ERRATA

Erratum: Parametric excitation of spin waves in uniaxial ferrites [Tech. Phys. 43, 539 (1998)]

A. V. Nazarov and A. G. Gurevich
Zh. Tekh. Fiz. **69**, 143 (August 1999)

[S1063-7842(99)03008-1]

Formula (9) should read:

$$M_z = M_0 - \frac{m_+ m_-}{2M_0}.$$

Formula (12) should read:

$$\omega_H = \gamma H_{c0} \cos(\theta_H - \theta_0).$$

Formula (24) should read:

$$Y_k = \lambda_k^2 \left(-\omega_M \frac{k_+ k_z}{k^2} - i \frac{\omega_M}{4\pi} N_{23}^a \right) a_0^+ + \mu_k^2 \left(-\omega_M \frac{k_+^* k_z}{k^2} + i \frac{\omega_M}{4\pi} N_{23}^a \right) a_0^{-*} - \lambda_k \mu_k$$
$$\times \left(2\gamma h_z - a_0^+ \omega_M \frac{k_+ k_z}{k^2} - a_0^{-*} \omega_M \frac{k_+ k_z}{k^2} + i \frac{\omega_M}{2\pi} N_{23}^a (a_0^+ - a_0^{-*}) \right).$$

Formula (27) should read:

$$W_k = \frac{1}{4}(A_k + \xi_a + \omega_k) \left(2 \exp(i\varphi_k) \sin 2\theta_k + i \frac{N_{23}^a}{\pi} \right) a_{01} + \frac{1}{4}(A_k + \xi_a - \omega_k)$$
$$\times \frac{[B_k \exp(2i\varphi_k) - \eta_a]^2}{|B_k \exp(2i\varphi_k) - \eta_a|^2} \left(2 \exp(-i\varphi_k) \sin 2\theta_k - i \frac{N_{23}^a}{\pi} \right) a_{02} + [B_k \exp(2i\varphi_k) - \eta_a]$$
$$\times \left[\frac{\cos\theta_h}{\omega_M} - \frac{a_{01}}{2} \exp(-i\varphi_k) \sin 2\theta_k - \frac{a_{02}}{2} \exp(i\varphi_k) \sin 2\theta_k + i \frac{N_{23}^a}{2\pi} (a_0^+ - a_0^{-*}) \right].$$

Formula (30) should read:

$$h_{\text{thr}} = \min \left\{ \frac{\Delta H_k}{4\pi M_0 \sin^2 \theta_k \exp(2i\varphi_k) + 2H_{\text{A1}} \sin^2 \theta_0} \right\} \frac{\omega_k}{\cos\theta_h}.$$

Translated by R. M. Durham

Interaction of contact electrohydrodynamic waves in current-carrying semiconductors

R. A. Brazhe and T. A. Novikova

Ulyanovsk State Technical University, 432027 Ulyanovsk, Russia
(Submitted July 23, 1997; resubmitted March 24, 1998)
Zh. Tekh. Fiz. **69**, 1–4 (August 1999)

It is shown that contact electrohydrodynamic waves similar to the waves on an interface between two immiscible liquids can exist on an interface between two identical semiconductors with different concentrations of free charge carriers. When a current flows, instabilities of such waves, particularly an analog of the Kelvin–Helmholtz hydrodynamic instability, appear in one of the layers. The possibility of practical utilization of contact waves interacting with currents is discussed. © *1999 American Institute of Physics.* [S1063-7842(99)00108-7]

INTRODUCTION

Let us consider a structure consisting of two semiconductor layers with different concentrations of free charge carriers. Such a structure can be obtained, for example, by the autoepitaxial growth of a high-resistance overlayer on a semiconductor substrate.[1] It was shown in Ref. 2 that contact electrohydrodynamic waves of free charge carriers, similar to the waves on an interface between two immiscible liquids,[3] can propagate in the structure under consideration.

The ensemble of free charge carriers in a semiconductor can be regarded as a charged, weakly compressible quasiliquid, if 1) the semiconductor is monopolar and nondegenerate; 2) the free carriers are thermally excited and are characterized by a constant and isotropic effective mass m^*; 3) their motion in a transverse electric field is collisionless (conforms to a ballistic regime[4]), which is realized when the thicknesses of the high-resistance layer (d_1) and the low-resistivity layer (d_2) are smaller than the so-called ballistic length L_b,

$$d_{1,2}<L_b, \quad L_b=\tau_s\left(\frac{2\hbar\omega_0}{m^*}\right)^{1/2}, \tag{1}$$

where $\tau_s=10^{-12}-10^{-13}$ s is the spontaneous emission time for optical phonons and ω_0 is their cutoff frequency; 4) the junction layer thicknesses $d_{1,2}$ exceed the Debye screening radius

$$d_{1,2}>r_D, \quad r_D=\left(\frac{\varepsilon_0\varepsilon k_BT}{n_{1,2}e^2}\right)^{1/2}, \tag{2}$$

and 5) the velocity distribution of the free carriers is insignificant. In such a case they are located near the bottom of the conduction band (electrons) or the top of the valence band (holes), and their velocity $v=(2E/m^*)^{1/2}\approx$ const. Therefore, when a current of majority carriers flows in one of the semiconductor layers, the behavior of the charged particles in the contact region can be described as the relative motion of two liquids of different density. This paper discusses the possibilities of the appearance of hydrodynamic instabilities in such a structure when the contact electrohydrodynamic waves interact with the current of free charges.

THEORETICAL MODEL

Figure 1 shows a structure consisting of two semiconductor layers with different electron concentrations. The contact potential difference in an $n-n^+$ junction equals[5]

$$U_c=\frac{k_BT}{e}\ln\frac{n_2}{n_1}, \tag{3}$$

where n_1 and n_2 are the electron concentrations in the high- and low-resistivity layers ($n_2>n_1$).

Electrohydrodynamic waves of free charges can be excited by an external source in the junction region. Their frequencies ω lie in the range[2]

$$\omega_c\leqslant\omega\leqslant\omega_d; \tag{4}$$

$$\omega_c=\frac{en_{1,2}\mu}{\varepsilon_0\varepsilon}, \quad \omega_d=\frac{c_0^2}{D}. \tag{5}$$

Here ω_c is the Maxwell relaxation frequency, ω_d is the electron diffusional frequency, c_0 is the velocity of propagation of a wave disturbance, $D=\mu k_BT/e$ is the diffusion coefficient, and μ is the electron mobility in the semiconductor.

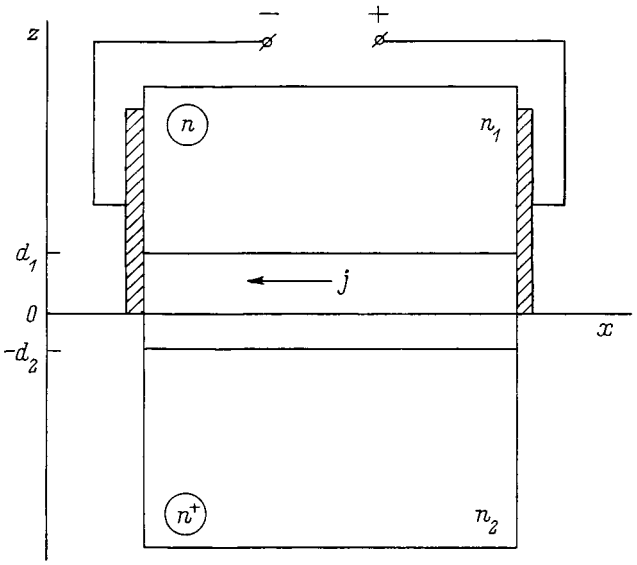

FIG. 1. Semiconductor model investigated.

Let us consider cadmium sulfide (CdS) at $T=300$ K as an example of an electronic semiconductor. The upper layer of the structure used, which was grown epitaxially, has a higher resistivity than does the lower layer: $\varrho_1=5.0\times10^{-2}\ \Omega\cdot$m and $\varrho_2=1.0\times10^{-2}\Omega\cdot$m. The contact potential difference (3) across the boundary between the two layers is roughly $U_c\simeq0.04$ V. The total thickness of the $n-n^+$ junction under these conditions is $d\simeq(2\varepsilon_0\varepsilon\mu\varrho_1U_c)^{1/2}=0.11\,\mu$m [in CdS $\mu=0.035\text{m}^2/(\text{V}\cdot\text{s})$], the thicknesses of the transition regions in the upper and lower layers being $d_1=\varrho_1 d/(\varrho_1+\varrho_2)=0.09\ \mu$m and $d_2=\varrho_2 d/(\varrho_1+\varrho_2)=0.02\,\mu$m, respectively. According to (1) and (2), the ballistic length in CdS is $L_b=0.12\ \mu$m, the Debye screening radius in the upper transition layer is $r_{D1}\simeq0.06\ \mu$m, and the value in the lower layer is $r_{D2}\simeq0.027\ \mu$m.

As the above estimates show, the condition for quasihydrodynamic phenomena is satisfied in the structure under consideration, since the thicknesses of the junctions in the layers are smaller than the ballistic length and greater than the Debye radius (or approximately equal to it).

We shall assume everywhere below that a longitudinal current of free electrons with a density $j=en_1u_d$ (u_d is the mean electron drift velocity) flows in the upper layer of the semiconductor. We can use the equations of hydrodynamics to determine how this current influences the propagation of contact electrohydrodynamic waves and ascertain the types of instabilities appearing.

DISPERSION RELATION

Let us use the model adopted above to examine the behavior of two ideal incompressible electron quasiliquids in the transition region of the semiconductor, the lower of which (with a large concentration of free charges) is static and occupies the region $z<0$, while the upper liquid moves with the velocity u_d (Fig. 1).

In the hydrodynamic approximation the system of equations describing the medium under consideration[3] includes the equations of motion of the particles comprising the electron quasiliquids of density $n_{1,2}m^*$ in each of the layers

$$\frac{\partial u_{1,2}}{\partial t}+u_{1,2}\frac{\partial u_{1,2}}{\partial x}+u_{1,2}\frac{\partial u_{1,2}}{\partial z}+U_{1,2}\frac{\partial u_{1,2}}{\partial x}=-\frac{1}{n_{1,2}m^*}\frac{\partial p_{1,2}}{\partial x}, \tag{6}$$

$$\frac{\partial v_{1,2}}{\partial t}+u_{1,2}\frac{\partial v_{1,2}}{\partial x}+v_{1,2}\frac{\partial v_{1,2}}{\partial z}+U_{1,2}\frac{\partial v_{1,2}}{\partial x}=-\frac{1}{n_{1,2}m^*}\frac{\partial p_{1,2}}{\partial z}-\frac{eE_{1,2}}{m^*}, \tag{7}$$

the continuity equation

$$\frac{\partial u_{1,2}}{\partial x}+\frac{\partial v_{1,2}}{\partial z}=0, \tag{8}$$

the condition for the pressures at the interface

$$p_1=p_2, \tag{9}$$

the boundary condition for the velocities

$$v_1|_{z=d_1}=0,\quad v_2|_{z=-d_2}=0, \tag{10}$$

and the evolution relation for the displacement $\eta(x,t)$ of the interface ($z=0$)

$$\frac{\partial\eta}{\partial t}+U_{1,2}\frac{\partial\eta}{\partial x}=v_{1,2}-u_{1,2}\frac{\partial\eta}{\partial x}. \tag{11}$$

The following notation is used in the equations just written down: $u_{1,2}$ and $v_{1,2}$ are the horizontal and vertical components of the particle velocities in the n layer (with the subscript 1) and n^+ layer (with the subscript 2); $U_{1,2}$ are the relative velocities of the electron quasiliquids, where $U_2\equiv0$ and U_1 is equal to the electron drift velocity [$U_1=u_d=j/(en_1)$]; and $E_{1,2}$ are the contact electric field intensities in the transition layers of the semiconductor.

A disturbance of the interface of the $n-n^+$ junction propagates in the form of plane traveling waves. Let us obtain the dispersion relation for such waves in the linear approximation. We seek the solution of Eq. (11) together with (6)–(10) in the form

$$\eta=\eta_0\exp[i(kx-\omega t)], \tag{12}$$

$$u_{1,2}=A_{1,2}\cosh[k(z\mp d_{1,2})]\exp[i(kx-\omega t)], \tag{13}$$

$$v_{1,2}=-iA_{1,2}\sinh[k(z\mp d_{1,2})]\exp[i(kx-\omega t)], \tag{14}$$

$$p_{1,2}=\frac{n_{1,2}m^*(\omega-kU_{1,2})}{k}u_{1,2}-n_{1,2}m^*\frac{eE_{1,2}}{m^*}z, \tag{15}$$

where k is the wave number.

Substituting expressions (12)–(15) into (11) and allowing for the boundary condition (9), we obtain the dispersion relation for small fluctuations of the interface between the electron quasiliquids:

$$(\omega-u_dk)^2n_1\tanh kd_2+\omega^2n_2\tanh kd_1=(n_2-n_1)k\omega\tanh kd_1\tanh kd_2. \tag{16}$$

Here $\omega=eE_{\max}/m^*$ is the effective acceleration of electrons in the contact electric field, whose maximum intensity is given by the expression $E_{\max}=(E_{1,2})_{\max}=en_1d_1/(\varepsilon\varepsilon_0)$. We write the solution of the dispersion relation

$$\omega_{1,2}=k\frac{u_dn_1\tanh kd_2\pm(\tanh kd_1\tanh kd_2[(n_2-n_1)\omega N_{1,2}/k-n_1n_2u_d^2])^{1/2}}{N_{1,2}}, \tag{17}$$

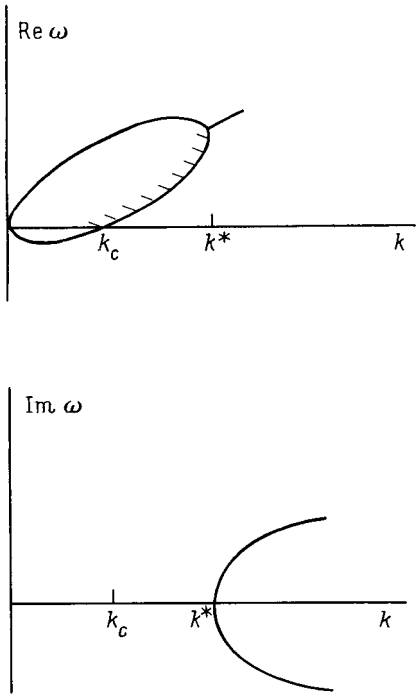

FIG. 2. Dispersion characteristics of contact electrohydrodynamic waves in an $n-n^+$ junction.

where $N_{1,2}=n_1 \tanh kd_2+n_2 \tanh kd_1$.

ANALYSIS OF INSTABILITIES

In the long-wavelength approximation the dispersion relation (16) takes the form

$$a(\omega-u_d k)^2+b\omega^2-b(1-a)\omega k^2 d_2=0, \tag{18}$$

where $a=n_1/n_2$ and $b=d_1/d_2$.

