$. When ω_n is increased above $\omega'_{\min}$ we arrive again at relation (6).

Relations (6') and (8) are valid for $\omega''_{\min} \geqslant 0.61a/(2\rho_0^2)$, i.e., $\omega'_{\min}/\omega''_{\min} \approx 1.7$. Consequently, the frequency band is increased by a factor of ~ 1.7, and $d_{\max}$, by a factor of 1.3, by decreasing $K_{\min}$ by a factor of 1.3. Taking additional terms of the asymptotic series into account leads to a complication of the formulas for calculating the coefficient a from the measured frequencies ω_n, an insubstantial gain in $d_{\max}$, and in practice to the appearance of phase shifts requiring diagrammatic solutions with a smaller value of $K_{\min}$ and are therefore without purpose.

2. Here we consider the case in which z and x in relation (4) are small, and $x^2 \leqslant 4(\nu + 1)/e$. Expanding $I_\nu(\mu)$ in a power series and retaining the first two terms $I_0(z) \approx 1 + (z/2)^2$, $I_1(x) \approx (x/2)[1 + \frac{1}{2!}(x/2)^2]$ transforms the approximate expression (4) into the following form

$$T - t_0 = \frac{Q_1}{\lambda}\frac{4a}{\rho_0}\cos(\omega\tau - \Phi),$$

$$\Phi = \arctan\left[\left(\frac{4}{\varphi_0^2} + \frac{1}{2}\varphi^2\right)/(1 - 2(\rho/\rho_0)^2)\right], \qquad (9)$$

where Φ is the phase shift between the signal at the $p-n$ junction and the reference voltage of the radiation source.

We set $\Phi_2 - \Phi_1 = \arctan[(b-c)/(1+bc)] = 2\pi$ in Eq. (9). Hence $\tan\arctan\Omega = \tan 2\pi = 0$ and $a = \pm i\omega\rho_0^2/4\sqrt{2}$, i.e., under the prescribed conditions on z and x in the frequency range $\omega \leqslant \omega_{\max} = (8/e)(a/2\rho_0^2)$, thermal wavelengths that are multiples of $\rho_2 - \rho_1$ are absent. Expanding $\arctan\Omega$ in a power series and neglecting terms higher than second order in ρ_i, we arrive at an approximate expression for the measured phase difference $\Delta\Phi \approx 3(\varphi_2^2 - \varphi_1^2)/(1 + 3\varphi_1^2\varphi_2^2)$, from which we determine the coefficient a_g, the thermal wavelength

$$\lambda_g = \sqrt{8\pi}\sqrt{\frac{a_g}{\omega_g}}$$

(in terms of the thermal accommodation coefficient $\sqrt{a_g/\omega_g}$; Ref. 2), and the layer thickness $d_g \approx 1/\lambda_g$. By varying the frequency ω_g and the magnification coefficient K, we obtain the sequences $a_g^{(m)}$ and $d_g^{(m)}$ in analogy with the above result and get the corresponding thermographic image of subsurface layers.

The difference between the procedure in Sec. 2 and the procedure in Sec. 1 is that the frequency ω_g varies smoothly while ω_n varies discretely and it forms the sets $a_n^{(m)}$ and $d_n^{(m)}$. The drawback of the procedure in Sec. 2 is that the measurements are carried out near the center of the thermal object, where the effect of instrument errors on the measurements is higher since the temperature gradient is much larger, and averaging over a_g takes place in addition over a wide temperature range in comparison with the procedure outlined in Sec. 1, where the measurements are carried out in the ''tail'' of the heat wave.

Strictly speaking, as K varies, so do the surface regions from which the photodiode receives radiation. The given procedure is therefore best suited for studying samples with layerwise-homogeneous thermal diffusivity a, e.g., multilayer metallic coatings on hybrid circuit boards, to reveal flaws in the joining of the layers. Alternatively, K may be varied in such a way that a partial overlap of two neighboring regions of the surface takes place upon successive variations of K, so that an averaged image of a in each layer is obtained.

[1] V. I. Turinov, Zh. Tekh. Fiz. **62**, 175 (1992) [Sov. Phys. Tech. Phys. **37**, 714 (1992)].

[2] A. V. Lykov, *Theory of Heat Conduction* [in Russian], Vysshaya Shkola, Moscow (1967).

[3] G. N. Watson, *Treatise on the Theory of Bessel Functions*, 2nd ed. (Cambridge University Press, Cambridge, 1952).

[4] E. V. Bernikov, S. S. Gaponov, and V. I. Turinov, *Abstracts of the Seventh All-Union Conference ''Photometry and Its Metrological Basis''* [in Russian], VNIIOFI, Moscow (1990), p. 114.

Translated by Paul F. Schippnick

Effect of dispersion of orthorhombic anisotropy fields on the ferromagnetic resonance linewidth in iron garnet films

A. M. Zyuzin and V. V. Radaĭkin

N. P. Ogarev Mordovian State University, 430000 Saransk, Russia

The effect of the dispersion of anisotropy fields on the width of the ferromagnetic resonance line is calculated in films with orthorhombic anisotropy. The results of the calculation, based on a model of linear variation of the anisotropy fields with thickness and on the assumption of additivity of the different contributions to the resulting linewidth, are in good agreement with experimental results. © *1997 American Institute of Physics.* [S1063-7842(97)02408-2]

The ferromagnetic resonance linewidth in magnetic films depends on many factors. Besides relaxation mechanisms, the linewidth is affected, for example, by variation of the equilibrium orientation of the magnetization[1] occurring during recording of the absorption line. Broadening can take place due to the presence of defects (pores, dislocations) both in metallic films and in ferro-insulator films.[2] The mechanism associated with the magneto-elastic contribution to the anisotropy field and leading to broadening of the resonance line in polycrystalline films was investigated in Ref. 3. References 4 and 5 investigated peculiarities of the broadening associated with angular dispersion of the anisotropy. In this type of dispersion the maxima in the angular dependence of the resonance field are observed in intermediate (between hard and easy) directions of the static magnetic field.

Studies of epitaxial iron garnet films of the composition $(YBiTmGd)_3(FeGa)_5O_{12}$ grown on substrates of neodymium gallium garnet with (110) orientation revealed an angular dependence of the resonance linewidths, shown in Fig. 1a. This same figure shows the angular dependence of the resonance field $H_p(\varphi_H)$. The angle φ_H between the field **H** and the hard axis was varied in the film plane. Note that the films were grown in the regime of linearly decreasing temperature growth and possessed orthorhombic anisotropy. The orthorhombic anisotropy fields

$$H_{ku}=\frac{2K_u}{M}-4\pi M_s \quad \text{and} \quad H_{kr}=\frac{2K_r}{M}$$

were equal to 1098 and 1411 Oe, respectively. In the above expressions K_u and K_r are the constants of growth-induced uniaxial and rhombic (describing the anisotropy in the film plane) components of the anisotropy.[6–8] The thickness of the original films was 2.5 μm, the saturation magnetization $4\pi M_s$ was equal to 713 G, and the gyromagnetic ratio γ was equal to 1.39×10^7 $Oe^{-1}\cdot s^{-1}$. The latter was determined from the resonance fields for the static field **H** oriented along the hard and soft directions in the film plane and in the direction normal to it.[6] Measurements of the ferromagnetic resonance parameters were carried out at a frequency of 9.34 GHz. As follows from Fig. 1a, the linewidth varies strongly in the angle interval from 0 to 180°. A peculiarity of the observed dependence $2\Delta H(\varphi_H)$ is that the maxima of $2\Delta H$ are observed for **H** oriented along the hard and soft directions, and the minima, at intermediate orientations. Such behavior of $2\Delta H$ cannot be explained by the above-enumerated broadening mechanisms.