Its solution (17) is also modified:

$$\omega_{1,2}=k\frac{u_d\pm([(1-a)(1+b/a)\omega d_2-u_d^2]b/a)^{1/2}}{1+b/a}. \tag{19}$$

The dispersion curve $\omega(k)$ is presented in Fig. 2. The upper branch of the dispersion curve corresponds to a positive-energy wave [the plus sign in (19)], and the lower branch corresponds to a negative-energy wave [the minus sign in (19)], whose domain is located between the critical points $\omega=0$ and $v_{gr}=\infty$ (v_{gr} is the group velocity).[6] The coupling of negative- and positive-energy waves gives rise to an instability after the bifurcation point $v_{gr}=\infty$. In this region the frequency of the wave becomes complex. In hydrodynamics a similar instability of internal waves in the shear flow of a liquid is known as the Kelvin–Helmholtz instability.[7]

Figure 2 shows the region in which the existence of stable negative-energy electrohydrodynamic waves is possible. It corresponds to the range of wave numbers satisfying the inequality

$$k_c<k<k^*. \tag{20}$$

Here k_c corresponds to the condition $\omega=0$, and k^* corresponds to the point where $v_{gr}=\infty$. At this point a region of Kelvin–Helmholtz instability is adjacent to the region of stable waves. Contact waves of either type (both positive- and negative-energy waves) will be unstable, if their wave number $k>k^*$ or if the electron drift velocity exceeds u_d^*:

$$u_d^*=[\omega d_2(1-a)(1+b/a)]^{1/2}. \tag{21}$$

In other words, electrohydrodynamic waves are unstable when the current density satisfies the inequality

$$j>en_1u_d^*. \tag{22}$$

The condition for the wave number of negative-energy waves (20) corresponds to the following range of electron drift velocities:

$$u_d^c<u_d<u_d^*. \tag{23}$$

In the present case the value of u_d^c is given by the expression

$$u_d^c=u_d^*(1+b/a)^{-1/2}. \tag{24}$$

We present some numerical estimates for the two-layer CdS structure having the parameters indicated above. The electron drift velocities corresponding to the condition for the excitation of negative-energy contact electrohydrodynamic waves in the semiconductor lie in the range $9.5\times10^4<u_d<4.5\times10^5$ m/s. This corresponds to a range of intensities of the external longitudinal electric field from $E^c\simeq2.7\times10^6$ V/m to $E^*\simeq1.3\times10^7$ V/m, i.e., below the breakdown value. According to (22), when $E>E^*$ (or $u_d>4.5\times10^5$ m/s) the contact waves experience Kelvin–Helmholtz instability.

CONCLUSION

We note that the range of frequencies of contact electrodynamic waves specified by the condition (4) for CdS extends from $f_{min}=1.9\times10^{11}$ Hz to $f_{max}=1.5\times10^{12}$ Hz. The phase velocity of linear contact waves

$$c_0=\left(\frac{1-a}{a+b}b\omega d_2\right)^{1/2} \tag{25}$$

amounts to 9.3×10^4 m/s, which corresponds to a wavelength $\lambda=0.093$ μm and a frequency $f=10^{12}$ Hz.

Thus contact electrohydrodynamic waves with microwave frequencies can be excited by properly selecting the parameters of a two-layer semiconductor structure. The propagation conditions for these waves can be varied by varying the current of charge carriers in one of the layers. Positive- and negative-energy contact waves are excited in a definite range of electron drift velocities and exhibit instability at $u_d>u_d^*$ [see (22)].

The growth of instabilities within an $n-n^+$ junction is interesting from the standpoint of the possibility of finding a regime of amplification of contact electrohydrodynamic waves in their interaction with the current of majority carriers in the high-resistance layer of the semiconductor.

Signal amplifiers (and generators) based on the principles described can be promising for work in the terahertz

frequency range, which has been exploited to only a small extent. The development of classical microwave electronic devices (klystrons, magnetrons, traveling-wave tubes, etc.) runs up against difficulties associated with their small geometric dimensions. Acoustoelectronic devices cannot operate in this range because of the excessively strong damping of acoustic waves. Free-electron lasers are still the only optoelectronic devices which permit the generation and amplification of signals at such frequencies, but they require cumbersome, expensive, and risky accelerator techniques. Thus, semiconductor structures with contact electrohydrodynamic waves of free carriers can offer a starting point for the development of terahertz electronics.

[1] S. M. Sze (ed.), *VLSI Technology* [McGraw-Hill, New York (1983); Mir, Moscow (1986), Vol. 1, 404 pp.].
[2] R. A. Brazhe and V. V. Sadulin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **40**, 1164 (1997).
[3] L. A. Ostrovskiĭ and Yu. A. Stepanyants, Izv. Akad. Nauk SSSR Ser. Mekh. Zhidk. Gaza, No. 4, 63 (1982).
[4] Yu. A. Ermolaev and A. L. Sanin, *Electronic Synergetics* [in Russian], Izd. LGU, Leningrad (1988), 248 pp.
[5] P. T. Oreshkin, *Physics of Semiconductors and Dielectrics* [in Russian], Vysshaya Shkola, Moscow (1977), 448 pp.
[6] L. A. Ostrovskiĭ, S. A. Rybak, and L. Sh. Tsimring, Usp. Fiz. Nauk **150**, 417 (1986) [Sov. Phys. Usp. **29**, 1040 (1986)].
[7] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* [Pergamon Press, London; Addison–Wesley, Reading, Mass. (1959); Gostekhizdat, Moscow (1953)].

Translated by P. Shelnitz

Investigation of the electromagnetic fields at a jump in the dielectric constant

I. P. Kozlov

Moscow Aviation Institute, Scientific-Research Institute of Applied Mechanics and Electrodynamics, 125871 Moscow, Russia

(Submitted June 24, 1997; resubmitted October 20, 1998)
Zh. Tekh. Fiz. **69**, 5–9 (August 1999)

The passage of a plane wave through an inhomogeneous flat insulator layer of arbitrary thickness without absorption is considered. A method is given for solving the problem in terms of elementary functions, which reduces the number of independent parameters of the layer. A similarity principle for layers having equivalent reflectivities is described. It is shown that the electric field intensity can increase to infinity near the critical point where $\varepsilon=0$.
 [S1063-7842(99)00208-1]

There are several bottlenecks to research on the passage of electromagnetic waves in a plane-layer medium without absorption in the vicinity of a zero of the dielectric constant. For example, as was shown in Ref. 1, in the case of a linear function $\varepsilon(z)$ the requirement that the solution vanish at infinity is incompatible with the requirement that the field be finite at the zero of ε. A singular plane on which the magnetic field $H=0$ can be identified near the point where $\varepsilon=0$. It was demonstrated in Ref. 2 that the behavior of E remains unchanged in character near the zero of ε when the variation of ε deviates from a linear dependence. Investigations of the solution for a linear lossless layer in terms of Airy functions[3] show that the electric field (E) is greater than zero in the region where $\varepsilon<0$, i.e., a wave penetrates into that region. We previously[4] refined this solution and showed on a model of a layer of arbitrary thickness with $\varepsilon>0$ that a wave does not pass into the region where $\varepsilon<0$. A critical point corresponding to $\varepsilon=0$, which is associated with a qualitative dependence of the solution on small changes in the parameters of the medium, was identified.

Because of the critical point corresponding to $\varepsilon=0$, the exact solution of the mathematical problem of the passage of a normally incident plane electromagnetic wave through an inhomogeneous flat insulator layer without absorption takes on special significance. This basic solution from a bifurcation set of solutions corresponds to the limiting case of $\psi_1\Rightarrow 0$, $\Theta_1\Rightarrow 0$, $r\Rightarrow\infty$, $\varepsilon''\Rightarrow 0$, and $H^0\Rightarrow 0$, where ψ_1 is the angular width of the flow of electromagnetic energy (the beam) impinging on the layer, Θ_1 is the angle of incidence of the beam on the layer, r is the radius of curvature of the $\varepsilon=\text{const}$ level of the inhomogeneous medium, H^0 is the amplitude of the constant magnetic field, and $\varepsilon''=\text{const}$ is the imaginary part of the dielectric constant for $\varepsilon>0$.

This paper describes an investigation of the fields at a jump in ε assigned by a flat linear layer of arbitrary thickness, which continues as a nonlinear layer with continuity of the function $\varepsilon(z)$ in the system, so that the phase shift of a wave passing through the layer $\Delta\phi\ll 1$ and $\varepsilon>0$ (in this case the conservation law holds for the fields). The nonlinear portion, which plays an auxiliary role, is employed to eliminate the reflected wave from the edge of the linear layer, and the function $\varepsilon(z)$ in the nonlinear portion is chosen on considerations of simplicity in obtaining a solution. Treating the case of normal incidence does not entail a loss of generality, since the case of oblique incidence can be reduced to normal incidence simply by a change of variables.[3] Moreover, the ''resonance'' properties of the singular point at $\varepsilon=0$ are most pronounced under the presumed condition $\varepsilon''\ll\varepsilon$.

In accordance with the proposed solution method, a continuously inhomogeneous layer can be represented by a set of homogeneous sublayers of variable thickness, so that in adjacent sublayers $\delta^2\equiv\varepsilon_n/\varepsilon_{n+1}=\text{const}$ and $(\delta^2-1)\ll 1$ (Ref. 4). The employment of finite differences with a variable step for the argument, i.e., the sublayer thickness, which is matched to the variation of the wave impedance of the medium, permits reduction of the number of independent parameters of the layer. The theoretical investigations are thereby simplified, and an efficient calculation algorithm can be devised.

1. Let a plane wave impinge normally on an inhomogeneous flat layer $(\varepsilon_1,\varepsilon_N)$ of a lossless insulator of arbitrary thickness extending from z_1 to z_N in the case of $\Delta\phi\ll 1$ and $\varepsilon>0$ (Fig. 1). The layer consists of two parts: a linear part from z_1 to z_L (in region *2*) with $\varepsilon(kz)=\varepsilon(kz_1)+\alpha_1 k(z-z_1)$ and a nonlinear part from z_L to z_N (in region *3*) with $gr(\varepsilon)\equiv\alpha/(\varepsilon)^{3/2}=gr(\varepsilon_L)=\text{const}$, where $\alpha\equiv d\varepsilon/d(kz)$, $\alpha_1<0$, $k=2\pi/\lambda$, and λ is the wavelength in free space. In the layer the function $\varepsilon(kz)$ is continuous along with its derivative and $|gr(\varepsilon_1)|\gg 1$ (since $\Delta\phi\ll 1$). In front of the layer (in region *1*) at $z<z_1$ (where $\varepsilon=\varepsilon_1$) and beyond the layer (in region *4*) at $z>z_N$ (with $\varepsilon=\varepsilon_N$) the medium is homogeneous, and there is only a transmitted wave beyond the layer.

Let us assign the solution of the wave equation

$$\frac{d^2E}{d(kz)^2}+\varepsilon(kz)E=0 \tag{1}$$

in the homogeneous sublayers (in regions *2* and *3*), whose thicknesses are specified by the condition $\delta^2=\varepsilon_n/\varepsilon_{n+1}=\text{const}$, in the form

$$E_x^n=A^n\exp(-i\sqrt{\varepsilon_n}kz)+C^n\exp(i\sqrt{\varepsilon_n}kz), \tag{2}$$

where $n=1,2,\ldots,N$; $A^1=1$, and $C^N=0$.

Then, after satisfaction of the boundary conditions on the edges of the homogeneous sublayers, for the reflectivities $R^n=C^n/A^n$ and the wave amplitudes A^n we obtain the following expressions:

$$R^n\exp\{2i\rho_n\}=R_0^n=\frac{r_n+R_0^{n+1}\exp\{i\Delta\rho_{n+1}\}}{1+r_nR_0^{n+1}\exp\{i\Delta\rho_{n+1}\}}, \tag{3}$$

$$A^{n+1}=\exp\left\{-i\left(1-\frac{1}{\delta}\right)\sum_{p=1}^{n}\rho_p\right\}P^{n+1},$$

$$P^n=\prod_{m=1}^{n-1}\left[1+\frac{\delta-1}{2}(1-R_0^m)\right], \tag{4}$$

where

$$n=1,\ldots,N-1,\quad R_0^N=0,\quad \Delta\rho_N=0,$$

$$\rho_n=\sqrt{\varepsilon_n}kz_n,\quad \varepsilon_L/\varepsilon_1=\delta^{-2(L-1)},$$

$$r_n=\frac{\delta-1}{\delta+1};\quad \Delta\rho_{n+1}=\frac{2\cdot(\delta^2-1)}{\nu_n gr(\varepsilon_1)};$$

$$\nu_n=\begin{cases}\delta^{3n} & \text{for } n<L,\\ 1 & \text{for } n>L.\end{cases} \tag{5}$$

We note that the reflectivity R_0^n is assigned in a local coordinate frame with a center 0_n, which is specified by the condition $z_n=0$. In the recurrence relations (3) the multiplier $\rho_n=\sqrt{\varepsilon_n}kz_n$ has been moved over to the left-hand side, and $gr(\varepsilon_n)$ is represented in terms of $gr(\varepsilon_1)$ (owing to the partitioning method used); therefore, R_0^n does not depend explicitly on ε_n and kz_n.

2. The solution (2)–(5) is exact for the discrete model of the medium. The finite-difference recurrence relations (3) reduce to a Riccati differential equation when $N\Rightarrow\infty$ ($\delta^2\Rightarrow 1$), demonstrating the transition to the continuous model of the medium. Also, since δ and N are interrelated by the equality $\varepsilon_1/\varepsilon_N=\delta^{2(N-1)}$, if the form of $\varepsilon(kz)$ is assigned, the reflectivity R_0^1 becomes a function of two variables as $N\Rightarrow\infty$: $R_0^1=f(gr(\varepsilon_1),(\varepsilon_N/\varepsilon_1))$. The solution in region *3* is easily obtained from the Riccati equation (it is given in the Appendix). Therefore, using (A5), we can take $N=L$ and $R_0^N=R_0$ in (3) (if $\varepsilon_N\Rightarrow 0$, then $R_0=\gamma+i\beta$).

3. The dependence of R_0^1 on two parameters allows us to describe a similarity principle for layers involving the identification of, for example, similar linear layers, in which these two parameters are equal, but the values of α_1 and ε_1 ($gr(\varepsilon_1)=\alpha_1/(\varepsilon_1)^{3/2}$) differ.[5] If $\Delta\phi\ll 1$ in the case under consideration, we can neglect the phase factor in the curly brackets in (4), i.e., we can take $A^n\cong P^n$, $n=1,2,\ldots,N$, and then in similar layers the reflectivities are equivalent and the wave amplitudes are equal. The similarity principle holds for two-part layers under the additional condition of equality of the values of $\varepsilon_L/\varepsilon_N$ for the two layers. Figure 1 shows two similar layers.

4. Let us perform some investigations of the solution for a two-part layer when $\varepsilon_N\Rightarrow 0$. In this case $R_0^1=f(gr(\varepsilon_1),\varepsilon_L/\varepsilon_1)$, and from (3) we find that $|R_0^1|\Rightarrow 1$ as $\varepsilon_L/\varepsilon_1\Rightarrow 0$. The latter is, of course, valid when $\varepsilon_1\Rightarrow 0$ ($\Delta\phi\Rightarrow 0$). Thus the transmission coefficient $S_N/S_1=1-|R_0|^2\Rightarrow 0$ as $\varepsilon_L/\varepsilon_1\Rightarrow 0$, and the flux density of the wave tends to vanish as the point where $\varepsilon=0$ is approached. Hence, $|H_y|\Rightarrow 0$ as $\varepsilon_L\Rightarrow 0$.

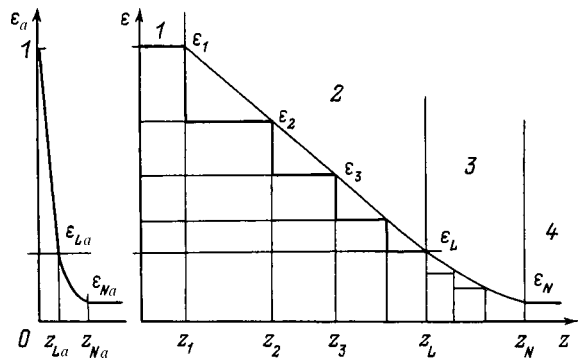

FIG. 1. Similar layers *a* and *b* have equivalent reflectivities when $gr(\varepsilon_{1a})=gr(\varepsilon_1)$, $\varepsilon_{1a}/\varepsilon_{Na}=\varepsilon_1/\varepsilon_N$, and $\varepsilon_{1a}/\varepsilon_{La}=\varepsilon_1/\varepsilon_L$, (the partitioning into homogeneous sublayers presumes that $\varepsilon_n/\varepsilon_{n+1}=\text{const}$).

Now let us determine the convergence of the product (4), taking $L\Rightarrow\infty$ ($\varepsilon_L\Rightarrow 0$) and using d'Alembert's convergence test for the corresponding series. Then, for sufficiently large *m* we find that

$$q=\frac{r_{m+1}}{r_m}[1+r_m(1+R_0^{m+1})]>1,$$

i.e., the product (4) diverges, and thus it is possible that $|E_x^L|\Rightarrow\infty$. Similarly, it was found from the equation for *H* in Ref. 6 that it is possible to have $|1/H_y|\Rightarrow\infty$ as $\varepsilon_L\Rightarrow 0$.