Results of layered etching experiments show that as the film thickness is decreased the amplitude of variation of the linewidth in the angular dependence $2\Delta H(\varphi_H)$ decreases (Figs. 1b and 1c). The interval of variation of the resonance field also decreases. Hence it follows that the observed dependence $2\Delta H(\varphi_H)$ can also not be explained by a possible anisotropy of the damping parameter.

To explain the angular dependence of $2\Delta H$ we begin by assuming the presence of a nonuniformity simultaneously in H_{ku} and H_{kr} over the thickness of the film, which is connected with the unsteady kinetics of its growth. Dispersion of H_{ku} and H_{kr} should lead to the appearance of an additional contribution to the linewidth. In the case of sufficiently large values of the damping parameter α the absorption line in a nonuniform film can be considered as a superposition of lines from different layers. Therefore the resulting linewidth was considered under the assumption of additivity of the intrinsic linewidth $2\Delta H_e$ and the contribution due to dispersion of the anisotropy fields $2\Delta H_d$. The latter can be calculated by representing it in the form

$$2\Delta H_d=2\left(\frac{\partial H_{res}}{\partial H_{kr}}\Delta H_{kr}+\frac{\partial H_{res}}{\partial H_{ku}}\Delta H_{ku}\right), \qquad (1)$$

where ΔH_{kr} and ΔH_{ku} are the dispersions of the corresponding anisotropy fields.

However, the expression for $2\Delta H_d$ in this representation is very cumbersome. Therefore, the contribution $2\Delta H_d$ for a given value of φ_H was determined as the difference of the maximum and minimum fields in the family of resonance curves constructed for given intervals of values of H_{ku} and H_{kr}.

The analysis was based on a model of linear variation of H_{ku} and H_{kr} over the thickness of the film (Fig. 2). This corresponds to an equiprobable distribution of the magnetic moments of the film relative to the fields H_{ku} and H_{kr} in the intervals of variation of these fields. The resonance curves were calculated with the help of the method of effective demagnetizing factors.[9] The energy density of the anisotropy of the magnetic film with allowance for the energy of the demagnetizing field can be written in the form
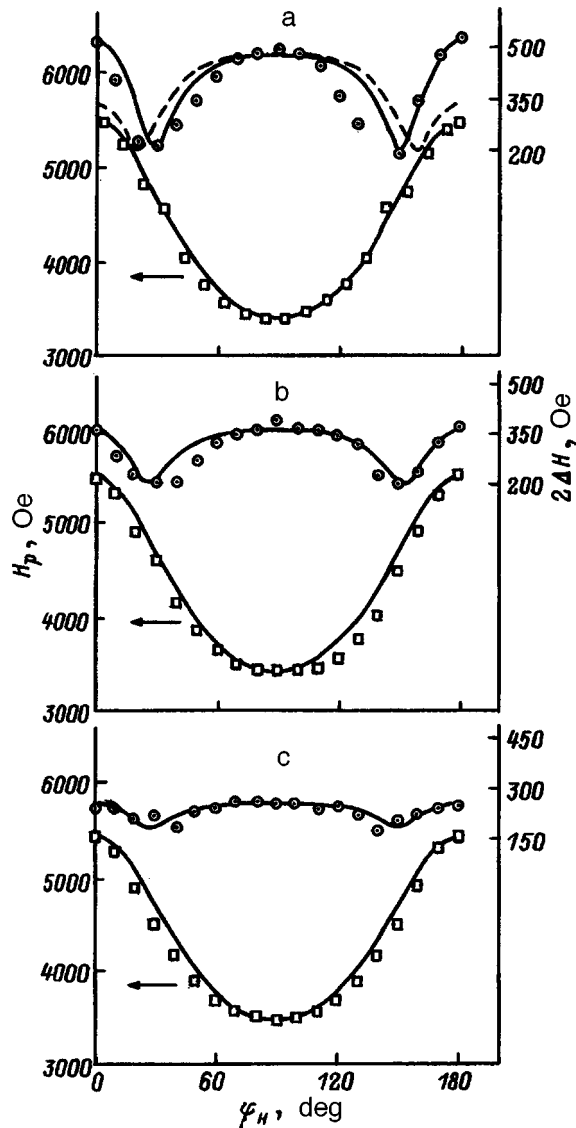
FIG. 1. Angular dependence of $2\Delta H$ and H_p: a) original film ($h_1 = 2.5$ μm), b) after first stage of etching ($h_2 = 1.76$ μm), c) after second stage of etching ($h_3 = 1.0$ μm). Circles and squares correspond to experiment, solid lines to calculation, and the dashed line to calculation with the dispersion only of H_{kr} taken into account.

$$U_a = K_u\left(1 - \frac{M_{z'}^2}{M^2}\right) + K_r\frac{M_{y'}^2}{M^2} + 2\pi M_{z'}^2, \tag{2}$$

where $M_{z'}$ and $M_{y'}$ are the components of the magnetization in the coordinate system whose z' and y' axes coincide with the anisotropy axes of the film.

We found the values of the components of the tensor of effective demagnetizing factors by comparing the expressions for the effective field[9]

$$\mathbf{H}_a^{\mathrm{eff}} = -\frac{\partial U}{\partial \mathbf{M}} \quad \text{and} \quad \mathbf{H}_a^{\mathrm{eff}} = -N^{\mathrm{eff}}\mathbf{M}. \tag{3}$$

To start with, we determined the components of the tensor $(N^{\mathrm{eff}})'$ in the above-mentioned coordinate system. They have the following form:
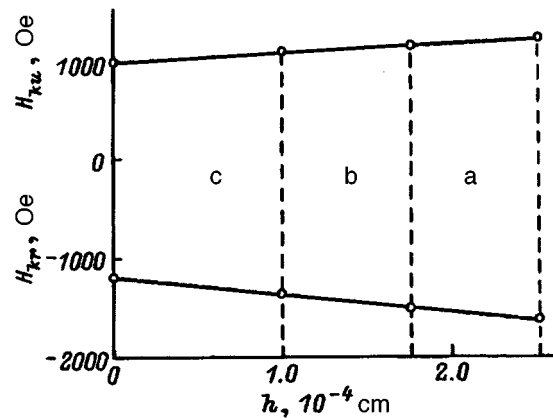

FIG. 2. Distribution of fields of the uniaxial component H_{ku} and rhombic component H_{kr} of the anisotropy over film thickness. The points of intersection with the vertical dashed lines determine the values of H_{ku} and H_{kr} on the surface: a) original film; b and c) after first and second stages of etching, respectively.

$$\left(\overset{\leftrightarrow}{N^{\mathrm{eff}}}\right)' = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \dfrac{2K_r}{M^2} & 0 \\ 0 & 0 & -\dfrac{2K_u}{M^2} + 4\pi \end{pmatrix}. \tag{4}$$

We found expressions for the components of N^{eff} in the coordinate system where the z axis coincides with the magnetization with the help of formulas for transforming the components of a tensor when transforming from one coordinate system to another.[10] The resonance relation for the case when $\mathbf{H}$ lies in the plane of a film with orientation (110) has the form

$$\left(\frac{\omega}{\gamma}\right)^2 = [H\cos(\varphi_H - \varphi_M) + H_{kr}\cos 2\varphi_M]$$

$$\times [H\cos(\varphi_H - \varphi_M) - H_{ku} + 4\pi M$$

$$- H_{kr}\sin^2\varphi_M]. \tag{5}$$

Here φ_M is the angle between $\mathbf{M}$ and the hard axis, whose value was found from the condition of equilibrium orientation of the magnetization. As analysis of the experimental angular dependences of the resonance field showed, the cubic anisotropy constant K_1 in the investigated films was roughly an order of magnitude lower than K_u and K_r. Therefore we neglected cubic anisotropy in the calculations of $2\Delta H(\varphi_H)$.