5. If we represent R_0^n in (3) in the form $R_0^n=|R_0^n|(1+i\Delta\phi^n)$ and $\Delta\phi^n\ll 1$, the variables can be separated and the solution, which does not depend on the form of $\varepsilon(kz)$ in this case, can be represented in the form

$$|R_0^n|=\frac{\sqrt{\varepsilon_1/\varepsilon_N}-1}{\sqrt{\varepsilon_1/\varepsilon_N}+1}. \tag{6}$$

In a first approximation formula (6) is valid for $|R_0^1|$, and formula (A9) is valid in the next approximation. The magnetic field intensity, like the reflectivity, does not depend very critically on $\varepsilon(kz)$, permitting the use of formula (A9) from the Appendix for a linear layer. However, the situation is different for the electric field *E*. For example, it follows from formula (A8) that *E* depends importantly on $\varepsilon(kz)$ in the layer. Therefore, it is best to use the relations (3)–(5) for calculating a linear layer.

6. We note the instability of the solution with respect to Θ in the half space $\varepsilon>0$. Going over to oblique incidence by performing the replacement $\varepsilon\Rightarrow\varepsilon-\varepsilon_1\sin^2\Theta_1$, we can easily show that an infinitesimally small value of $\Delta\Theta_1$ (which may be caused, for example, by nonparallelism of the sublayers) causes the direction Θ of the Poynting vector to tend to $\pi/2$ as $\varepsilon\Rightarrow\varepsilon_1\Delta\Theta_1^2$. Thus, the wave does not pass into the region where $\varepsilon<0$. The result which we obtained is consistent with the solution considered above for the case of $\Theta\equiv 0$ and refines the solution in terms of Airy functions that was given previously[3] for a linear layer. The approximate character of

the latter solution for $\alpha_1 \ll 1$ and $\varepsilon_1 = 1$ was noted long ago by Fock.[7] The error in the calculation of the electric field intensity in it increases with decreasing ε and, on the basis of energy arguments, is determined by $\varepsilon^{1/4}E$. In many practical cases, for example, in calculations of an atmospheric plasma, this error can be neglected.

7. From the two solutions obtained for a linear layer with $\varepsilon_L \Rightarrow 0$ it follows that $|R| \Rightarrow 1$ and thus $S_L/S_1 = 1 - |R|^2 \Rightarrow 0$. The similarity principle for linear layers allows us to extend the results of the investigation of the fields at a jump in ε to the vicinity of the point where ε equals zero and the function $\varepsilon(kz)$ changes sign. As a result, the exact solution enables us to identify a singular plane where $\varepsilon = 0$, so that $S_L/S_1 \Rightarrow 0$, $E \Rightarrow \infty$, and $H \Rightarrow 0$ as $\varepsilon \Rightarrow 0$. The point where $\varepsilon = 0$ is a critical point, near which the behavior of the solution depends qualitatively on small changes in the parameters of the medium.

The curvature of an $\varepsilon = \text{const}$ surface in the physical problem leads to ''reduction'' of that surface to a point as $\varepsilon \Rightarrow 0$, since we then have $\sqrt{\varepsilon} kr \Rightarrow 0$, which causes diffraction losses when $\sqrt{\varepsilon} kr \sim 1$. The occurrence of absorption eliminates this phenomenon and causes E to become finite at $\varepsilon = 0$ and the singular plane to ''spread'' into a layer, whose thickness (Δz) obeys the law $k\Delta z \sim |\mathrm{Im}(\varepsilon - i\varepsilon'')/\alpha_1|$. Depending on the degree of absorption, the wave will either propagate along the layer or pass into the region where $\varepsilon < 0$. A sharp increase in E can then lead to nonlinear effects.

Branching of the solution is possible in the vicinity of the point where $\varepsilon = 0$ (the solution is potentially multiple-valued). This is allowed by the similarity principle described above for the layers $(\varepsilon_1, \varepsilon_N)$, in which the reflectivities are equivalent and beyond which the fields are approximately equal. The solution of the mathematical problem under consideration is elementary, and it corresponds to the limiting transition, under which $\varepsilon'' \Rightarrow 0$, $\Theta \Rightarrow 0$, and $r \Rightarrow \infty$, in the physical problem of the passage of a wave through an inhomogeneous insulator layer. There is a bifurcation set of solutions, which branch off from the basic solution in response to small changes in the parameters.

8. When the fundamental parameter $|gr(\varepsilon)| \ll 1$, the short-wavelength asymptotic expression is valid. For example, a quantitative criterion for its applicability for a given reflectivity was put forth in Ref. 5. When $|gr(\varepsilon)| \gg 1$, the long-wavelength asymptotic expression is applicable for an assigned value of S/S_1, and in this case qualitatively different solutions can be described by the proposed method of solution. The familiar approximate solution for a linear layer in terms of Airy functions is one of the possible solutions.

9. Owing to the reduction of the number of independent parameters in the expression for the reflectivity, the proposed method of solution, in addition to analytical investigations, is effective in computational investigations associated with, for example, the design of antenna systems for spacecraft (see Ref. 8, Suppl. 2).

The results of computer calculations of the reflectivities R and the fields E are presented in Figs. 2 and 3 for values of $gr(\varepsilon_1)$ equal to 10 and 20, respectively. Here $\varepsilon_1 = 1$, z_0 corresponds to the point where $\varepsilon = 0$, the solid curves show the results of calculations for the linear layer, and the dashed curves show the results for the nonlinear layer under investigation. As follows from formula (4), in the range of variation of the parameters considered $A^n \cong P^n$; therefore, the solution does not depend on ε_1, i.e., $R = f(gr(\varepsilon_1), \varepsilon_1/\varepsilon_N)$ and

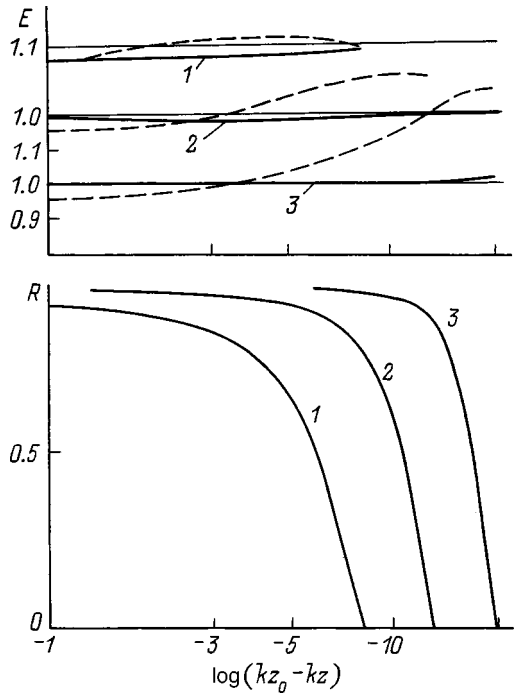

FIG. 2. Dependence of the reflectivity R and the electric field intensity E on the coordinate z for $gr(\varepsilon_1) = 10$ in inhomogeneous layers of various thickness: $\varepsilon_1/\varepsilon_N = 0.2 \times 10^4$ (*1*), 0.2×10^6 (*2*), and 0.2×10^9 (*3*).

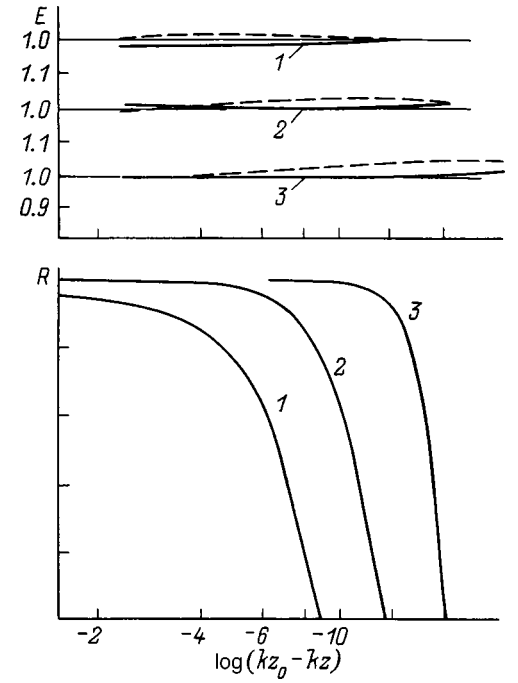

FIG. 3. Dependence of the reflectivity R and the electric field intensity E on the coordinate z for $gr(\varepsilon_1) = 20$ in inhomogeneous layers of various thickness (*1–3* — same as in Fig. 2).

$E=f(gr(\varepsilon_1),\varepsilon_1/\varepsilon_N)$. This is confirmed by the numerical investigations. We note that the solution for E in the case of the linear layer is less critical toward variation of the parameters than in the case of the nonlinear layer, since ρ_n is smaller in the former case than in the latter. A quantitative criterion for applicability of the long-wavelength approximation can be given from a comparison of the fields for the forms of $\varepsilon(kz)$ considered. The energy conservation law is used to monitor the accuracy of the calculations.

On the basis of the calculations, we shall use the condition $gr(\varepsilon)>20$ as a criterion for applicability of the long-wavelength approximation $gr(\varepsilon)\gg 1$ when $\varepsilon_1/\varepsilon_N\gg 1$.

10. When the influence of absorption is taken into account in the physical problem for a linear layer intersecting the point where $\varepsilon=0$, we introduce $\varepsilon''=\text{var}$ in the nonlinear portion $(\varepsilon_L,\varepsilon_N)$ in such a way that $\varepsilon'=\varepsilon-i\varepsilon''$ satisfies formula (A1) from the Appendix. Depending on the value of ε_L'', the function $\varepsilon(kz)$ either intersects the z axis when $\varepsilon_L''\sim\varepsilon_L$ or asymptotically approaches the z axis $(\varepsilon>0)$ when $\varepsilon_L''\ll\varepsilon_L$. We can give the following estimatefor the suitability of the model under consideration as applied to the latter case: $\varepsilon''\ll 0.1\alpha_1^{2/3}$.

The main results of the work are as follows. A method for solving the wave equation (1) in terms of elementary functions that employs finite differences with a variable step, which is matched to the variation of the wave impedance of the medium without absorption, has been proposed. Reduction of the number of independent parameters of the layer and application of the conservation law permitted simplification of the investigation of the fields at a jump in ε, which is justified for small losses. The similarity principle for inhomogeneous layers described allowed the results of investigations of the fields at a jump in ε to be extended to the vicinity of the zero of ε.

A critical point corresponding to $\varepsilon=0$, near which the solution depends qualitatively on small changes in the parameters, was discovered. For the transmission coefficient of the layer and the intensities of the electric and magnetic fields we have $S_N/S_1\Rightarrow 0$, $E\Rightarrow\infty$, and $H\Rightarrow 0$ as $\varepsilon\Rightarrow 0$.

The model considered in this work can easily be generalized to the case of a medium with losses, and the proposed method for solving the wave equation can be widely employed in mathematical physics.

APPENDIX A:

Let us consider the solution of the wave equation (1) in the nonlinear layer in region *3* (Fig. 1), where the function $\varepsilon(kz)$ is assigned by the condition $gr(\varepsilon)=gr(\varepsilon_L)=\text{const}$ and is represented for $z_L=0$ in the coordinate frame with a center 0_n by the expression

$$\varepsilon=\frac{\beta^2}{4(kz+b)^2}. \tag{A1}$$

Here

$$b=\frac{\beta}{-2\sqrt{\varepsilon_L}}=\frac{2\varepsilon_L}{-\alpha_1},\quad \beta=4/gr(\varepsilon_L),\quad |\beta|\ll 1,$$

and the wave equation (1) has an exact solution in elementary functions, which was first obtained by Rayleigh:[9]

$$E_3=C_1(kz+b)^{r_1}+C_2(kz+b)^{r_2}, \tag{A2}$$

where $r_{1,2}=0.5(1\pm\gamma)$ and $\gamma=\sqrt{1-\beta^2}$.

We note that $E_3\Rightarrow 0$ as $kz\Rightarrow -b$, where ε has a singularity (a pole), and the solution for the case of $\beta^2>1$ was considered in Ref. 3. The fields in regions *2* and *4*, respectively, have the form

$$E_4=A_N\exp(-i\sqrt{\varepsilon_N}kz),$$
$$E_2=\exp(-i\sqrt{\varepsilon_L}kz)+R_0\exp(i\sqrt{\varepsilon_L}kz). \tag{A3}$$

The following conditions should be satisfied on the boundary at $z=z_L$:

$$E_2=E_3,\quad \frac{dE_2}{d(kz)}=\frac{dE_3}{d(kz)}.$$

Using them and the analogous boundary conditions at $z=z_N$, as well as the expressions (A2) and (A3), we obtain the following expression for the reflectivity R_0:

$$R_0+i\beta=\gamma\frac{(c+1)}{(c-1)},$$

where

$$C=-\frac{(\gamma+i\beta)}{(\gamma-i\beta)}h,\quad h=\left(-\frac{\varepsilon_1}{\varepsilon_N}\right)^{\gamma/2}, \tag{A4}$$

or, after some transformations,

$$R_0=\frac{(h-1)[\gamma(h+1)-i\beta(h-1)]}{(h+1)^2-4h\beta^2},$$
$$|R_0|^2=\frac{(h-1)^2}{(h+1)^2-4h\beta^2}. \tag{A5}$$

When $h\gg 1$, we have $R_0=\gamma+i\beta$. Next, taking $r_1=1-\beta^2/4$ and $r_2=\beta^2/4$, we obtain

$$E_2=\left(\frac{\sqrt{\varepsilon_L}}{\sqrt{\varepsilon}}\right)^{\beta^2/4}\left[\frac{\sqrt{\varepsilon_N}}{\gamma\sqrt{\varepsilon}}(-3/2\beta^2+\beta^3+i\beta(1-3/2\beta^2))\right.$$
$$\left.+2-\beta^2/2-i\beta\right], \tag{A6}$$

and from the condition $E_3=E_4$ on the boundary at $z=z_N$ we have

$$E_2|_{z=z_N}=\left(\frac{\sqrt{\varepsilon_L}}{\sqrt{\varepsilon_N}}\right)^{\beta^2/4}[2-2\beta^2-i\beta(2-2\beta^2)],$$
$$A_N=\exp\left[i\frac{\beta}{2}\left(1-\frac{\sqrt{\varepsilon_N}}{\sqrt{\varepsilon_1}}\right)\right]E_2|_{z=z_N}. \tag{A7}$$

The mathematical solution of the problem is completed by substituting A_N into (A3).

The solution for the reflectivity R_0 from the nonlinear layer under consideration can be obtained directly from the Riccati equation, and in this case it is equivalent to (A4). If the medium is lossless, the conservation law can be used.

Then the transmission coefficient, which is the ratio between the mean values of the Poynting vector for the transmitted and incident waves, is given by the relation $S_N/S_1=1-|R_0|^2$, whence for the fields in region *4* we obtain

$$|E_4|^2=\frac{\sqrt{\varepsilon_L}}{\sqrt{\varepsilon_N}}(1-|R_0|^2)=4\left(\frac{\varepsilon_L}{\varepsilon_N}\right)^{\beta^2/4}(1-\beta^2), \tag{A8}$$

$$|H_4|^2=\frac{\sqrt{\varepsilon_N}}{\sqrt{\varepsilon_L}}(1-|R_0|^2)=4\left(\frac{\varepsilon_N}{\varepsilon_L}\right)^{1-\beta^2/4}(1-\beta^2), \tag{A9}$$

where

$$\frac{\varepsilon_L}{\varepsilon_N}=\frac{|\alpha_1|kz}{2\varepsilon_L}+1.$$

It follows from (A5), (A8), and (A9) that $|R_0|\Rightarrow 1$, $E_4\Rightarrow\infty$, and $H_4\Rightarrow 0$ as $\varepsilon\Rightarrow 0$. This is a first approximation for a linear layer characterized by $gr(\varepsilon_1)=4/\beta$. We note that E depends very significantly on $\varepsilon(kz)$ when $\beta\ll 1$, unlike H and R_0 from (A5). Although the exact solutions (A8) and (A9) describe the field only after passage through the layer, unlike (A6), they are attractive because of the simplicity of their form. An analysis of (A6) shows that $dE_2/dkz=0$ when $\varepsilon\cong 10\varepsilon_N$ and then increases.