The family of resonance curves for the original film with the corresponding intervals of variation of H_{ku} and H_{kr} is shown in Fig. 3. As follows from the calculations, the spread of the resonance fields for $\varphi_H = 0$ and 90° is roughly equal. The minimum is observed at angles φ_H (between $\mathbf{H}$ and the hard axis) close to 30°.

As the film thickness is decreased by means of layered etching, the intervals of variation of the anisotropy fields H_{ku} and H_{kr} are observed to decrease (Fig. 2). In this connection, the contribution to $2\Delta H$ due to this factor should also de-
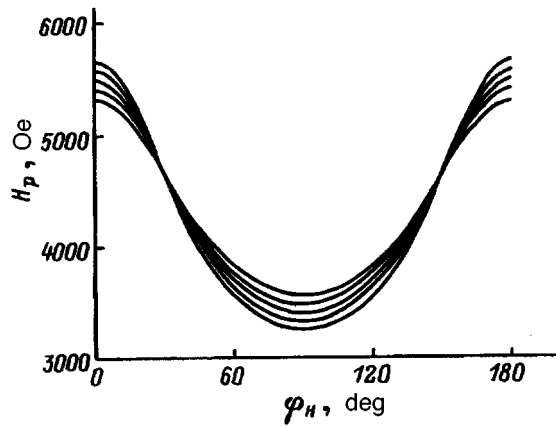
FIG. 3. Family of resonance curves calculated for values of H_{ku} and H_{kr} corresponding to different layers of the original film.

FIG. 4. Temperature dependence of $2\Delta H$ (a) and H_p (b). $1,2$ — for the field **H** parallel to the hard and easy axes, respectively, 3 — for **H** parallel to the direction at which $2\Delta H$ takes it minimum.

crease. Experiments and calculations confirm this conclusion. As follows from Figs. 1b and 1c, as the film thickness h is decreased from 1.76 to 1.0 μm the amplitude of variation of $2\Delta H$ in the angular dependence decreases substantially. The variation of the minimum values of $2\Delta H$ here does not exceed the error of measurement. This is explained by the fact that the minimum of the contribution $2\Delta H_d$ is insignificant and, consequently, the minimum value of $2\Delta H$ is determined mainly by the intrinsic linewidth.

As h is decreased, the maximum resonance field in the angular dependence $H_p(\varphi_H)$ is observed to decrease and the minimum resonance field to increase. This indicates that the absolute values of H_{ku} and H_{kr} increase from the substrate to the free surface of the film, and not *vice versa*. The calculated dependences $H_p(\varphi_H)$ (Fig. 1) were obtained from the mean values of H_{ku} and H_{kr} for the given thickness. It can be seen that there is quite good agreement between the theoretical and experimental dependences both of $2\Delta H(\varphi_H)$ and $H_p(\varphi_H)$.

Some of the difference between them may be due to a deviation from a linear distribution of the magnetic moments of the sample on the prescribed intervals of variation of H_{ku} and H_{kr}, and also to effects of exchange and dipole narrowing of the inhomogeneously broadened line.

Good agreement was observed for the calculated and experimental polar-angle dependences $2\Delta H(\theta_H)$ when the angle θ_H between **H** and the film normal was varied in the planes passing through the hard and easy directions in the film plane and the film normal. Note that taking only the dispersion of the rhombic component of the anisotropy field into account does not make it possible to obtain satisfactory agreement between the calculated (dashed curve in Fig. 1a) and experimental results. In these calculations H_{ku} was taken to be equal to its average value over the corresponding interval.

Temperature studies also confirm the hypothesis of the presence of dispersion of the anisotropy fields and its effect on $2\Delta H$. Figures 4a and 4b plot the temperature dependence of $2\Delta H$ and H_p for different orientations of **H** relative to the hard axis. It can be seen that there is a clear correlation in the behavior of $2\Delta H$ and H_p for all three orientations. With
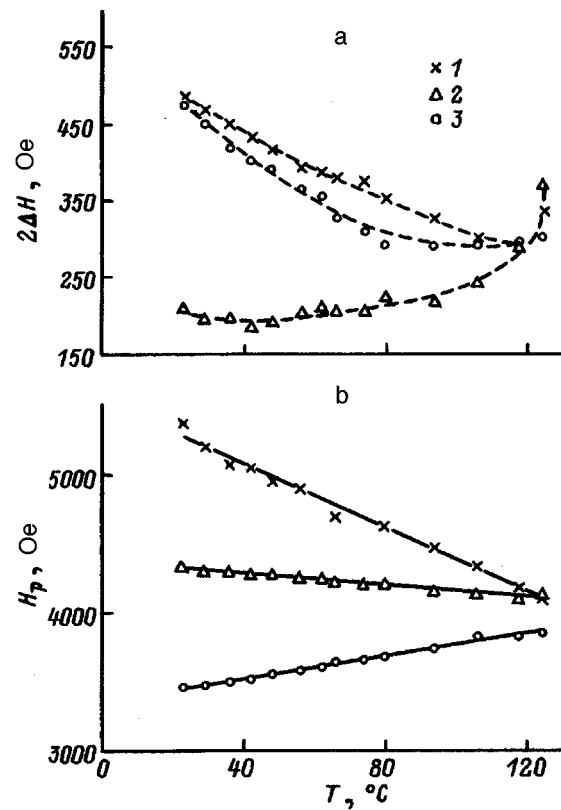
decrease of the absolute values of H_{ku} and H_{kr} upon increase of the temperature the magnitude of their dispersion decreases. This in turn leads to a decrease of $2\Delta H_d$.

As the Curie temperature T_k is approached, the resonance fields for different orientations of the magnetic field converge, which indicates that H_{ku} and H_{kr} approach zero in this limit. In this limit $2\Delta H$ also tends to one value for different orientations. The minimum value of $2\Delta H$ varies weakly all the way to temperatures near T_k. For this orientation the contribution to H_p due to the fields H_{ku} and H_{kr} is minimal. Consequently, the contribution to $2\Delta H$ due to dispersion of these parameters is also minimal.

To summarize, we may make the following conclusions.

1. The effect of dispersion of the anisotropy fields on the ferromagnetic resonance linewidth has been taken into account.

2. A model based on simultaneous variation of H_{ku} and H_{kr} over the thickness of the film and the assumption of additivity of the various contributions to the linewidth makes it possible to explain the angular and temperature dependences of $2\Delta H$ and also the results of layered etching experiments.