[1] L. A. Zhekulin, Zh. Éksp. Teor. Fiz. **4**(1), 76 (1934).
[2] V. A. Permyakov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. No. 8, 1264 (1969).
[3] V. L. Ginzburg, *The Propagation of Electromagnetic Waves in Plasmas*, 2nd ed. [Pergamon Press, Oxford–New York (1970); Nauka, Moscow (1967), 683 pp.].
[4] I. P. Kozlov, *Propagation and Diffraction of Electromagnetic Waves* [in Russian], MFTI, Moscow (1993), pp. 104–113.
[5] I. P. Kozlov, Radiotekh. Elektron. No. 2, 142 (1997).
[6] I. P. Kozlov, Élektrodin. Tekh. SVCh KVCh (Moscow) **4**(4), 63 (1996).
[7] V. A. Fock, *Electromagnetic Diffraction and Propagation Problems*, 1st ed. [Pergamon Press, Oxford–New York (1965); Sov. Radio, Moscow (1970), 560 pp.].
[8] I. P. Kozlov, *Forty-Fifth Congress of the International Astronomy Federation* (IAF-94-U.2.469), Jerusalem (1994), 4 pp.
[9] J. W. Strutt (Baron Rayleigh), *The Theory of Sound*, 2nd ed. [Dover, New York (1945); Gostekhizdat, Moscow (1955)].

Translated by P. Shelnitz

Exact solutions of the forced Burgers equation

S. V. Petrovskiĭ

P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences, 117851 Moscow, Russia
(Submitted November 28, 1997)
Zh. Tekh. Fiz. **69**, 10–14 (August 1999)

Two new methods for obtaining exact solutions of the initial-value problem on an unbounded straight line (the Cauchy problem) for the inhomogeneous Burgers equation are considered. They are applied to the cases of a stationary and a transient external force. A self-similar solution and a solution which describes the localization (blocking) of solitary traveling waves are obtained as examples. © *1999 American Institute of Physics.* [S1063-7842(99)00308-6]

INTRODUCTION

The Burgers equation[1)]

$$v_t - 2vv_x - v_{xx} = F \tag{1}$$

(where $v = v(x,t)$, x is the coordinate, and t is the time), which was originally proposed for describing turbulence,[1] has turned out to be an effective model of the dynamics of nonlinear dissipative media of diverse physical nature.[2–4] In particular, it was shown in Ref. 2 that Eq. (1) holds for an extensive class of processes in hydrodynamics, nonlinear acoustics, and plasma physics. The Burgers equation with a nontrivial right-hand side, i.e., with an external force $F(x,t) \not\equiv 0$ (the forced Burgers equation), describes the dynamics of a physical system immersed in an external field (a system with energy ''pumping'') and is a natural generalization of the homogeneous equation corresponding to autonomous motions.

Although the literature devoted to Eq. (1) is indeed enormous, it is still the subject of many studies (see, for example, Refs. 5–8). Attention has recently focused specifically on the forced Burgers equation.[6–8] Exact solutions of the Cauchy problem for the forced Burgers equation in the case where the coordinate dependence on the right-hand side of Eq. (1) is singular, i.e., is described either by a Dirac δ function or by its derivative, were obtained in Refs. 6 and 7. This paper proposes a method which provides special exact solutions for the case where $F(x,t)$ is a continuous function.

As we know, the Hopf–Cole substitution[9,10]

$$v = u_x/u \tag{2}$$

transforms the forced Burgers equation into a linear equation for the new function $u(x,t)$

$$u_t - u_{xx} = u \int_{x_0}^{x} F(x',t)\, dx'. \tag{3}$$

If the right-hand side of Eq. (1) $F(x,t) \equiv 0$ (just this case was considered in Refs. 1, 9 and 10), Eq. (3) reduces to an ordinary diffusion equation, and its solution for arbitrary initial conditions is written in quadratures, allowing us to also write a solution for the Burgers equation with allowance for (2). However, if $F \not\equiv 0$, the construction of an analytical solution of equation (3) is a very complicated problem. A general algorithm for constructing a solution in the form of a series with insignificant restrictions on the form of $F(x,t)$ was proposed in Ref. 11, but the expressions appearing as a result are very cumbersome, making it difficult to use them in practice. Nevertheless, because of the physical significance of Eq. (1) there is unquestionable interest in obtaining exact solutions for it which are expressed by relatively simple formulas and have a clear physical meaning.[2] This is the goal of the present work.

STATIONARY EXTERNAL FORCE: BLOCKING OF A SOLITARY WAVE

We begin with the case in which the right-hand side of Eq. (1) does not depend on t, i.e., $F = F(x)$. Instead of the classical substitution (2) we consider the modified form

$$v = (u_x/u) + k(x), \tag{4}$$

where $k(x)$ is a certain function.

Substituting (4) into (1), after some transformations we have

$$((u_t - 2ku_x - u_{xx})/u)_x = (k'(x) + k^2 + F(x) - C(t))_x, \tag{5}$$

where $C(t)$ is an arbitrary function of time (the minus sign was chosen for convenience).

Integrating over x, from Eq. (5) we obtain

$$u_t = 2ku_x - u_{xx} = (k'(x) + k^2 + \psi(x) - C(t))u, \tag{6}$$

where $\psi'(x) = F(x)$.

Thus, in cases where the function $k(x)$ is a solution of the Riccati equation

$$k'(x) + k^2 = -\psi(x) + C, \tag{7}$$

the function $u(x,t)$ satisfies a linear equation of the diffusion–convection type:

$$u_t - 2ku_x = u_{xx}. \tag{8}$$

In Eq. (7) the time t appears only in the form of a parameter, i.e., as the argument of the arbitrary function C. We note that if we set $C(t) = \text{const}$ (just this case will be considered below), the solutions of Eq. (7) are stationary solutions

of the Burgers equation. The meaning of the substitution (4) is thereby made clear: Eq. (8) for the new unknown $u(x,t)$ describes a process evolving on top of the ''background'' of the stationary profile $k(x)$.

The solution of the Riccati equation (7) with a right-hand side of arbitrary form is a complicated problem, which does not always admit writing a solution in a closed form. However, at least in some cases the system (7) and (8) turns out to be more convenient for obtaining exact solutions of the forced Burgers equation than the approach based on Eqs. (2) and (3). In particular, unlike Eq. (3), Eq. (8) has a family of solutions in the form of a traveling wave [for $k(x)$ of a certain form].

In fact, we seek a solution in the form $u(x,t)=U(\xi)$, where $\xi=x-y(t)$.

Making the substitution in (8), we obtain the equation for $U(\xi)$

$$-(y'(t)+2k(x))U'(\xi)=U''(\xi). \tag{9}$$

The transition to the coordinates of a traveling wave is correct if the variables x and t enter into in Eq. (9) only in terms of the variable ξ, i.e., if $y'(t)+2k(x)=\phi(\xi)$, where ϕ is a certain function. Clearly, since the variables x and ξ are related by a linear transformation, this is possible only if k and ϕ are linear functions of their arguments, i.e., if $k(x)=Bx+B_1$ and $\phi(\xi)=\beta\xi+\gamma$, where B, B_1, β, and γ are constants. Then, Eq. (7) yields

$$B+(Bx+B_1)^2=-\psi(x)+C, \tag{10}$$

whence, differentiating with respect to x, we obtain $F(x)=-2B(Bx+B_1)$.

Thus, the inhomogeneous Burgers equation has a traveling-wave solution only if the external force is a linear function of the coordinate,[3)] i.e., if

$$F(x)=-2B^2x \tag{11}$$

(if, with no loss of generality, we set $B_1=0$, which can always be accomplished by selecting the origin of coordinates). Equation (7) takes the form

$$k'(x)+k^2=B^2x^2+C. \tag{12}$$

It is not difficult to prove [for example, by direct substitution into (12)] that Eq. (12) has the required linear solution $k(x)=Bx$ only for the choice of integration constant $C=B$.

We note that for the Riccati equation knowledge of one special solution permits finding its general solution (see, for example, Ref. 12). The general solution of Eq. (12) (for $C=B$) has the form

$$k(x)=Bx+\exp(-Bx^2)/(C_1+(\pi/4B)^{1/2}\mathrm{erf}(B^{1/2}x)), \tag{13}$$

where $\mathrm{erf}(z)$ is the error function, and C_1 is an integration constant or, more specifically, $C_1\subseteq\mathbb{R}U\{\infty\}$.

Any function of the one-parameter family (13) is a stationary solution of the Burgers equation with a right-hand side in the form (11).

From Eq. (9) and the condition of linearity of the functions $k(x)$ and $\phi(x)$ we obtain

$$y'(t)+2Bx=\beta\xi+\gamma=\beta(x-y(t))+\gamma, \tag{14}$$

where β and γ are constant coefficients.

Separating the variables, we have

$$(2B-\beta)x=0, \tag{15}$$

whence $\beta=2B$ and

$$y'(t)+\beta y-\gamma=0. \tag{16}$$

Setting $y(0)=0$ (which corresponds to the natural choice of the initial condition $\xi(x,0)=x$), we obtain

$$y(t)=\delta[1-\exp(-2Bt)], \tag{17}$$

where $\delta=\gamma/2B$.

Let us find the expression which describes the form of the wave. From (13)–(15) we have

$$-(2B\xi+\gamma)U'(\xi)=U''(\xi). \tag{18}$$

Introducing the new variable $U'(\xi)=p(\xi)$, we arrive at the first-order equation

$$p'(\xi)=-(2B\xi+\gamma)p, \tag{19}$$

which can easily be integrated to obtain

$$p\equiv U'(\xi)=\mathrm{const}\exp(-B\xi^2-\gamma\xi). \tag{20}$$

The case where $B<0$ is apparently not of interest, since the solution [which describes the deviation from the stationary profile $k(x)=Bx$] increases without bound as $|x|\to\infty$. Therefore, we henceforth set $B>0$. The solution of Eq. (20) is found without difficulty:

$$U(\xi)=a+b(1+\mathrm{erf}[B^{1/2}(\xi+\delta)]), \tag{21}$$

where a and b are constants.

Returning to the original variables x, t, and u, we have

$$u(x,t)=a+b(1+\mathrm{erf}[B^{1/2}(x+\delta\exp(-2Bt))]). \tag{22}$$

Finally, introducing the notation $\mu=(\pi/4B)^{1/2}(a+b)/b$ and taking into account (4), we obtain the following family of solutions of Eqs. (1) and (11):

$$v(x,t)=Bx+\exp[-B(x+\delta\exp(-2Bt))^2]/(\mu+(\pi/4B)^{1/2}\mathrm{erf}[B^{1/2}(x+\delta\exp(-2Bt))]). \tag{23}$$

Obviously, when $|\mu|>(\pi/4B)^{1/2}$, the function (23) is continuous for any x and t [when $|\mu|<(\pi/4B)^{1/2}$, expression (23) has discontinuities of the second kind; solutions of this type, which have also attracted considerable attention,[13,14] will not be considered in this paper]. It is not difficult to see that the two-parameter family of solutions $v(x,t;\delta,\mu)$ describes the retarded motion of a dome-shaped asymmetric wave along the linear profile $k(x)=Bx$. The value of δ characterizes the initial position of the wave, and $1/\mu$ characterizes its amplitude; the position of the wave relative to the origin of coordinates and the straight line $k(x)=Bx$ also depends on the signs of δ and μ. The final position (which is established asymptotically as $t\to\infty$) is described by (13) (when $C_1=\mu$).

We note that because of the linearity of Eq. (8) a linear combination of solutions of the form (23) is also a solution.

Thus, the more general N-wave solution of the forced Burgers equation with the external force (11) has the form

$$v(x,t)=Bx+\left(\sum_{i=0}^{N}\varepsilon_i\exp[-B(x+\delta_i\exp(-2Bt))^2]\right)\Bigg/\left(\mu+(\pi/4B)^{1/2}\sum_{i=0}^{N}\varepsilon_i\times\operatorname{erf}[B^{1/2}(x+\delta_i\exp(-2Bt))]\right). \quad (24)$$

If the constants δ_i differ strongly enough in absolute value, the solution (24) is a set of N individual waves, which ''gather'' from the right and left toward the origin of coordinates. In this case, if the constants δ_i are of the same order of magnitude, the individual waves merge, and the solution (24) describes the evolution of the initial disturbance (which, with an appropriate set of values for the parameters μ, ε_i, δ_i, and N, can now have a fairly complicated form) to the stationary distribution (13).

TRANSIENT EXTERNAL FORCE: AUTOTRANSFORMATIONS

In the case considered above, the right-hand side of the Burgers equation (1) depended only on the variable x. In the more general case $F=F(x,t)$. It is not difficult to directly verify that in this case the transformation (4), where we now have $k=k(x,t)$, also leads to a linear equation of the form (8) for the new function $u(x,t)$. The coupling equation, however, is no longer the Riccati equation, but coincides with the original equation (1). Thus, in the case of a transient external force the substitution (4) describes ''autotransformation'' of the solutions of the forced Burgers equation (1), and if the function $k(x,t)$ is a solution of the Burgers equation (with a right-hand side of arbitrary form), then the function

$$v(x,t)=(u_x/u)+k(x,t) \quad (25)$$

is a solution (which corresponds to a different initial condition) of the same equation (1), and $u(x,t)$ satisfies Eq. (8).

The relation (25) can be used to construct exact solutions in the case where the right-hand side of the Burgers equation depends on time. As an example let us consider the model equation describing the dynamics of a certain system in a linear field decaying with time

$$v_t-2vv_x-v_{xx}=-ax/(t+t_0)^2, \quad (26)$$

where a and t_0 are constants and in order to avoid singularities at $t>0$ we set $t_0>0$.

It is not difficult to see that the function $k(x,t)=bx/(t+t_0)$ is a solution of Eq. (26) under the condition $b+2b^2=a$. The solutions of Eq. (26) have different properties, depending on the signs of a and b. Since the purpose of this section is to present an example of how to obtain exact solutions of Eq. (1) using the autotransformation (25), rather than a detailed investigation of Eq. (26), we confine ourselves here to the single case of $a>0$ and $b=(1/4)((1+8a)^{1/2}-1)>0$.

The solution $k(x,t)$ is not of interest because of its simplicity. However, it can be used to construct another solution, which has very interesting properties. By virtue of (25) the expression $v=k+(u_x/u)$ will also be a solution, if $u(x,t)$ satisfies the following equation:

$$u_t-(2bx/(t+t_0))u_x=u_{xx}. \quad (27)$$

We seek the solution of Eq. (27) in the self-similar form $u(x,t)=U(\xi)$, where $\xi=x\phi(t)$, and the form of the functions U and ϕ is subject to determination. Making the substitution in (27), we have

$$(x\phi^{-2}\phi'(t)-2b\phi^{-1}x/(t+t_0))U'(\xi)=U''(\xi). \quad (28)$$

The transition to self-similar variables is correct if the expression in brackets is a function of ξ. To satisfy this condition we require that

$$\phi^{-2}\phi'(t)=\lambda\phi, \quad 1/[(t+t_0)\phi]=\nu^{-2}\phi, \quad (29)$$

where λ and η are coefficients.

From (29) we find that $\phi(t)=\eta(t+t_0)^{-1/2}$ and $\lambda=-0.5\eta^{-2}$. Setting $\xi(x,0)=x$, we have $\eta=t_0^{1/2}$.

Equation (28) takes the form

$$-(2/t_0)(b+0.25)\xi U'(\xi)=U''(\xi). \quad (30)$$

Introducing the notation $(b+0.25)/t_0=\alpha^2$ (since $b>0$; see above), from (30) we obtain

$$-2\alpha^2\xi U'(\xi)=U''(\xi). \quad (31)$$

Equation (31) is easily integrated to obtain

$$U(\xi)=A_1\operatorname{erf}(\alpha\xi)+A_2, \quad (32)$$

where $\operatorname{erf}(z)$ is the error integral, and A_1 and A_2 are constants.

Taking into account the relation (25), we arrive at the following exact solution of Eq. (26);

$$v(x,t)=bx/(t+t_0)+[t_0/(t+t_0)]^{1/2}\times\exp(-\alpha^2\xi^2)/(\varkappa+(\sqrt{\pi}/2\alpha)\operatorname{erf}(\alpha\xi)), \quad (33)$$

where $\xi=x[t_0/(t+t_0)]^{1/2}$.

When $|\varkappa|>\sqrt{\pi}/2\alpha$, the function (33) is continuous at all x and t. Expression (33) describes the self-similar diffusion and fading of a dome-shaped initial disturbance of a linear field. Thus, the simple ''seed'' solution $k(x,t)$, which is not of great interest in itself, can be used to generate more complex solutions through the transformation (25).

We note, in conclusion, that an attempt to construct a ''many-particle'' solution similar to the multiwave solution (24) of the preceding section for Eq. (26) was unsuccessful: the expression obtained coincides exactly with (33).

CONCLUSION

The new method proposed in this paper for constructing exact solutions of the forced Burgers equation on the basis of the modified Hopf–Cole transformation (2) was considered separately for the cases of stationary and transient external forces. However, there is essentially a single method for both cases: the substitution (2) reduces the forced Burgers equation (1) to the linear equation (8), which describes the evo-

lution of an initial disturbance of some ''original'' solution, which can be either stationary or transient, depending on the type of external force.

The special exact solutions (23), (24), and (33) are unbounded as $|x|\to\infty$, raising some doubt as to their physical significance at first glance. We note in this regard that the examination of solutions which are unbounded at infinity has a long-standing tradition in mathematical physics (the textbook example is the field of an infinitely long charged wire). The fact is that in a real physical system ''infinity'' signifies satisfaction of the condition $L/l\gg 1$, where L is the external scale of the system and l is the characteristic length of the process being described (in the example with a wire this means that a solution which increases logarithmically to infinity correctly describes the distribution of the field near the axis of a real wire and is known to be incorrect at distances from the axis greater than or of the order of the length of the wire).

The properties of the solutions (23) and (24) completely correspond to the situation described. The deviation described by them from the stationary profile tends to zero as $|x|$ increases. This means, in particular, that the internal structure of the solutions determines the characteristic length l, which depends only on the parameters of the problem, so that the evolution of the initial disturbance takes place in the region from $-l$ to l. If L is a scale of the physical system which indicates the distances at which Eq. (1) is valid [particularly the distance at which the field $F(x)$ can be considered linear], when the condition $L\gg l$ is satisfied, it can be claimed that the dynamics of the system (for the corresponding initial conditions) are described by Eq. (23) or (24).

The foregoing statements also apply to the self-similar solution (33) except that now the length l also depends on the duration of the treatment of the process (as a consequence of the spreading of the initial disturbance): $l\sim T^{1/2}$. This additionally restricts the applicability of (33) to not excessively long times.

We note, in addition, that all the solutions obtained are suitable for describing processes in a system of finite size in the case where conditions of special form, which can be specified by the form of (23), (24), or (33), are satisfied on its boundaries.

The last remark refers to autotransformations. Equation (25) actually specifies only one ''link'' in an entire hierarchy of solutions. More specifically, after a certain ''seed'' solution k_0, we arrive at $k_1=k_0+(u_x^{(0)}/u^{(0)})$, then $k_2=k_1+(u_x^{(1)}/u^{(1)})$, $k_3=k_2+\ldots$, etc. The study of this hierarchy is a fascinating problem and will be the subject of future investigations.

[1)]The left-hand side of the Burgers equation is usually written in the form $v_t+vv_x-v_{xx}$, and the coefficient -2 in Eq. (1) was chosen for convenience.

[2)]For further details on the importance of the exact solutions, see Ref. 5.

[3)]This applies specifically to the inhomogeneous equation (1). The existence of solutions in the form of a traveling wave for the homogeneous Burgers equation is a well-known fact.[3]

[1] J. M. Burgers, Adv. Appl. Mech. **1**, 171 (1948).
[2] C. S. Su and C. S. Gardner, J. Math. Phys. **10**, 536 (1969).
[3] G. B. Whitham, *Linear and Nonlinear Waves* [Wiley, New York (1974); Mir, Moscow (1977), 621 pp.].
[4] M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *Wave Theory* [in Russian], Nauka, Moscow (1979), 384 pp.
[5] S. Hood, J. Math. Phys. **36**, 1971 (1995).
[6] M. J. Ablowitz and S. De Lillo, Phys. Lett. A **156**, 483 (1991).
[7] M. J. Ablowitz and S. De Lillo, Physica D **92**, 245 (1996).
[8] A. Chekhlov and V. Yakhot, Phys. Rev. E **52**, 5681 (1995).
[9] E. Hopf, Commun. Pure Appl. Math. **3**, 201 (1950).
[10] J. D. Cole, Q. Appl. Math. **9**, 225 (1951).
[11] V. A. Il'in, A. S. Kalashnikov, and O. A. Oleĭnik, Usp. Mat. Nauk **17**(3), 3 (1962).
[12] E. Kamke, *Differentialgleichungen, Lösungsmethoden und Lösungen. I. Gewöhnliche Differentialgleichungen*, 5th ed. [Akademische Verlagsgesellschaft, Leipzig (1959); Nauka, Moscow (1965), 704 pp.].
[13] A. M. Samsonov, Appl. Anal. **57**, 85 (1995).
[14] K. A. Volosov, V. G. Danilov, and A. M. Loginov, Teor. Mat. Fiz. **101**, 189 (1994).

Translated by P. Shelnitz

Excited hydrogenic atom in a strong low-frequency electromagnetic field

O. B. Prepelitsa

Institute of Applied Physics, Moldovan Academy of Sciences, 277028 Kishinev, Moldova
(Submitted May 22, 1998)
Zh. Tekh. Fiz. **69**, 15–20 (August 1999)

The multiphoton ionization of a bound electron state which is twofold degenerate with respect to its orbital angular momentum is considered in a quasiclassical approximation. It is shown that the ionization probability increases strongly in an intense electromagnetic field, in which nonresonant mixing of the levels forming the degenerate state is significant, in comparison to the case described by the Keldysh formula. It is also shown that such degeneracy leads to a sharp increase in the intensity of the radiation scattered by the bound electron, and the high-frequency cutoff of the emission spectrum is shifted to higher frequencies. © *1999 American Institute of Physics.* [S1063-7842(99)00408-0]

INTRODUCTION

This paper is devoted to an investigation of the physical consequences of the accidental degeneracy of bound levels during interactions with an electromagnetic field. This motivates the choice specifically of an excited hydrogenic atom, which is known to have l-fold degeneracy (l is the orbital angular momentum quantum number). Two important effects arising when a strong electromagnetic wave acts on a bound electronic system are considered: multiphoton ionization and higher-harmonic generation. The multiphoton ionization of excited atoms was considered theoretically and experimentally in Refs. 1–6. However, it was assumed in the theoretical description of this phenomenon that the state being ionized does not have an ''internal'' structure, in the sense that the mixing of closely positioned energy levels by the applied electromagnetic field was not taken into account. This is justified for a hydrogenic atom only in fairly weak electromagnetic fields. In fact, fundamental reorganization of the quasienergy spectrum takes place in strong fields owing to such mixing. This, in turn, leads to qualitative differences in the characteristics of the electrodynamic processes in comparison to the ordinary case. Here we should mention Refs. 7 and 8, in which it was shown that the intrinsic dipole moment of the excited state of a two-level system has a significant influence on the rate of its excitation, as well as on the intensity and line shape of the emission spectrum.

This paper consists of an Introduction followed by two other sections. The first of these sections describes the derivation of an analog of the Keldysh formula[9] for the probability of the multiphoton ionization of a twofold degenerate electronic state. It is shown that owing to the accidental degeneracy of the level, the ionization rate increases exponentially in comparison with the ionization rate of a nondegenerate state.

The last section explores the spectrum of higher-order harmonics emitted by an excited hydrogenic atom. The generation of high-order harmonics by atoms has been investigated in many studies (see, for example, Refs. 10–15). According to the experimental results, the spectrum of harmonics contains a broad plateau with a sharp high-frequency cutoff, which is located at photon energies close to $I+2-3U_p$, where I is the ionization potential and U_p is the ponderomotive energy of the electron. This phenomenon is of interest in connection with the possibility of obtaining coherent high-frequency radiation (up to the x-ray range). Unfortunately, because of the rigid restrictions on the electromagnetic field intensity and the small ionization potential, an excited atom is a less promising source than a ground-state atom. Nevertheless, all other conditions being equal, the intensity of the radiation scattered by an excited hydrogenic atom is much greater than in the ordinary case, i.e., in the absence of degeneracy. In addition, the strong mixing of the levels forming a degenerate state leads to an increase in the length of the plateau in the emission spectrum.

1. MULTIPHOTON IONIZATION OF AN EXCITED HYDROGENIC ATOM

Let us consider a bound electron state which is twofold degenerate with respect to the orbital angular momentum in the field of a linearly polarized wave of frequency ω and intensity $\mathbf{E}(t)=\mathbf{E}_0\cos\omega t$. In the Keldysh–Faisal–Reiss approximation the ionization probability amplitude is found using the formula

$$A_{if}=-\frac{i}{\hbar}\int dt\,d\mathbf{r}\,\Psi_f^*(\mathbf{r},t)e\mathbf{E}(t)\cdot\mathbf{r}\Psi_i(\mathbf{r},t), \tag{1}$$

$$\Psi_f(\mathbf{r},t)=\frac{1}{(2\pi\hbar)^{3/2}}\exp\left[\frac{i}{\hbar}\left(\left(\hbar\mathbf{k}+\frac{e\mathbf{E}_0}{\omega}\sin\omega t\right)\cdot\mathbf{r}-\int_0^t d\tau\frac{\left(\hbar\mathbf{k}+\frac{e\mathbf{E}_0}{\omega}\sin\omega\tau\right)^2}{2M}\right)\right],$$

$$\Psi_i(\mathbf{r},t)=\Psi_1(\mathbf{r},t)\exp(-i\rho\sin\omega t)+\Psi_2(\mathbf{r},t)\exp(i\rho\sin\omega t),$$

$$\Psi_{1,2}(\mathbf{r},t)=\frac{1}{\sqrt{2}}(\Psi_s(\mathbf{r})\pm\Psi_p(\mathbf{r}))\exp\left(\frac{i}{\hbar}It\right). \tag{2}$$

Here $\Psi_f(\mathbf{r},t)$ is the wave function of a free electron with momentum $\hbar\mathbf{k}$ in the field of a plane electromagnetic wave; $\Psi_i(\mathbf{r},t)$ is the wave function of the state being ionized in the wave field;[16] $\Psi_s(\mathbf{r})$ and $\Psi_p(\mathbf{r})$ are the wave functions of the degenerate state in the absence of an electromagnetic field; I is the ionization potential;

$$\rho=\frac{|\mathbf{E}_0\cdot\mathbf{d}|}{\hbar\omega},\quad \mathbf{d}=e\int d\mathbf{r}\,\Psi_s^*(\mathbf{r})\mathbf{r}\Psi_p(\mathbf{r}).$$

We present some explanations regarding the choice of the wave function of the initial state in the form (2). It is generally known that even a comparatively weak electromagnetic field leads to intense mixing of states with the same energy between which transitions are allowed by the selection rules. In our case the parameter ρ is a measure of that mixing. Under conditions where the inequality $\rho\gtrsim 1$ holds, the transitions between levels forming a degenerate state should be taken into account exactly, i.e., in all orders of perturbation theory. Only if $\rho\ll 1$ can the treatment be confined to a finite number of perturbation terms. It can easily be seen that even for fields with an intensity significantly below the atomic level there is no bound on the value of ρ in the general case. This stipulates the choice of the wave function of the initial state in the form (2), where mixing of the $\Psi_s(\mathbf{r})$ and $\Psi_p(\mathbf{r})$ levels by an external electromagnetic field was taken into account exactly in Ref. 16.

Utilizing the formula for expanding an exponential function in Bessel functions of real argument

$$\exp(i\rho\sin\omega t)=\sum_{m=-\infty}^{\infty}J_m(\rho)\exp(im\omega t),$$

we rewrite the expression (2) in the form

$$\Psi_i(\mathbf{r},t)=\sum_{m=-\infty}^{\infty}\Psi_i^{(m)}(\mathbf{r})J_m(\rho)\exp\left[\frac{i}{\hbar}(I-m\hbar\omega)t\right],$$

$$\Psi_i^{(m)}(\mathbf{r})=\Psi_1(\mathbf{r})+(-1)^m\Psi_2(\mathbf{r}).$$

As follows from the last formulas, a bound quasistationary electron state in the field of an electromagnetic wave can be interpreted as a superposition of states with the quasienergies

$$\varepsilon_m=-(I+m\hbar\omega);\quad m=0,\pm1,\pm2,\ldots. \tag{3}$$

Taking into account the properties of Bessel functions, we can easily see that the maximum of the quasienergy distribution corresponds to the values $m\sim\pm[\rho]$, where [ ] denotes the integer part of the number. This means that only the half-levels with $\varepsilon_{\pm[\rho]}=-(I\pm[\rho]\hbar\omega)$ are substantially populated. Thus the electron is partially localized in a state lying closer to the continuum edge than in the case of $\rho=0$. Here the width of the barrier separating the quasistationary (discrete) states and the states of the continuous spectrum is smaller than for $\rho=0$. As will be shown below, this leads to an effective increase in the probability of the passage of the electron through the potential barrier and ionization of the atom. We also note that since the electron is found mainly in states with the quasienergies $I\pm[\rho]\hbar\omega$, the usual condition for applicability of the quasiclassical approximation[9] $I/(\hbar\omega)\gg 1$ is replaced by the more rigid condition $I/(\hbar\omega)-\rho\gg 1$. This means that the condition for a multiphoton transition should be satisfied for the quasilevel which is closest to the continuum and is most strongly populated.