3. The presence of dispersion of H_{ku} and H_{kr} in films with orthorhombic anisotropy leads to a qualitatively different angular dependence of the ferromagnetic resonance linewidth.

[1] A. M. Zyuzin, Fiz. Tverd. Tela **31**, No. 7, 109 (1989) [Sov. Phys. Solid State **31**, 1161 (1989)].

[2] S. Krupichka, *Physics of Ferrites and Related Magnetic Oxides* [Russian translation, Mir, Moscow, 1976], Vol. 2.

[3] V. N. Van'kov, A. M. Zyuzin, and Yu. V. Starostin, Pis'ma Zh. Tekh. Fiz. **18**(21), 66 (1992) [Tech. Phys. Lett. **18**, 713 (1992)].

[4] G. G. Kirsanov, Yu. V. Kornev, D. I. Sementsov, and V. V. Sidorenkov, Fiz. Met. Metalloved. **61**, 750 (1986).

[5] D. I. Sementsov and V. V. Sidorenkov, Fiz. Met. Metalloved. **65**, 219 (1988).

[6] H. Makino and Y. Hidaka, Mater. Res. Bull. **16**, 957 (1981).

[7] R. M. Sabitov, R. M. Vakhitov, and E. G. Shanina, Mikroélektronika **18**, 266 (1989).

[8] V. N. Van'kov and A. M. Zyuzin, Zh. Tekh. Fiz. **62**, No. 5, 119 (1992) [Sov. Phys. Tech. Phys. **37**, 548 (1992)].

[9] A. G. Gurevich, *Magnetic Resonance in Ferrites and Antiferromagnets* [in Russian], Nauka, Moscow (1973).

[10] N. E. Kochin, *Vector Calculus and Elements of Tensor Calculus* [in Russian], Nauka, Moscow (1965).

Translated by Paul F. Schippnick

Magnetic anisotropy of (100) and (110) oriented $(Gd,Bi)_3Fe_5O_{12}$ films

V. V. Randoshkin, V. I. Kozlov, V. Yu. Mochar, N. V. Vasil'eva, and V. V. Voronov

Magnetooptoélektronika Joint Self-Supporting Laboratory of the Institute of General Physics of the Russian Academy of Sciences at the N. P. Ogarev Mordovian State University, 430000 Saransk, Russia

One way of increasing the speed of magnetooptical devices based on single-crystal bismuth-containing iron garnet films is to use films with rhombic magnetic anisotropy.[1–5] Rhombic magnetic anisotropy occurs, in particular, for such films with orientations other than (111) if they contain Y^{3+} ions in addition to the Bi^{3+} ions.[2,6–9] Rhombic anisotropy has also been observed in films also containing Gd^{3+} ions.[10,11] However, it remains unclear whether the Gd^{3+} ions contribute to the anisotropy or if it is completely determined by the Bi^{3+} and Y^{3+} ions. The present paper seeks to answer this question.

$Bi_xGd_{3-x}Fe_5O_{12}$ films were grown by liquid-phase epitaxy from a supercooled PbO-Bi_2O_3-B_2O_3 based fluxed melt on nearly (100) and (110) oriented $Nd_3Ga_5O_{12}$ substrates. The crystallographic orientation (the angle θ of the deviation of their plane from the basal plane) was monitored on a DRON-2.0 diffractometer with an accuracy of $0.1°$.

The lattice mismatch of the film and substrate $\delta a/a$ was determined by the standard method from the rocking curves measured on a two-crystal x-ray spectrometer. The spectrometer assembly consisted of a DRON-3M diffractometer and a monochromator unit using a high-quality germanium crystal and (333) reflection. To identify the peaks from the film and substrate, the rocking curves were recorded for two orders of reflections and the ratios of peak intensities compared.

For the investigated films only one essentially undistorted diffraction peak was observed. Calculation of the linear absorption coefficient for $Bi_xGd_{3-x}Fe_5O_{12}$ films with $0.5 < x < 1$ gives the value $\mu \approx 1500$ cm^{-1} for Cu K_α radiation. The maximum film thickness h_{max} allowing one to record the rocking curves from the film and substrate was estimated from the formula[12–14]

$$h_{max} = k(\sin\theta)/(2\mu), \qquad (1)$$

where $k = 4.61$ for the case in which the fraction of radiation scattered from the film is 99%; for this case $h_{max} \approx 7$ μm.

The absence of a second peak may be interpreted as due to a small value of the mismatch, preventing us from resolving the peaks of the film and substrate. The upper limit of this mismatch can be estimated as half the half-width of the observed rocking curve $\Delta\theta = 20''$, which gives $\delta a/a \approx 8 \times 10^{-5}$. The values of θ and $\delta a/a$ are given in Table I. Noting the absence of a lattice mismatch between the film and substrate according to the data in Table 9.4 in Ref. 12, we obtain $x = 0.78$.

To estimate the saturation magnetization $4\pi M_s$ of the investigated films, we used the data of Table 9.3 in Ref. 15.

We assumed that in the system $Bi_xGd_{3-x}Fe_5O_{12}$ the saturation magnetization increases linearly with growth of x from the value corresponding to $Gd_3Fe_5O_{12}$ to the value corresponding to $Bi_3Fe_5O_{12}$. The saturation magnetization for the latter was taken to be the same as for $Y_3Fe_5O_{12}$. An estimate gives $4\pi M_s = 500$ G.

The magnetic anisotropy parameters were investigated by the ferromagnetic resonance method at a frequency of 9.34 GHz. Here we recorded the resonance fields for the external magnetic field oriented perpendicular (H_n, perpendicular resonance) and parallel (H_{in}, parallel resonance) to the plane of the film, and also the corresponding values of the resonance line width $2\Delta H_n$ and $2\Delta H_{in}$. To determine the anisotropy in the film plane, we recorded the azimuthal dependences $H_{in}(\varphi)$ for parallel resonance. The results of these measurements are listed in Table I, where $H_{in\,min}$ and $H_{in\,max}$ are the minimum and maximum values of the resonance field for parallel resonance, and δH_{in} is the difference of these values.

Since the investigated films do not contain rapidly relaxing magnetic ions, and the gyromagnetic ratio of the slowly relaxing ions Gd^{3+} and Fe^{3+} is the same and equal to $\gamma_0 = 1.76 \times 10^7$ Oe$^{-1} \cdot$s^{-1}, for an iron garnet according to the formula of Wangness[1,16]

$$\gamma = (M_{Gd} + M_{Fe})/(M_{Gd}/\gamma_{Gd} + M_{Fe}/\gamma_{Fe}), \qquad (2)$$

the effective value of the gyromagnetic ratio γ is equal to γ_0. Here M_{Gd} is the total magnetic moment of the Gd^{3+} ions in the dodecahedral sublattice of the garnet structure, M_{Fe} is the total magnetic moment of the Fe^{3+} ions in the tetrahedral and octahedral sublattices. This circumstance allows us, in contrast to Refs. 17 and 18, to determine the effective mag-

TABLE I. Results of x-ray and resonance measurements for $(Gd,Bi)_3Fe_5O_{12}$ films.