After calculating the integral over the coordinate and some transformations associated with expansion of the functions in Fourier series, we represent the expression (1) in the form

$$\begin{aligned}A_{if}=-\frac{ie}{\hbar}\int_{-\pi}^{\pi}d\varphi\Bigg(&\mathbf{E}_0\cdot\mathbf{D}_1(\mathbf{k}(\varphi))\exp\Bigg[\frac{i}{\hbar\omega}\Bigg(\int_0^{\varphi}d\varphi'\\&\times\frac{\left(\hbar\mathbf{k}+\dfrac{e\mathbf{E}_0}{\omega}\sin\varphi'\right)^2}{2M}+I\varphi-\rho\hbar\omega\sin\varphi\Bigg)\Bigg]\\&+\mathbf{E}_0\cdot\mathbf{D}_2(\mathbf{k}(\varphi))\exp\Bigg[\frac{i}{\hbar\omega}\Bigg(\int_0^{\varphi}d\varphi'\\&\times\frac{\left(\hbar\mathbf{k}+\dfrac{e\mathbf{E}_0}{\omega}\sin\varphi'\right)^2}{2M}+I\varphi+\rho\hbar\omega\sin\varphi\Bigg)\Bigg]\Bigg)\\&\times\sum_{l=-\infty}^{\infty}\delta\left(\frac{\hbar^2\mathbf{k}^2}{2M}+I+(\beta-l)\hbar\omega\right),\end{aligned} \tag{4}$$

where

$$\mathbf{D}_{1,2}(\mathbf{k}(\varphi))=\frac{1}{(2\pi\hbar)^{3/2}}\int d\mathbf{r}\exp(i\mathbf{k}(\varphi)\mathbf{r})\mathbf{r}\Psi_{1,2}(\mathbf{r}),$$

$$\mathbf{k}(\varphi)=\mathbf{k}+\frac{e\mathbf{E}_0}{\hbar\omega}\sin\varphi,\quad \beta=\frac{e^2\mathbf{E}_0^2}{4M\hbar\omega^3}.$$

We assume that the inequality $I/(\hbar\omega)-\rho\gg 1$ is always satisfied. In this case the exponential functions in (4) are rapidly oscillating functions, and the saddle-point (stationary phase) method can be used to calculate the integral. The saddle points are found from the relation

$$\frac{(\hbar\mathbf{k}+(e\mathbf{E}_0/\omega)\sin\varphi^*)^2}{2M}+I\pm\rho\hbar\omega\cos\varphi^*=0. \tag{5}$$

As in the derivation of the Keldysh formula,[9] it can be shown in our case that the main contribution to the ionization probability amplitude is made by small $\mathbf{k}$, which are such that $\hbar^2\mathbf{k}^2/(2M)\ll I$. Therefore, confining ourselves to exponential accuracy in the calculation, we can set $\mathbf{k}=0$ in (4) and (5) and thereafter obtain the following values for the saddle points:

$$\varphi^* = i\sinh^{-1}\gamma + \begin{cases} \pi, \\ 0, \end{cases}$$

$$\gamma^2 = \gamma_0^2 + \gamma_\rho^2 - \sqrt{2\gamma_\rho^2(\gamma_0^2+1)+\gamma_\rho^4},$$

$$\gamma_0^2 = \frac{I}{2\beta\hbar\omega}, \quad \gamma_\rho^2 = \frac{\rho^2}{8\beta^2}.$$

Now there is no difficulty in finding the ionization rate. After some ordinary transformation we obtain

$$w \sim \exp\left[-\frac{2I}{\hbar\omega} f(\gamma) \right], \tag{6}$$

$$f(\gamma) = \left(1 + \frac{1}{2\gamma_0^2} \right) \sinh^{-1}\gamma - \frac{\gamma(1+2\gamma_0^2-\gamma^2)}{2\gamma_0^2\sqrt{1+\gamma^2}}. \tag{7}$$

In the case of $\rho = 0$ ($\gamma = \gamma_0$) formula (6) transforms into the expression for the multiphoton ionization rate of a nondegenerate level obtained by Keldysh[9]

$$w_0 \sim \exp\left[-\frac{2I}{\hbar\omega} f_0(\gamma_0) \right],$$

$$f_0(\gamma_0) = \left(1 + \frac{1}{2\gamma_0^2} \right) \sinh^{-1}\gamma_0 - \frac{\sqrt{1+\gamma_0^2}}{2\gamma_0}. \tag{8}$$

Since the inequality $f(\gamma) < f_0(\gamma_0)$ always holds for $\rho \neq 0$, it is clear that, all other conditions being equal, degeneracy of the state being ionized with respect to the orbital angular momentum leads to an exponential increase in the ionization probability in comparison to the case of $\mathbf{d} = 0$. For example, for the parameter values $|\mathbf{E}_0| \sim 10^7$ V/cm and $\omega \sim 10^{14}\,\mathrm{s}^{-1}$, as well as $I \sim 3.45$ eV and $|\mathbf{d}| \sim 10$ D, which correspond to the ionization potential and intrinsic dipole moment of the $n^* = 2$ state of the hydrogen atom, we have $\ln(w/w_0) \sim 11$. In the limiting cases in which ionization has a multiphoton or tunneling character, formula (6) undergoes considerable simplification and takes the form

$$w \sim (2\gamma)^{-\frac{2I}{\hbar\omega}} \exp\left[\frac{2I}{\hbar\omega} \frac{2\gamma_0^2 - \gamma^2}{2\gamma_0^2} \right], \quad \gamma \gg 1, \tag{9}$$

$$w \sim \exp\left[-\frac{4}{3} \frac{I-\rho\hbar\omega}{\hbar\omega} \sqrt{\frac{I-\rho\hbar\omega}{2\beta\hbar\omega}} \right], \quad \gamma_0 \ll 1. \tag{10}$$

It follows from (9) that in the multiphoton regime the ionization probability, as in the case of $\rho = 0$, is proportional to $\mathbf{E}_0^{2I/(\hbar\omega)}$. The numerical value of the ionization rate is then $(\gamma_0/\gamma)^{2I/(\hbar\omega)} \exp(I(\gamma_0^2-\gamma^2)/(\gamma_0^2\hbar\omega))$ times greater than in the ordinary case. Formula (10) formally coincides with the corresponding limiting case of the Keldysh formula (8)

$$w_0 \sim \exp\left[-\frac{4}{3} \frac{I}{\hbar\omega} \sqrt{\frac{I}{2\beta\hbar\omega}} \right], \quad \gamma_0 \ll 1,$$

where the difference $I - \rho\hbar\omega$ appears instead of the ionization potential I.

Thus, in the tunneling regime ionization takes place mainly from the quasienergy level $\varepsilon_{[\rho]} = -(I - [\rho]\hbar\omega)$, confirming the qualitative arguments regarding the nature of the ionization of a degenerate state made at the beginning of this section.

2. HIGHER-HARMONIC GENERATION BY AN EXCITED HYDROGENIC ATOM IN THE FIELD OF A STRONG ELECTROMAGNETIC WAVE

Considering a pulse of an electromagnetic field short enough that significant depletion of the bound state due to ionization does not occur during its action, we represent the wave function of the electron in the form

$$\Psi(\mathbf{r},t) = \Psi_i(\mathbf{r},t) + \int dt'\, d\mathbf{r}'\, G(\mathbf{r},t,\mathbf{r}',t') e\mathbf{E}(t') \cdot \mathbf{r}' \Psi_i(\mathbf{r}',t'),$$

$$G(\mathbf{r},t,\mathbf{r}',t') = -\frac{i\Theta(t-t')}{(2\pi^2)^{3/2}\hbar} \int d\mathbf{k} \exp\left[i\left(\left(\mathbf{k} + \frac{e\mathbf{E}_0}{\hbar\omega}\sin\omega t \right) \cdot \mathbf{r} - \left(\mathbf{k} + \frac{e\mathbf{E}_0}{\hbar\omega}\sin\omega t' \right) \cdot \mathbf{r}' - \int_{t'}^{t} d\tau \frac{(\hbar\mathbf{k} + (e\mathbf{E}_0/\omega)\sin\omega\tau)^2}{2M\hbar} \right) \right], \tag{11}$$

where $\Theta(x)$ is a Heaviside step function: $\Theta(x) = 1$ for $x > 0$ and $\Theta(x) = 0$ for $x < 0$.

The intensity of the dipole radiation at the frequency $\Omega = n\omega$ is proportional to $|\mathbf{r}_n|^2$, where $\mathbf{r}_n$ is the corresponding Fourier component of the mean value of the coordinate operator

$$\mathbf{r}_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} d(\omega t) \exp(in\omega t) \langle \mathbf{r}(t) \rangle,$$

$$\langle \mathbf{r}(t) \rangle = \int dt'\, d\mathbf{r}'\, d\mathbf{r}\, \Psi_i^*(\mathbf{r},t) \mathbf{r} G(\mathbf{r},t,\mathbf{r}',t') e\mathbf{E}(t') \cdot \mathbf{r}' \Psi_i(\mathbf{r}',t') + \text{h.c.} \tag{12}$$

After several transformations with consideration of formulas (2) and (11), we represent (12) in the form

$$\mathbf{r}_n=-\frac{e\hbar^3}{\sqrt{2\pi^4}}\sum_{l=-\infty}^{\infty}\int d\mathbf{k}\int_{-\pi}^{\pi}d\varphi_1\,d\varphi_2\exp\left[i\left(\boldsymbol{\alpha}\cdot\mathbf{k}(\cos\varphi_1-\cos\varphi_2)+\frac{\beta}{2}(\sin 2\varphi_1-\sin 2\varphi_2)\right)\right](\mathbf{D}_1(\mathbf{k}(\varphi_1))\exp(i\rho\sin\varphi_1)$$
$$+\mathbf{D}_2(\mathbf{k}(\varphi_1))\exp(-i\rho\sin\varphi_1))\frac{\mathbf{E}(\varphi_1)\exp(i(l\varphi_1+(n-l)\varphi_2))+\mathbf{E}(\varphi_2)\exp(i(l\varphi_2+(n-l)\varphi_1))}{\hbar^2\mathbf{k}^2/(2M)+I+(\beta-l)\hbar\omega}$$
$$\times(\mathbf{D}_1(\mathbf{k}(\varphi_2))\exp(-i\rho\sin\varphi_2)+\mathbf{D}_2(\mathbf{k}(\varphi_2))\exp(i\rho\sin\varphi_2)), \tag{13}$$

where

$$\boldsymbol{\alpha}=e\mathbf{E}_0/(M\hbar\omega^2),\quad \mathbf{E}(\varphi)=\mathbf{E}_0\cos\varphi.$$

The exact calculation of the integrals runs up against serious difficulties. Therefore we use an approximate calculation method based on the so-called pole approximation.[17] In this approximation it is assumed that the integral, in the principal value sense, in the expression

$$\lim_{\varepsilon\to 0}\frac{1}{ix+\varepsilon}=-\pi\delta(x)+i\frac{p}{x},$$

where in our case

$$x=\frac{\hbar^2\mathbf{k}^2}{2M}+I+(\beta-l)\hbar\omega,$$

is small compared with the contribution of the delta function. Although there is no rigorous proof for the correctness of using the pole approximation, this approximation is probably reasonable for multiphoton transitions.[17,18] At worst, it can be expected that the corrections to the results will be of the same order of magnitude as the results within the pole approximation themselves. Utilizing the properties of the delta function formed in this approximation, we bring (13) into the form

$$|\mathbf{r}_n|^2=\frac{1}{2\pi^2}\left\{l\hbar^3\int d\mathbf{k}\int_{-\pi}^{\pi}d\varphi_1\,d\varphi_2\,\mathbf{A}_n^*(\varphi_1)\mathbf{E}(\varphi_2)\cdot\mathbf{A}_0(\varphi_2)\right.$$
$$\left.\times\sum_{l=-\infty}^{\infty}\delta\left(\frac{\hbar^2\mathbf{k}^2}{2M}+I+(\beta-l)\hbar\omega\right)\right\}^2,$$

$$\mathbf{A}_n(\varphi)=\left(\mathbf{D}_1(\mathbf{k}(\varphi))\exp\left[\frac{i}{\hbar\omega}\left(\int_0^{\varphi}d\varphi'\right.\right.\right.$$
$$\left.\left.\times\frac{(\hbar\mathbf{k}+(e\mathbf{E}_0/\omega)\sin\varphi')^2}{2M}+I\varphi-\rho\hbar\omega\sin\varphi\right)\right]$$
$$+\mathbf{D}_2(\mathbf{k}(\varphi))\exp\left[\frac{i}{\hbar\omega}\left(\int_0^{\varphi}d\varphi'\right.\right.$$
$$\left.\left.\left.\times\frac{(\hbar\mathbf{k}+(e\mathbf{E}_0/\omega)\sin\varphi')^2}{2M}+I\varphi+\rho\hbar\omega\sin\varphi\right)\right]\right.$$
$$\left.\times\exp(-in\varphi)\right)$$

(in the expression for $|\mathbf{r}_n|^2$ we have taken into account only the term making the leading contribution).

We assume that $I/(\hbar\omega)-\rho\gg 1$ and $n-I/(\hbar\omega)-\rho\gg 1$. Then the exponential functions are rapidly oscillating functions, and the integrals can be calculated by the saddle-point method. The following equations must be solved to find the saddle points:

$$\frac{(\hbar\mathbf{k}+(e\mathbf{E}_0/\omega)\sin\varphi_1^*)^2}{2M}+I\pm\rho\hbar\omega\cos\varphi_1^*-n=0, \tag{14}$$

$$\frac{(\hbar\mathbf{k}+(e\mathbf{E}_0/\omega)\sin\varphi_2^*)^2}{2M}+I\pm\rho\hbar\omega\cos\varphi_2^*=0. \tag{15}$$

We note that the integral over φ_2 is equal to the multiphoton ionization amplitude of the degenerate level (4) to within exponential accuracy [consequently, the equations for the saddle points (5) and (15) coincide]. Here, as in (4), the main contribution is made by small $\hbar\mathbf{k}$. On the other hand, for $I/(\hbar\omega)+\rho<n<I/(\hbar\omega)-\rho+2\beta$ and small $\hbar\mathbf{k}$ the roots of Eq. (14) are real; therefore, as can easily be seen, the integral over φ_1 is expressed in terms of an oscillating function. Thus, its value depends comparatively weakly on $\mathbf{k}$ in the sense that it does not decay exponentially with increasing $\mathbf{k}$, unlike the integral over φ_2. Therefore, for the values of n considered the main contribution to $\mathbf{r}_n$ is made by small $\hbar\mathbf{k}$, which are much smaller than the Bohr momentum. Now as an approximation we can set $\mathbf{k}=0$ in Eqs. (14) and (15), and then they are easily solved. For the plus sign in front of ρ in Eqs. (14) and (15)

$$\varphi_1^*=\pm\sin^{-1}\gamma_n^{(-)};\quad \pm\sin^{-1}\gamma_n^{(+)}\mp\pi,$$
$$\varphi_2^*=i\sinh^{-1}\gamma+\pi$$

and for the minus sign in front of ρ in (14) and (15)

$$\varphi_1^*=\pm\sin^{-1}\gamma_n^{(+)};\quad \pm\sin^{-1}\gamma_n^{(-)}\mp\pi,$$
$$\varphi_2^*=i\sinh^{-1}\gamma,$$

where

$$\gamma_n^{(\pm)}=\sqrt{\frac{n}{2\beta}-\gamma_0^2-\gamma_\rho^2\pm\sqrt{2\gamma_\rho^2\left(\gamma_0^2+1-\frac{n}{2\beta}\right)+\gamma_\rho^4}}. \tag{16}$$

We suppose that the applied magnitude field is strong enough that $\gamma\ll 1$. We also assume that $I/(\hbar\omega)\ll n/(2\beta)\ll 1$. In this case we can set $\gamma_n^{(\pm)}\approx\sqrt{n/(2\beta)-\gamma_0^2\pm\sqrt{2}\gamma_\rho}$ in the exponential expressions containing $\gamma_n^{(\pm)}$, and we can set $\gamma_n^{(\pm)}\approx\sqrt{n/(2\beta)}$ in the preexponential expressions. Then we can obtain a fairly simple expression for the square of the modulus of the matrix element

$$|\mathbf{r}_n|^2 \sim \frac{1}{\sqrt{n}}\left[\mathbf{D}_1\left(\sqrt{\frac{2nM\omega}{\hbar}}\right)+(-1)^{\left[\frac{I}{\hbar\omega}+\beta-n\right]}\mathbf{D}_2\right.$$
$$\times\left(\sqrt{\frac{2nM\omega}{\hbar}}\right)\Bigg]^2\exp\left[-\frac{4}{3}\frac{I-\rho\hbar\omega}{\hbar\omega}\sqrt{\frac{I-\rho\hbar\omega}{2\beta\hbar\omega}}\right]$$
$$\times\left(\cos\left(\frac{2}{3}\left(n-\frac{I}{\hbar\omega}+\rho\right)\sqrt{\frac{n-\frac{I}{\hbar\omega}+\rho}{2\beta}}-\frac{\pi}{4}\right)\right.$$
$$\left.\left.+\cos\left(\frac{2}{3}\left(n-\frac{I}{\hbar\omega}-\rho\right)\sqrt{\frac{n-\frac{I}{\hbar\omega}-\rho}{2\beta}}-\frac{\pi}{4}\right)\right)^2. \quad (17)$$

As follows from Eq. (17), the intensities of the harmonics depend comparatively weakly on n, i.e., expression (17) obtained describes a plateau in the emission spectrum of the atom. Relation (16) permits finding the high-frequency cutoff of the emission spectrum. The high-frequency cutoff of the spectrum is characterized by an abrupt drop in the intensity of the emitted harmonics as the order of the harmonic increases. According to the relations obtained, this is achieved when all the roots φ_1^* become complex (since in that case the value of the integral over φ_1 transforms from an oscillating function into an exponentially decaying function), which occurs, in turn, when the $\gamma_n^{(\pm)}$ are complex numbers or $|\gamma_n^{(\pm)}|>1$. After a simple analysis of (16) with consideration of these conditions we find the high-frequency cutoff of the emission spectrum:

$$n_{\max}\sim\frac{I}{\hbar\omega}+2\beta+\frac{\rho^2}{8\beta}.$$

The results obtained allow us to conclude that degeneracy of the level promotes an increase in the intensity of the harmonics in comparison to the ordinary case of $\mathbf{d}=0$. In this case, as well as in multiphoton ionization, this increase has an exponential character. In addition, as can be seen from the last formula, because of the mixing of the levels forming the degenerate state, the high-frequency cutoff of the spectrum is shifted to higher frequencies by $\rho^2/(8\beta)$ in comparison to the case of $\mathbf{d}=0$ under the same conditions.