Orientation	(100)	(110)
θ, deg	2.3	1.6
$\delta a/a \times 10^4$	<0.8	<0.8
$H_{in\,min}$, Oe	3130	2240
$H_{in\,max}$, Oe	3490	3330
δH_{min}, Oe	360	1090
$2\Delta H_{min}$, Oe	170	70
H_n, Oe	3390	4300
$2\Delta H_n$, Oe	170	70
H_{eff}, Oe	−207	−870

FIG. 1. Azimuthal dependence $H_{in}(\varphi)$ of the resonant field in parallel resonance for films with (100) orientation.
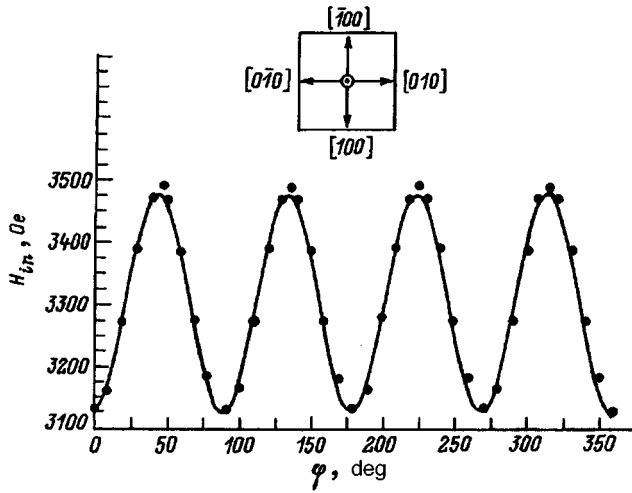

FIG. 2. Azimuthal dependence $H_{in}(\varphi)$ of the resonant field in parallel resonance for films with (110) orientation.

netic anisotropy field H_{eff} only from the perpendicular resonance data. For this field the resonance relation can be written as

$$\omega/\gamma = H_{eff} + H_n, \tag{3}$$

where ω is the angular frequency of the ferromagnetic resonance.

The fact that the Gd^{3+} and Fe^{3+} ions behave identically in the dynamics is corroborated by the results of studies near the angular momentum compensation point.[5,18,19] In particular, when Gd^{3+} ions are added to the film the level of substitution of iron by nonmagnetic ions required to bring about compensation of the angular momentum is lowered. Values of H_{eff} are given in Table I.

Typical dependences of $H_{in}(\varphi)$ for (100) and (110) oriented films are shown in Figs. 1 and 2, respectively, where the insets show the arrangement of the crystallographic axes in the film plane. It can be seen that for (100) oriented films the presence of four equivalent maxima on the $H_{in}(\varphi)$ curve (Fig. 1) reflect the crystallographic anisotropy of the film, while their positions correspond to (100) type axes. Here the value of δH_{in} is relatively small (see Table I).

In (110) oriented films in comparison with (100) oriented films the magnetic anisotropy in the film plane more than three times higher (compare the values of δH_{in} in Table I). In this case only two peaks are observed in the azimuthal dependence of the resonance field (Fig. 2), which is characteristic of films with rhombic magnetic anisotropy.

A comparison of the above estimate of $4\pi M_s$ and the values of H_{eff} listed in Table I allows one to conclude that the demagnetization fields deliver the main contribution to the effective field.

To summarize, in this paper we have shown that a rhombic magnetic anisotropy is induced by epitaxial growth in (110) oriented iron garnet films containing only Gd^{3+} and Bi^{3+} ions in their dodecahedral sublattice.
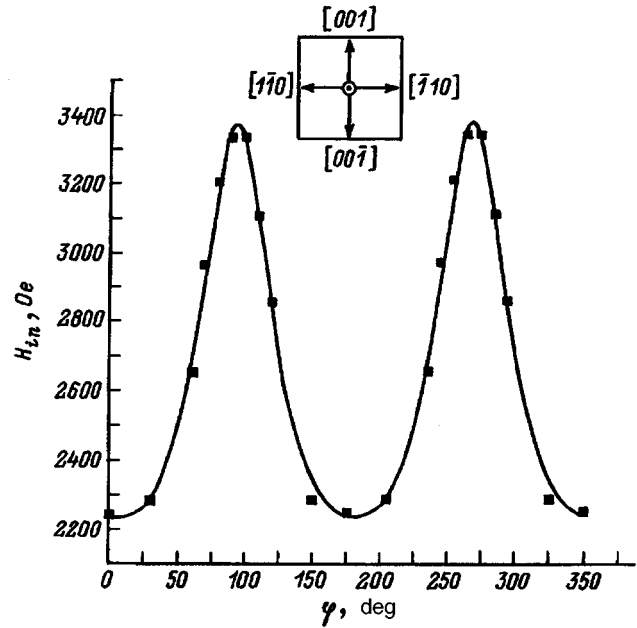
[1] V. V. Randoshkin and A. Ya. Chervonenkis, *Applied Magnetooptics* [in Russian], Energoatomizdat, Moscow (1990).

[2] V. V. Randoshkin, V. I. Chani, M. V. Logunov *et al.*, Pis'ma Zh. Tekh. Fiz. **15**(14), 42 (1989) [Tech. Phys. Lett. **15**, 553 (1989)].

[3] V. V. Randoshkin, *Proceedings of SPIE*, Vol. 1307, 10 (1990).

[4] V. V. Randoshkin, *Proceedings of SPIE*, Vol. 1469, 796 (1991).

[5] V. V. Randoshkin, *Magnetooptical Iron Garnet Films and Their Applications* [in Russian], Trudy IOFAN, Vol. 35 (Nauka, Moscow, 1992), pp. 49–107.

[6] T. Hibiya, H. Makino, and S. Konishi, J. Appl. Phys. **52**, 7347 (1981).

[7] S. Kikukawa, S. Isomura, and S. Iwata, J. Appl. Soc. Jpn. **7**(2), 83 (1983).

[8] V. V. Randoshkin, V. B. Sigachev, V. I. Chani, and A. Ya. Chervonenkis, Fiz. Tverd. Tela **31**(7), 70 (1989) [Sov. Phys. Solid State **31**, 1138 (1989)].

[9] M. V. Logunov, V. V. Randoshkin, and Yu. N. Sazhin, Fiz. Tverd. Tela **32**, 1456 (1990) [Sov. Phys. Solid State **32**, 849 (1990)].

[10] M. Hamamoto, S. Iwata, S. Kikukawa, IEEE Trans. Magn. **MAG-20**, 1105 (1984).

[11] I. Yu. Kukushkina, V. V. Randoshkin, V. B. Sigachev, and M. I. Timoshechkin, ''The present level of development of magnetic bubble memory and logic devices,'' *Abstracts of All-Union Seminar*, Moscow (1985), p. 37.

[12] S. S. Gorelik, Yu. A. Skakov, and L. N. Rastorguev, *X-Ray and Electron-Optical Analysis* [in Russian], Moscow (1994).

[13] Yu. A. Tkhorik and L. S. Khazan, *Plastic Deformation and Misfit Dislocations in Epitaxial Heterostructures* [in Russian], Naukova Dumka, Kiev (1983).

[14] L. I. Mirkin, *Handbook on X-Ray Structural Analysis of Polycrystals* [in Russian], GIFMI, Moscow (1961).

[15] A. M. Balbashov, F. V. Lisovskiĭ, V. K. Raev *et al.*, *Handbook of Magnetic Bubble Devices and Device Components*, edited by N. N. Evtikhiev and B. N. Naumov [in Russian], Radio i Svyaz', Moscow, 1987.

[16] R. K. Wangness, Phys. Rev. **91**, 1085 (1953).

[17] K. Gangulee and R. J. Kobliska, J. Appl. Phys. **51**, 3333 (1980).

[18] N. A. Loginov, M. V. Logunov, and V. V. Randoshkin, Fiz. Tverd. Tela **31**(10), 58 (1989) [Sov. Phys. Solid State **31**, 1684 (1989)].

[19] V. V. Randoshkin and V. B. Sigachev, JETP Lett. **42**, 41 (1985).

Translated by Paul F. Schippnick

Transformation of beams of dia-, para-, and ferromagnetic particles by the magnetic field of a linear current

N. I. Shtepa

Chernigov State Pedagogical Institute, 250038 Chernigov, Ukraine
(Submitted June 8, 1996)
Zh. Tekh. Fiz. **67**, 138–141 (August 1997)

Some possibilities of transformation and trapping of beams of small ($10^{-1}-10^{-4}$ cm) dia-, para-, and ferromagnetic particles by the magnetic field of a linear, constant current are investigated. © *1997 American Institute of Physics.* [S1063-7842(97)02608-1]

INTRODUCTION

We consider homogeneous (before transformation) beams of small, spherical ($10^{-1}-10^{-4}$ cm) particles of dia-, para-, and ferromagnetic materials, moving in vacuum with the same initial velocity ($\mathbf{v}_0 = \text{const}$). The beams are sufficiently rarefied that interactions between their particles can be neglected. The beams are transformed by the magnetic field of a constant, rectilinear, cylindrical current $I = \text{const}$ of radius a. The motion of the particles is described in cylindrical coordinates (z, φ, z), where the z axis is directed along the current axis, r is the distance from this axis to the particle, and φ is the angle of rotation about the z axis.

This study is restricted to two cases: longitudinal and transverse transformation. In longitudinal transformation, up until the transformation the particle beam has a hollow cylindrical shape, coaxial with the current, bounded by an inner radius q and an outer radius Q ($a < q < Q$). The initial velocities of the beam particles are directed parallel to the current with an offset distance relative to the current axis $r_0 = p$ ($q \leqslant p \leqslant Q$) which is different for each particle. The particle trajectories remain planar during the transformation process, lying in the $\varphi = \text{const}$ plane.

In transverse transformation, a ribbon-shaped beam prior to the transformation, at a large distance from the current axis $r_0 - l \gg Q$ at which the action of the field may be neglected, moves in a direction crossed with the current and perpendicular to it. The offset distance of a particle (the shortest distance from its initial direction of motion to the current axis) we denote by p. The offset distances of the particles bounding the ribbon beam we denote by q and Q, respectively ($a < q \leqslant p \leqslant Q$); thus, the thickness of the beam before the transformation is $Q - q$.[1)] The particle trajectories in this transformation also remain planar, but lying in planes $z = \text{const}$. The polar angle φ is the angle between the half-line extending from the current axis, antiparallel to the initial velocity of the particle $\mathbf{v}_0$, and the polar radius r.

TRANSFORMATION OF BEAMS OF DIA- AND PARAMAGNETS

If it is assumed that the magnetic field $\mathbf{H}$ of the linear current is quasistationary over the extent of a particle of radius R and with permeability μ, the magnetic moment[1] of the particle will be equal to

$$\mathbf{M} = R^3 \frac{\mu - 1}{\mu + 2} \mathbf{H}. \tag{1}$$

It is customary to express the force $\mathbf{F}$ acting on the particle in the first approximation as[2]

$$\mathbf{F} = R^3 \frac{\mu - 1}{\mu + 2} (\mathbf{H} \cdot \nabla) \mathbf{H}. \tag{2}$$

The magnetic field of a linear current,

$$\mathbf{H} = \frac{2I}{cr} \mathbf{e}_\varphi \tag{3}$$

($\mathbf{e}_\varphi$ is the unit vector in the $\mathbf{H}$ direction) has the property that the force according to relation (2) is equal to zero; we therefore go to the next following approximation:

$$\mathbf{F} = \mathbf{M} \cdot \frac{\partial \mathbf{H}}{\partial \mathbf{r}} \mathbf{e}_r, \tag{4}$$

where $\mathbf{e}_r$ is the unit radial vector outward from the current axis.

Equations (1), (2), and (4) yield

$$\mathbf{F} = -\frac{4\pi R^3 I^2}{c^2} \frac{\mu - 1}{\mu + 2} r^{-3} \mathbf{e}_r. \tag{5}$$

a) *Longitudinal transformation.* In this transformation of the beam the motion of a particle of density ρ in the plane of its motion ($\varphi = \text{const}$) is described by the differential equation

$$\ddot{r} = -\alpha r^{-3}, \tag{6}$$

where

$$\alpha = \frac{3I^2}{\pi c^2 \rho} \frac{\mu - 1}{\mu + 2} \tag{7}$$

and

$$z = v_0 t, \tag{8}$$

$\alpha < 0$ for diamagnets and $\alpha > 0$ for paramagnets.

Integrating Eq. (6) with initial conditions $t_0 = 0$, $r_0 = p$, $\dot{r}_0 = 0$ and eliminating t via Eq. (8), we find the equation of the particle trajectory

$$r^2 = p^2 - \frac{\alpha}{p^2 v_0^2} z^2. \tag{9}$$

Trajectories of paramagnets, according to Eq. (9), curve toward the current, and as a result the beam particles are trapped by the current — they reach the current surface ($r=a$). Trapping in the z coordinate takes place within the interval $z_q \leqslant z \leqslant z_Q$, whose end-points z_q and z_Q are given by the formula

$$z_\gamma = \gamma a \sqrt{\frac{\gamma^2 - a^2}{\alpha}} \tag{10}$$

with γ replaced by q or Q, respectively. The time T the particles are trapped at the offset distance p is given by

$$T = p \sqrt{\frac{p^2 - a^2}{\alpha}}. \tag{11}$$

The trapped particle density $\eta(z)$ decreases with increasing z

$$\eta(z) = \eta(z_q) \frac{z \sqrt{\frac{a^2}{2} + \sqrt{\frac{a^4}{4} + \frac{\alpha}{v_0^2} z_q^2}} \sqrt{\frac{a^4}{4} + \frac{\alpha}{v_0^2} z_q^2}}{z_q \sqrt{\frac{a^2}{2} + \sqrt{\frac{a^4}{4} + \frac{\alpha}{v_0^2} z^2}} \sqrt{\frac{a^4}{4} + \frac{\alpha}{v_0^2} z^2}}. \tag{12}$$