Generally speaking, calculations in the Keldysh–Faisal–Reiss approximation (due to neglect of the influence of the Coulomb potential on states in the continuous spectrum) provide only exponential accuracy in the results. Mainly for this reason, the final formulas (6) and (17) were also written to within a pre-exponential factor (with respect to the field). This, of course, makes it difficult to compare directly the numerical results with experimental results. Only their qualitative correspondence can be tested. For this reason we should discus the conditions under which the model considered in this paper is faithful to the real physical situation. The correctness of its use presumes that the intensity of the applied electromagnetic field is considerably less than the atomic intensity for the level from which ionization or harmonic generation occurs: $|\mathbf{E}_0|\ll|\mathbf{E}_{\rm at}|/n^{*4}$, where $|\mathbf{E}_{\rm at}|\sim 5\times10^9$ V/cm and n^* is the principal quantum number of the level. The time of action τ of the pulse of the electromagnetic field must be considerably shorter than the spontaneous relaxation time of the excited state: $\tau\ll\tau_{\rm sp}$ ($\tau_{\rm sp}\sim10^{-8}\,{\rm s}^{-1}$). In order to eliminate the resonant mixing of atomic states by the exciting field, its frequency should not be a multiple of the frequencies of transitions to other bound states of the atom. In addition, the relative shift of the levels forming the degenerate state should be considerably smaller than the frequency of the electromagnetic field.

We note that the results obtained are applicable not only to hydrogenic atoms, but also to other excited hydrogenic systems, for example, excitonic series in solid inert gases[19] or molecules which have an intrinsic dipole moment.

[1] N. B. Delone and V. P. Kraĭnov, Izv. Akad. Nauk SSSR Ser. Fiz. **45**, 2331 (1981) [Bull. Acad. Sci. USSR, Phys. Ser. **45**(12), 77 (1981)].
[2] I. Ya. Berson, Zh. Éksp. Teor. Fiz. **83**, 1276 (1982) [Sov. Phys. JETP **56**, 731 (1982)].
[3] M. V. Fedorov and A. M. Movsecian, J. Opt. Soc. Am. B **6**, 928 (1989).
[4] M. V. Fedorov and A. M. Movsecian, J. Opt. Soc. Am. B **6**, 1504 (1989).
[5] I. Bakosh, A. Kish, and M. L. Nachaeva, in *Multiphoton Ionization of Atoms (Proceedings of the Lebedev Physics Institute, Academy of Sciences of the USSR, Vol. 115)* [in Russian], Nauka, Moscow (1980), p. 96.
[6] N. B. Delone, V. P. Kraĭnov, and D. L. Shepelyanskiĭ, Usp. Fiz. Nauk **140**, 355 (1983) [Sov. Phys. Usp. **26**, 551 (1983)].
[7] V. A. Kovarskiĭ, Zh. Éksp. Teor. Fiz. **57**, 1217, 1613 (1969) [Sov. Phys. JETP **30**, 663, 872 (1970)].
[8] O. B. Prepelitsa, Opt. Spektrosk. **81**, 377 (1996) [Opt. Spectrosc. **81**, 340 (1996)].
[9] L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1965)].
[10] A. McPherson, G. Gibson, H. Jara *et al.*, J. Opt. Soc. Am. B **4**, 595 (1987).
[11] M. Ferray, A. L. Huillier, X. F. Li *et al.*, J. Phys. B **21**, L31 (1988).
[12] J. H. Eberly, Q. Su, and J. Jawainen, Phys. Rev. Lett. **62**, 1989 (1989).
[13] P. B. Corcum, Phys. Rev. Lett. **71**, 1994 (1993).
[14] A. L. Huillier and P. Balcou, Phys. Rev. Lett. **70**, 774 (1993).
[15] M. Lewenstein, P. Balcou, M. Ivanov *et al.*, Phys. Rev. A **49**, 2117 (1994).
[16] V. A. Kovarskiĭ, *Multiphoton Transitions* [in Russian], Shtiintsa, Kishinev (1974), p. 228.
[17] M. V. Fedorov, *Electrons in Intense Light Fields* [in Russian], Nauka, Moscow (1991), p. 223.
[18] D. F. Zaretskiĭ and É. A. Nersesov, Zh. Éksp. Teor. Fiz. **103**, 1191 (1993) [JETP **76**, 583 (1993)].
[19] G. Baldini, Phys. Rev. **128**, 1562 (1962).

Translated by P. Shelnitz

Thermophoretic transport of moderately large spherical and cylindrical particles in two-component gases

E. R. Shchukin

Joint Institute of High Temperatures, Russian Academy of Sciences, 127412 Moscow, Russia

N. N. Kareva, Yu. I. Yalamov, and Z. L. Shulimanova

Moscow Pedagogical University, Moscow, Russia

(Submitted October 3, 1997; resubmitted January 11, 1999)
Zh. Tekh. Fiz. **69**, 21–27 (August 1999)

The problem of the thermophoretic motion of moderately large solid spherical and cylindrical particles in a two-component gas is solved for $\mathrm{Re} \ll 1$. The formulas obtained permit direct estimation of the rate of thermophoretic motion of both single-layer and multilayer particles. Corrections which depend directly on the Knudsen number are taken into account in the derivation of these formulas. The thermal conductivity of the particles is assumed to be a function which depends on the radial coordinate. It is shown that thermal diffusion and the dependence of the thermal conductivity on the radial coordinate can have a significant influence on the rate of thermophoretic transport of particles. © *1999 American Institute of Physics.* [S1063-7842(99)00508-5]

In gases with an inhomogeneous temperature distribution, a thermophoretic force[1–4] acts on particles, causing ordered motion of the particles relative to the gaseous medium. The thermophoretic motion of particles takes place in channels of heat and mass exchangers,[4–7] in transilluminated regions of clouds and fog,[8] in the vicinity of drops washing particles,[9] and in filters intended for fine gas purification.[4–7] For this reason, the derivation of formulas which would permit estimation of the rate of the thermophoretic motion of particles is of not only scientific, but also practical interest.

The formulas which have been derived permit estimation of the thermophoretic velocity of moderately large spherical[1–3] and cylindrical[10] particles only in one-component gases with a constant value for the thermal conductivity of the particles. Under real conditions the thermophoretic motion of particles can occur not only in one-component, but also in multicomponent gases.[11,12] It should also be noted that particles which are either homogeneous or inhomogeneous with respect to their thermophysical properties appear in the composition of natural and industrial disperse systems in air.[11,13,14] Particles which are inhomogeneous with respect to their thermophysical particles are found, for example, in condensation aerosols from metallurgical plants[13] and chemical concerns[14] and in the emissions of chemical plants and motor vehicles,[11,15] and form as a result of the occurrence of natural processes such as volcanic activity. The formulas derived in this paper permit estimation of the thermophoretic velocity of moderately large solid spherical and cylindrical particles in a two-component gas with a thermal conductivity which depends on the radial coordinate. Particles with a Knudsen number in the range $0.01 \leqslant \mathrm{Kn} \leqslant 0.3$ are classified as moderately large particles.

Let us dwell first in greater detail on the solution of the problem of the thermophoretic motion of a spherical particle. In deriving the formula for the thermophoretic velocity we shall assume that the motion of a multilayer solid spherical particle with a radius R_N takes place in a two-component gaseous medium in the external field of a temperature gradient. The particle consists of N layers. The kth layer is bounded on its inner and outer sides by spherical surfaces with the radii R_{k-1} and R_k, respectively. Within each of the layers the thermal conductivity ε_k and its derivative with respect to the radial coordinate are continuous functions. The coefficient ε_k depends on the radial coordinate r. The subscript k in ε_k indicates the number of the layer. The inner core is noted by $k=1$, and the outermost layer is denoted by $k=N$. Motion of the particles occurs at Reynolds and Peclet numbers much smaller than unity. A restriction is imposed on the temperature gradient:

$$\frac{f_T M_p |\nabla T_{e\infty}|}{6\pi R^2 \rho_e T_{e\infty}} \ll 1,$$

where M_p is the mass of a particle, ρ_e is the density of the gas, and f_T is a scalar coefficient, which appears in the expression for the thermophoretic velocity.[16]

When this condition is satisfied, the thermophoretic transport of particles can be described in a quasistationary approximation by virtue of the short relaxation times of the temperature, concentration, and gas-dynamic fields, as well as the short characteristic time for viscous drag of a particle in comparison to the time needed for the particle to traverse a distance equal to its radius.

Under the conditions considered the distributions of the mass velocity $\mathbf{v}$, the gas temperature T_e, the temperature of the particle layers T_k, and the relative concentration c_1 of molecules of the first component in a particle–gas-medium system are described by the following system of equations:[1–3,16]

$$\operatorname{div}\mathbf{v}=0,\quad \Delta\mathbf{v}=0,\quad \Delta T_e=0,$$
$$\operatorname{div}(\varepsilon_k\nabla T_k)=0,\qquad \Delta c_1=0. \tag{1}$$

Here $c_1=n_1/n$; $n=n_1+n_2$, where n_1 and n_2 are the concentrations of molecules of the first and second kinds in the gaseous medium; and Δ is the Laplacian operator.

In writing the Stokes equations it was taken into account that in the case of steady thermophoretic motion of a particle, the total force acting on the particle is equal to zero.[1–3] In this case thermophoretic transport of the particle takes place at a constant pressure. Thus, $\nabla p=0$ in the hydrodynamic equations.

The system (1) was solved in a spherical coordinate system, whose origin coincides with the center of the particle. It was assumed that the polar $0Z$ axis is parallel to the temperature gradient $\nabla T_{e\infty}$. In the spherical coordinate system the boundary conditions under which the system (1) was solved have the following form:[2,3,17]

$$v_r\Big|_{r=R_N}=C_r^{(c)}\,\mathrm{Kn}\,\frac{D}{R}\,\operatorname{div}_\Theta c_1+C_r^{(T)}\,\mathrm{Kn}\,\frac{\nu_e}{RT_e}\,\operatorname{div}_\Theta T_e\Big|_{r=R_N}, \tag{2}$$

$$\begin{aligned} v_\Theta\Big|_{r=R_N}=&\,C_\Theta\,\mathrm{Kn}\,R\left[r\frac{\partial}{\partial r}\left(\frac{v_\Theta}{r}\right)+\frac{1}{r}\frac{\partial v_r}{\partial\Theta}\right]+(C_D^{(0)}\\ &+C_D'\,\mathrm{Kn})\frac{D}{R}\frac{\partial c_1}{\partial\Theta}+C_D^{(R)}\mathrm{Kn}\,D\frac{\partial^2 c_1}{\partial r\,\partial\Theta}\\ &+C_D^{(B)}\mathrm{Kn}\,D\left[\frac{\partial^2 c_1}{\partial r\,\partial\Theta}-\frac{1}{r}\frac{\partial c_1}{\partial\Theta}\right]+(C_T^{(0)}\\ &+C_T'\,\mathrm{Kn})\frac{\nu_e}{RT_e}\frac{\partial T_e}{\partial\Theta}+C_T^{(R)}\,\mathrm{Kn}\,\nu_e\frac{1}{T_e}\frac{\partial^2 T_e}{\partial r\,\partial\Theta}\\ &-C_T^{(B)}\,\mathrm{Kn}\,\frac{\nu_e}{T_e}\left[\frac{\partial^2 T_e}{\partial r\,\partial\Theta}-\frac{1}{r}\frac{\partial T_e}{\partial\Theta}\right]\Big|_{r=R_N}, \end{aligned} \tag{3}$$

$$T_e-T_N|_{r=R_N}=C_T^{(T)}\,\mathrm{Kn}\,R_N\frac{\partial T_e}{\partial r}+C_T^{(n)}\,\mathrm{Kn}\,R_N T_e\frac{\partial c_1}{\partial r}\Big|_{r=R_N}, \tag{4}$$

$$\begin{aligned} &-\kappa\frac{\partial T_e}{\partial r}+\varepsilon_N\frac{\partial T_N}{\partial r}-\kappa T_e C_q^{(k)}\frac{\partial c_1}{\partial r}\Big|_{r=R_N}\\ &\quad=-T_e C_q^{(c)}\mathrm{Kn}\frac{\kappa}{R}\operatorname{div}_\Theta c_1-C_q^{(T)}\mathrm{Kn}\frac{\kappa}{R}\operatorname{div}_\Theta T_e\Big|_{r=R_N}, \end{aligned} \tag{5}$$

$$\begin{aligned} &\left(\frac{\partial c_1}{\partial r}+K_T\frac{1}{T_e}\frac{\partial T_e}{\partial r}\right)\Big|_{r=R_N}\\ &\quad=C_\nu^{(c)}\,\mathrm{Kn}\,\frac{1}{R}\,\operatorname{div}_\Theta c_1+C_\nu^{(T)}\,\mathrm{Kn}\,\frac{1}{RT_e}\,\operatorname{div}_\Theta T_e\Big|_{r=R_N}, \end{aligned} \tag{6}$$

$$T_k|_{r=R_k}=T_{k+1}|_{r=R_k},$$

$$-\varepsilon_k\frac{\partial T_k}{\partial r}\Big|_{r=R_k}=-\varepsilon_{k+1}\frac{\partial T_{k+1}}{\partial r}\Big|_{r=R_k}, \tag{7}$$

$$v_r|_{r\to\infty}=-u_{TZ}\cos\Theta,\quad v_\Theta|_{r\to\infty}=u_{TZ}\sin\Theta, \tag{8}$$

$$T_e|_{r\to\infty}=T_{e\infty}+|\nabla T_{e\infty}|r\cos\Theta,\quad c_1|_{r\to\infty}=c_{1\infty}, \tag{9}$$

where

$$\mathrm{Kn}=\frac{\lambda}{R_N};\quad \lambda=\mu_e\left(\frac{\pi}{2n\rho_e kT_e}\right)^{1/2};$$

$$\operatorname{div}_\Theta=\frac{\partial^2}{\partial\Theta^2}+\cot\Theta\,\frac{\partial}{\partial\Theta};$$

κ is Boltzmann's constant; $\nu_e=\mu_e/\rho_e$ is the kinematic viscosity; μ_e is the dynamic viscosity; $\rho_e=n_1m_1+n_2m_2$ is the density of the gas; D and κ are the interdiffusion coefficient and thermal conductivity of the gas mixture; K_T is the thermal-diffusion ratio; u_{TZ} is the projection of the thermophoretic velocity onto the polar $0Z$ axis; and the index k in (7) takes values from $k=1$ to $k=N-1$.