The trajectories of diamagnetic particles curve away from the current axis — they are scatttered, and at large distances ($z \gg Q^2 v_0^2 / \sqrt{|\alpha|}$) they are transformed into a hollow diverging beam, bounded by the angles ψ_Q and ψ_q ($\psi_Q \leqslant \psi \leqslant \psi_q$), where the formula

$$\psi_\gamma = \arctan \frac{\sqrt{|\alpha|}}{\gamma v_0} \tag{13}$$

yields these angles upon substituting Q and q, respectively, for γ. By transforming, the beam becomes twisted, and the inner and outer (from the axis) beam trajectories change places. After this transformation, a uniform beam becomes nonuniform, decreasing in density $\eta(\psi)$ with growth of the scattering angle ψ. At large distances

$$\eta(\psi) \approx \eta(\psi_Q) \frac{\sin^2 \psi_Q}{\sin^2 \psi} \frac{p}{Q}. \tag{14}$$

b) *Transverse transformation.* In this transformation the motion of a particle in polar coordinates (r, φ) in the plane of motion ($r = \mathrm{const}$) under the action of the force (5) is described by the differential equations

$$\ddot{r} - r\dot{\varphi}^2 = -\alpha r^{-3} \tag{15}$$

and

$$r^2 \dot{\varphi} = p v_0. \tag{16}$$

Eliminating t, we obtain a differential equation for the particle trajectories

$$\frac{d^2 \left(\frac{1}{r}\right)}{d\varphi^2} + \left(1 - \frac{\alpha}{p^2 v_0^2}\right) \frac{1}{r} = 0. \tag{17}$$

We obtain the equation of the trajectory of a diamagnetic particle by integrating Eq. (17) for $\varphi_0 = 0$, $r_0 = p$, and

$$\left(\frac{1}{r^2} \frac{dr}{d\varphi}\right)_0 = -\frac{1}{p},$$

$$r = \frac{kp}{\sin kp}, \tag{18}$$

where

$$k = \sqrt{1 - \frac{\alpha}{p^2 v_0^2}}. \tag{19}$$

This trajectory is curved away from the current and passes at a minimum distance $r_m = kp$ from it for $\varphi_m = \pi/2k$. As a result of the transformation, the beam of diamagnets is transformed into a diverging beam, bounded by the surfaces

$$r = \frac{\sqrt{1 + \frac{|\alpha|}{\gamma^2 v_0^2}}}{\sin\left(\varphi \sqrt{1 + \frac{|\alpha|}{\gamma^2 v_0^2}}\right)} \tag{20}$$

with γ replaced by Q and q, respectively. These surfaces approach the asymptotic limits

$$\varphi_\gamma = \frac{\pi}{\sqrt{1 + \frac{|\alpha|}{\gamma^2 v_0^2}}} \tag{21}$$

with γ replaced by the same values as above. As a consequence of the transformation, the diamagnet beam becomes twisted and nonuniform. At large r ($r \gg kQ$) the density $\eta(\varphi)$ of the transformed beam decreases with increasing φ

$$\eta(\varphi) \approx \eta(\varphi_q) \left(\frac{\pi^2 - \varphi_q^2}{\pi^2 - \varphi^2}\right)^{3/2}. \tag{22}$$

The transverse transformation of a beam of paramagnets depends on the sign of the radicand in expression (19) for k. For $\alpha/(p^2 v_0^2) < 1$, expression (18) remains the solution of Eq. (17), but with a trajectory curved toward the current. If in this case $r_m < a$, i.e., $p^2 - a^2 < \alpha/v_0^2$, the particle is trapped by the current. If

$$Q^2 - a^2 < \frac{\alpha}{v_0^2}, \tag{23}$$

the entire beam is trapped.[2] The time a particle is trapped at the offset distance p is equal to

$$T = \frac{1}{2v_0} \ln \frac{l^2 + \sqrt{l^2 - p^2 + \frac{\alpha}{v_0^2}}}{a^2 + \sqrt{a^2 - p^2 + \frac{\alpha}{v_0^2}}}. \tag{24}$$

If, on the contrary, $r_m > a$ for all the particles, i.e.,

$$q^2 - a^2 > \frac{\alpha}{v_0^2}, \qquad (25)$$

then the beam, curving toward the current, is transformed into a diverging beam, bounded by the surfaces (20) and (21) with the $+$ sign replaced by a $-$ sign. In contrast to the transformation of a beam of diamagnetic particles, a beam of paramagnets is not twisted as it scatters, but becomes non-uniform in density. At large r $(r \gg kQ)$ its density $\eta(\varphi)$ decreases with increasing φ

$$\eta(\varphi) \approx \eta(\varphi_Q) \left(\frac{\varphi_Q^2 - \pi^2}{\varphi^2 - \pi^2} \right)^{3/2}. \qquad (26)$$

In the transformation of a paramagnet beam in the case $\alpha/(p^2 v_0^2) > 1$ Eq. (17) leads to trajectories

$$r = \frac{2p\lambda}{e^{\lambda\varphi} - e^{-\lambda\varphi}} \qquad (27)$$

with

$$\lambda = \sqrt{\frac{\alpha}{p^2 v_0^2} - 1}.$$

Moving along a helical trajectory, the paramagnetic particles (and consequently the entire beam) are trapped by the current. Trapping of a particle with offset distance p takes place during a time

$$T = \frac{1}{v_0} (\sqrt{l^2 + \lambda^2 p^2} - \sqrt{a^2 + \lambda^2 p^2}). \qquad (28)$$

TRANSFORMATION OF FERROMAGNETIC PARTICLE BEAMS

We represent the magnetization $\mathbf{J}$ of ferromagnetic particles, as in Ref. 3, as consisting of a remanent component $\mathbf{J}_0$ and an induced component $\mathbf{J}_i$, both directed along the magnetic field $\mathbf{H}$ of the linear current:

$$\mathbf{J} = \mathbf{J}_0 + \mathbf{J}_i = J_0 \frac{\mathbf{H}}{|H|} + \frac{3}{4\pi} \frac{\mu - 1}{\mu + 2} \mathbf{H}. \qquad (29)$$

We assume that during the transformation of the beams J_0 and μ remain constant. Then the magnetic moment of a particle is expressed by

$$\mathbf{M} = \frac{4}{3} \pi R^3 J_0 \frac{\mathbf{H}}{|H|} + R^3 \frac{\mu - 1}{\mu + 2} \mathbf{H}. \qquad (30)$$

We restrict the discussion to transformations of ferromagnet beams through the remanent magnetization when $J_0 \gg J_i$ and to transformations through the induced magnetization when $J_i \gg J_0$. In transformations of beams in the remanent magnetization we neglect the induced magnetization. Motions of such particles in the case of a longitudinal transformation are described by the differential equation

$$\ddot{r} = -\beta r^{-2}, \qquad (31)$$

where

$$\beta = \frac{6IJ_0}{c\rho}, \qquad (32)$$

and Eq. (8). Integrating Eq. (31) for $t_0 = 0$, $r_0 = p$, and $\dot{r}_0 = 0$, we find the equation of motion of the particle in the plane $(\varphi = \mathrm{const})$ of its motion

$$r \sqrt{\frac{p}{r} - 1} + 2p \arctan \sqrt{\frac{p}{r} - 1} = \sqrt{\frac{2\beta}{p}} t. \qquad (33)$$

This equation, together with Eq. (8), defines the trajectory of the particle

$$r \sqrt{\frac{p}{r} - 1} + 2p \arctan \sqrt{\frac{p}{r} - 1} = \sqrt{\frac{2\beta}{p}} \frac{z}{v_0}. \qquad (34)$$

Equation (34) shows that the particle trajectory is bent toward the current and the particle in moving along it according to Eq. (33) is trapped by the current. We obtain the trapping time T of a particle with offset distance p directly from Eq. (33) by setting $r = a$,

$$T = \sqrt{\frac{p}{2\beta}} \left(a \sqrt{\frac{p}{a} - 1} + 2p \arctan \sqrt{\frac{p}{a} - 1} \right). \qquad (35)$$

Trapping of particles in the z coordinate takes place in the interval $z_q \leq z \leq z_Q$, where

$$z_\gamma = v_0 \sqrt{\frac{\gamma}{2\beta}} \left(a \sqrt{\frac{\gamma}{a} - 1} + 2\gamma \arctan \sqrt{\frac{\gamma}{a} - 1} \right), \qquad (36)$$

and γ, as above, takes the values q and Q. The particle trapping density $\eta(z)$ decreases with increasing z in the rough approximation

$$\eta(z) \approx \eta(z_q) \sqrt[3]{\frac{z_q}{z}}. \qquad (37)$$

For transverse transformation of a ferromagnet beam through the remanent magnetization, motions of the particles in the planes $z = \mathrm{const}$ are described by differential equation (16) and

$$\ddot{r} - r\dot{\varphi}^2 = -\beta \dot{r}^{-2}. \qquad (38)$$

We find the differential equation of the trajectory by eliminating t from Eqs. (16) and (38),

$$\frac{d^2(1/r)}{d\varphi^2} + \frac{1}{r} = \frac{\beta}{p^2 v_0^2}. \qquad (39)$$

We find the particle trajectory by integrating Eq. (39) under the condition that $t_0 = 0$, $\varphi_0 \simeq 0$, $(1/z_0) \simeq 0$, $\dot{r}_0 \simeq -v_0$, and $r_0^2 \dot{\varphi}_0 = p v_0$ (according to Eq. (16)),

$$r = \frac{P}{1 + e \sin(\varphi + \vartheta)}, \qquad (40)$$

where

$$P = \frac{p^2 v_0^2}{\beta}, \qquad (41)$$

$$e = \sqrt{1 + \frac{p^2 v_0^4}{\beta^2}}, \qquad (42)$$

$$\tan \vartheta = -\frac{\beta}{pv_0^2}. \tag{43}$$

The curve (40) is a hyperbola subtending the current axis with minimum distance of approach r_m

$$r_m = \frac{P}{1+e}. \tag{44}$$

If $r_m < a$, then the particle becomes trapped by the current. The condition of trapping of the entire beam is[3]

$$\frac{Q^2 v_0^2}{\beta + \sqrt{\beta^2 + Q^2 v_0^4}} < a. \tag{45}$$

We obtain the trapping time T of a particle with offset distance p for $r_0 = l$ ($l \gg Q$) by integrating Eq. (31). To first order

$$T \approx \frac{1}{v_0} \left(l + \frac{\beta}{v_0^2} \ln \frac{p + a + \dfrac{\beta}{v_0^2}}{2l + \dfrac{\beta}{v_0^2}} \right). \tag{46}$$

If, however,

$$a < \frac{q^2 v_0^2}{\beta + \sqrt{\beta^2 + q^2 v_0^4}}, \tag{47}$$

then the beam is transformed into a diverging beam, contained within the angular interval $\varphi_Q \leq \varphi \leq \varphi_q$, whose endpoints are given by the formula

$$\varphi_\gamma = \pi + \arcsin \frac{1}{\sqrt{1 + \dfrac{\gamma^2 v_0^4}{\beta^2}}} + \arctan \frac{\beta}{\gamma v_0^2} \tag{48}$$

with γ replaced by Q and q, respectively. The transformed beam is nonuniform in density $\eta(\varphi)$ and at large r ($r \gg Q$) decreases in density with increasing scattering angle. Thus, in a weak ($\beta \ll pv_0^2$) field

$$\eta(\varphi \approx \eta(\varphi_Q)) \frac{\cos \varphi}{\cos \varphi_Q}, \tag{49}$$

and in a strong field ($\beta \gg pv_0^2$)

$$\eta(\varphi) \approx \eta(\varphi_Q) \frac{p^2 + \dfrac{\beta^2}{v_0^4}}{Q^2 + \dfrac{\beta^2}{v_0^4}}. \tag{50}$$

In the calculation of transformations of ferromagnet beams through the induced magnetization, we have neglected the remanent magnetization. Thus, with allowance for the adopted simplifications the mathematical formalism for describing the transformation of ferromagnet beams remains the same, and with the same notation, as for paramagnet beams. Thus, for longitudinal transformation all formulas obtained for paramagnet beams, including (9)–(13) and

those following from them, remain valid for ferromagnet beams. The same applies for transverse ferromagnet beams, in particular formulas (23)–(28) and their consequences.

However, the efficiency of transformation of ferromagnet beams differs substantially from that for paramagnet beams due to differences in them of several orders of magnitude in $\mu - 1$ and also by the presence of remanent magnetization in ferromagnets. A rough calculation shows that transformations of dia- and paramagnet beams by means of the magnetic field of a linear current are feasible at currents of the order of tens and hundreds of thousands of amperes, and for ferromagnet beams — at tens and hundreds of amperes. Therefore the application of the above transformations of dia- and paramagnet beams is hindered in practice and may be realized only at very high currents, for example in powerful line discharges and short-circuits. On the other hand, transformations of ferromagnet beams by the magnetic field of a linear current may be implemented under ordinary laboratory and production conditions.

Note that the dimensions of the particles do not enter into the equations of motion of magnetic particles and their trajectories in the magnetic field of a linear current; therefore their beam transformations do not depend on particle dimensions. In addition, as in Refs. 4 and 5, we may conclude that the transformations of ferromagnet beams through the remanent magnetization depend on the shape of the particle, and in the remaining cases (transformations of dia- and paramagnet beams through the induced magnetization) the shape of the particles only weakly influences the transformations.

In conclusion I would like to point out that the calculations presented here involve a number of major simplifications and idealizations. Thus, in reality for ferromagnets μ depends on $\mathbf{H}$, their magnetization process is hysteretic, and the remanent magnetization $\mathbf{J}_0$ varies with the magnetic field. Nevertheless, the results of our idealized calculations should prove helpful (as first approximations) for more accurate calculations of transformations of magnetic beams by the magnetic fields of constant linear currents.

---

[1] The width of the ribbon beam is arbitrary, but such that edge effects due to the finite length of the linear current do not affect the transformations of the beam.

[2] For $q^2 - a^2 < \alpha/v_0^2 < Q^2 - a^2$ partial trapping takes place.

[3] If $(Q^2 v_0^2/\beta + \sqrt{\beta^2 + Q^2 v_0^4}) > a > (q^2 v_0^2/\beta + \sqrt{\beta^2 + q^2 v_0^4})$, then partial trapping takes place.

---

[1] S. V. Izmailov, *Course in Electrodynamics* [in Russian], Uchpedgiz, Moscow, Leningrad (1952).

[2] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, Oxford, 1960) [Russian original, GIFML, Moscow, 1959].

[3] N. I. Shtepa, Zh. Tekh. Fiz. **49**(9), 1839 (1979) [Sov. Phys. Tech. Phys. **24**, 1034 (1979)].

[4] N. I. Shtepa, Zh. Tekh. Fiz. **63**(2), 182 (1993) [Tech. Phys. **38**, 147 (1993)].

[5] N. I. Shtepa, Zh. Tekh. Fiz. **65**(2), 203 (1995) [Tech. Phys. **40**, 224 (1995)].

Translated by Paul F. Schippnick