The boundary conditions on the particle surface (2)–(6) were written with allowance for corrections which depend on the Knudsen number.[2,3,17] In (2)–(6) $C_D^{(0)}$, $C_T^{(0)}$, and C_Θ are the diffusive, thermal, and isothermal slip coefficients; C_D' and C_T' are corrections for the surface curvature; $C_D^{(B)}$ and $C_T^{(B)}$ are the Barnett slip coefficients, which appear as a result of consideration of the terms proportional to the second derivatives of the relative concentrations of the gaseous components and the temperature (with respect to r and Θ) in the distribution function of the gas molecules in the Knudsen layer; $C_D^{(R)}$ and $C_T^{(R)}$ are coefficients associated with the inhomogeneity of the gradients of the relative concentrations of the gaseous components and the temperature (the appearance of these corrections is also attributed to the surface curvature); $C_T^{(T)}$ and $C_T^{(n)}$ are the temperature jump coefficients; $C_\nu^{(T)}$ and $C_\nu^{(c)}$, $C_q^{(T)}$ and $C_q^{(c)}$, and $C_r^{(T)}$ and $C_r^{(c)}$ are, respectively, the gas-kinetic coefficients of the diffusion and heat fluxes, as well as of the mass-average flux flowing in the Knudsen layer; $C_q^{(k)}=DknK_T/\kappa c_1c_2$ (Ref. 14); and $C_q^{(c)}=DknC_q/\kappa$ (Ref. 14). When estimates are obtained, the values of the gas-kinetic coefficients appearing in (2)–(6) must correspond to $c_1=c_{1\infty}$, $T_e=T_{e\infty}$, and $p=p_\infty$.

The boundary conditions (2)–(6) were written in accordance with the results of Refs. 2, 3, and 17. Not only expressions for the gas-kinetic coefficients appearing in (2)–(6), but also concrete values of these coefficients were given in those papers. In the general case the gas-kinetic coefficients appearing in the boundary conditions together with the Knudsen number are functions which depend on Kn.

The solutions of the system of equations (1) which satisfy the boundary conditions at infinity have the following form:

$$v_r=\left(\frac{E}{y^3}-u_{TZ}\right)\cos\Theta,\quad v_\Theta=\left(\frac{E}{2y^3}+u_{TZ}\right)\sin\Theta, \quad (10)$$

$$T_e=T_{e\infty}+R_N\left(y+\frac{1}{y^2}A\right)|\nabla T_{e\infty}|\cos\Theta,$$

$$T_k=T_{e\infty}+R_N(B_{k,1}\varphi_{k,1}+B_{k,2}\varphi_{k,2})|\nabla T_{e\infty}|\cos\Theta, \quad (11)$$

$$c_1=c_{1\infty}+R_N M\,\frac{1}{y^2}\frac{1}{T_{e\infty}}|\nabla T_{e\infty}|\cos\Theta, \quad (12)$$

where E, M, A, $B_{k,1}$, and $B_{k,2}$ are integration constants, and $y=r/R_N$.

The functions $\varphi_{k,1}$ and $\varphi_{k,2}$ are linearly independent particular solutions of the equation

$$\varepsilon_k y^2\frac{d^2\varphi_k}{dy^2}+\frac{d}{dy}(\varepsilon_k y^2)\frac{d\varphi_k}{dy}-2\varepsilon_k\varphi_k=0, \quad (13)$$

which are divergent at $y=0$ ($\varphi_{k,1}$) and nondivergent at $y=0$ ($\varphi_{k,2}$).

Because of the finite value of the temperature at the center of the particle we have $B_{1,1}=0$. In the general case the dependence of $\varphi_{k,1}$ and $\varphi_{k,2}$ on y can be found during the numerical solution of Eq. (13). If the expression for ε_k in an annular region bounded by circles of radii y_{k-1} and y_k can be represented in the form of a convergent power series

$$\varepsilon_k=\varepsilon_k^{(0)}\sum_{n=0}^{\infty}\alpha_n^{(k)}y^n,\quad \alpha_0^{(k)}=1 \quad (14)$$

with a radius of convergence in the complex region $a\geqslant y_k$, the solutions $\varphi_{k,1}$ and $\varphi_{k,2}$ have the following form:

$$\varphi_{k,1}=y\left\{\sum_{n=-3}^{\infty}\beta_{n,k}^{(1)}y^n+\Delta^{(k)}(\ln y)\sum_{n=0}^{\infty}\beta_{n,k}^{(2)}y^n\right\},$$

$$\varphi_{k,2}=y\sum_{n=0}^{\infty}\beta_{n,k}^{(2)}y^n,\quad \beta_{0,k}^{(2)}=1. \quad (15)$$

After substituting (14) and (15) into Eq. (13) and then setting the coefficients in front of identical powers y^n equal to zero, we obtain the following recurrence relations for the coefficients $\beta_{n,k}^{(1)}$ and $\beta_{n,k}^{(2)}$:

$$\beta_{\substack{n,k\\ n\geqslant -2,n\neq 0}}^{(1)}=-\frac{1}{n(n+3)}\left\{\Delta^{(k)}(2n+3)\beta_{n,k}^{(2)}+\sum_{m=1}^{n+3}[((n-m)(n+3)+m)\beta_{n-m,k}^{(1)}+\Delta^{(k)}(2n-m+3)\beta_{n-m,k}^{(2)}]\alpha_m^{(k)}\right\},$$

$$\beta_{\substack{n,k\\ n\geqslant 1}}^{(2)}=-\frac{1}{n(n+3)}\sum_{m=1}^{n}[(n-m)(n+3)+m]\alpha_m^{(k)}\beta_{n-m,k}^{(2)},\quad \beta_{0,k}^{(2)}=1. \quad (16)$$

In determining the value of $\beta_{n,k}^{(1)}$ it must be taken into account that

$$\beta_{-3,k}^{(1)}=1,\quad \beta_{0,k}^{(1)}=0,\quad \Delta^{(k)}=\frac{2}{3}\sum_{m=1}^{3}m\alpha_m^k\beta_{-m,k}^{(1)}.$$

For $\varepsilon_k=\text{const}$ we have $\varphi_{k,1}=1/y^2$ and $\varphi_{k,2}=y$, respectively. If $\varepsilon_k=\varepsilon_k^{(0)}\exp(\alpha^{(k)}y)$, $\varepsilon_k^{(0)}=\text{const}$, and $\alpha^{(k)}=\text{const}$, then

$$\varphi_{k,1}=\frac{1}{y^2}\exp(-\alpha^{(k)}y),$$

$$\varphi_{k,2}=3\left[\left(\frac{1}{\alpha^{(k)}}-\frac{2}{y\alpha^{(k)2}}+\frac{2}{y^2\alpha^{(k)3}}\right)-\frac{2}{y^2\alpha^{(k)3}}\exp(-\alpha^{(k)}y)\right].$$

In the case of $\varepsilon_k=\varepsilon_k^{(0)}(1+\alpha^{(k)}y)^{\gamma(k)}$ and $|\alpha^{(k)}|<1$,

$$\varphi_{k,2}=y\left\{\sum_{n=-3}^{\infty}\beta_{n,k}^{(1)}y^n+\Delta^{(k)}(\ln y)\sum_{n=0}^{\infty}\beta_{n,k}^{(2)}y^n\right\},$$

$$\varphi_{k,2}=y\sum_{n=0}^{\infty}\beta_{n,k}^{(2)}y^n,\quad \beta_{0,k}^{(2)}=1. \quad (17)$$

The recurrence relations for $\beta_{n,k}^{(1)}$ and $\beta_{n,k}^{(2)}$ are

$$\beta_{\substack{n,k\\ n\geqslant -2,n\neq 0}}^{(1)}=-\frac{1}{n(n+3)}\{\alpha^{(k)}(n^2+n(1+\gamma^{(k)})-2)\times\beta_{n-1,k}^{(1)}+\Delta^{(k)}(2n+3)\beta_{n,k}^{(2)}+\alpha^{(k)}\Delta^{(k)}(2n+1+\gamma^{(k)})\beta_{n-1,k}^{(2)}\},$$

$$\beta_{\substack{n,k\\ n\geqslant 1}}^{(2)}=-\alpha^{(k)}\frac{[n^2+n(1+\gamma^{(k)})-2]}{n(n+3)}\beta_{n-1,k}^{(2)},$$

$$\beta_{0,k}^{(2)}=1. \quad (18)$$

In finding the values of $\beta_{n,k}^{(1)}$ it must be taken into account that $\beta_{-3,k}^{(1)}=1$, $\beta_{0,k}^{(1)}=0$, and $\Delta^{(k)}=(2/3)\alpha^{(k)}\beta_{-1,k}^{(1)}$. It follows from (18) that the values of $\beta_{n,k}^{(2)}$ can be found using the formula

$$\beta_{\substack{n,k\\ n\geqslant 1}}^{(2)}=(-\alpha^{(k)})^n\prod_{m=1}^{n}\frac{[m^2+(1+\gamma^{(k)})m-2]}{m(m+3)},$$

$$\beta_{0,k}^{(2)}=1.$$

The substitution of (10)–(12) into the boundary conditions (2)–(7) followed by the elimination of $\cos\Theta$ and $\sin\Theta$ from them yielded a system of linear algebraic equations relative to the unknowns E, A, M, $B_{k,1}$, $B_{k,2}$, and u_{TZ}. In the course of solving this linear system we obtained an expression for the projection u_{TZ}. The formula for the thermophoretic velocity obtained with consideration of this expression has the following form:

$$\mathbf{u}_T=-f_T\frac{\nu_e}{T_{e\infty}}\nabla T_{e\infty}, \quad (19)$$

$$f_T=\frac{1}{(1+2C_\Theta \mathrm{Kn})(1-C_v^{(c)}\mathrm{Kn})d_1}$$
$$\times[d_2\{[(1-C_v^{(c)}\mathrm{Kn})-(C_q^{(k)}-C_q^{(c)}\mathrm{Kn})K_T]\kappa\Delta_1$$
$$+[(1-C_v^{(c)}\mathrm{Kn})C_T^{(T)}-K_TC_T^{(n)}]\mathrm{Kn}\varepsilon_N^{(2)}\Delta_2\}$$
$$+d_3\{[1-\mathrm{Kn}(C_v^{(c)}-2C_v^{(T)}C_T^{(n)}\mathrm{Kn})]\varepsilon_N^{(2)}\Delta_2$$
$$-2[C_q^{(T)}(1-C_q^{(c)}\mathrm{Kn})-C_v^{(T)}(C_q^{(k)}$$
$$-C_q^{(c)}\mathrm{Kn})]\mathrm{Kn}\,\kappa\Delta_1\}]. \qquad (20)$$

In Eq. (20) κ is the thermal conductivity of the gas,

$$d_1=\{2[(1-C_q^{(T)}\mathrm{Kn})(1-C_v^{(c)}\mathrm{Kn})+(C_q^{(k)}-C_q^{(c)}\mathrm{Kn})$$
$$\times(C_v^{(T)}\mathrm{Kn}-K_T)]\kappa\Delta_1+[(1+2C_T^{(T)}\mathrm{Kn})$$
$$\times(1-C_v^{(c)}\mathrm{Kn})+2(C_v^{(T)}\mathrm{Kn}-K_T)C_T^{(n)}\mathrm{Kn}]\varepsilon_N^{(2)}\Delta_2\},$$

$$d_2=\Bigg\{[C_T^{(0)}+\mathrm{Kn}((C_T'+C_T^{(B)})-(1+6C_\Theta\mathrm{Kn})C_r^{(T)})]$$
$$\times(1-C_v^{(c)}\mathrm{Kn})+\frac{D}{\nu_e}[C_D^{(0)}+\mathrm{Kn}((C_D'-3C_D^{(B)}$$
$$-2C_D^{(R)})-(1+6C_\Theta\mathrm{Kn})C_r^{(c)})]C_v^{(T)}\mathrm{Kn}\Bigg\},$$

$$d_3=\Bigg\{\mathrm{Kn}(C_T^{(R)}-C_T^{(B)}\mathrm{Kn})(1-C_v^{(c)})$$
$$+\frac{D}{2\nu_e}[C_D^{(0)}+\mathrm{Kn}((C_D'-3C_D^{(B)}-2C_D^{(R)})$$
$$-(1+6C_\Theta\mathrm{Kn})C_r^{(c)})]K_T\Bigg\}.$$

The expressions for the coefficients Δ_1 and Δ_2 have the form

$$\Delta_1=\det\|a_{im}\|,\quad \Delta_2=\det\|b_{im}\|,$$
$$1\leqslant i\leqslant 2N-1,\quad 1\leqslant m\leqslant 2N-1. \qquad (21)$$

The values of the matrix elements a_{im} and b_{im} are found using the formulas

$$a_{2k-1,2k-2}=\varphi_{k,1}^{(2)},\quad a_{2k-1,2k}=\varphi_{k,1}^{(2)},$$
$$a_{2k-1,2k}=-\varphi_{k+1,1}^{(1)},$$
$$a_{2k-1,2k+1}=-\varphi_{k+1,2}^{(1)},\quad 1\leqslant k\leqslant N-1,$$
$$a_{2k-1,m\leqslant 2k-3}=0,$$
$$a_{2k-1,m\geqslant 2k+2}=0,\quad a_{2k,2k-2}=\varepsilon_k^{(2)}\frac{d\varphi_{k,1}^{(2)}}{dy},$$
$$a_{2k,2k-1}=\varepsilon_k^{(2)}\frac{d\varphi_{k,2}^{(2)}}{dy},\quad a_{2k,2k}=-\varepsilon_{k+1}^{(1)}\frac{d\varphi_{k+1,1}^{(1)}}{dy},$$
$$a_{2k,2k+1}=-\varepsilon_{k+1}^{(1)}\frac{d\varphi_{k+1,2}^{(1)}}{dy},\quad 1\leqslant k\leqslant N-1,$$
$$a_{2k,m\leqslant 2k-3}=0,\quad a_{2k,m\geqslant 2k+2}=0,\quad a_{2N-1,m\leqslant 2N-3}=0,$$
$$a_{2N-1,2N-2}=\varphi_{N,1}^{(2)},\quad a_{2N-1,m\leqslant 2N-3}=0,\quad b_{im}=a_{im},$$
$$1\leqslant i\leqslant 2N-2,\quad 1\leqslant m\leqslant 2N-1,\quad b_{2N-1,m\leqslant 2N-3}=0,$$
$$b_{2N-1,2N-2}=\frac{d\varphi_{N,1}^{(2)}}{dy},\quad b_{2N-1,2N-1}=\frac{d\varphi_{N,2}^{(2)}}{dy}, \qquad (22)$$

in which the index k takes values from $k=1$ to $k=N-1$;

$$y_k=R_k/R_N,\quad \varphi_{k,1}^{(1)}=\varphi_{k,1},\quad \varphi_{k,2}^{(1)}=\varphi_{k,2}|_{y=y_{k-1}},$$
$$\frac{d\varphi_{k,1}^{(1)}}{dy}=\frac{d\varphi_{k,1}}{dy}\Bigg|_{y=y_{k-1}},\quad \frac{d\varphi_{k,2}^{(1)}}{dy}=\frac{d\varphi_{k,2}}{dy}{}_{y=y_{k-1}},$$
$$\varepsilon_k^{(1)}=\varepsilon_k|_{y=y_{k-1}},\quad \varphi_{k,1}^{(2)}=\varphi_{k,1}|_{y=y_k},\quad \varphi_{k,2}^{(2)}=\varphi_{k,2}|_{y=y_k},$$
$$\frac{d\varphi_{k,1}^{(2)}}{dy}=\frac{d\varphi_{k,1}}{dy}\Bigg|_{y=y_k},\quad \frac{d\varphi_{k,2}^{(2)}}{dy}=\frac{d\varphi_{k,2}}{dy}\Bigg|_{y=y_k},\quad \varepsilon_k^{(2)}=\varepsilon_k|_{y=y_k}.$$

In single-layer particles ($N=1$) the expressions for Δ_1 and Δ_2 are

$$\Delta_1=\varphi_{1,2}^{(2)},\quad \Delta_2=\frac{d\varphi_{1,2}^{(2)}}{dy}. \qquad (23)$$

When $\varepsilon_1=\mathrm{const}$, the coefficients (23) take the values $\Delta_1=1$ and $\Delta_2=1$. In the case of a two-layer particle

$$\Delta_1=\Bigg[\varepsilon_1^{(2)}(\varphi_{2,1}^{(1)}\varphi_{2,2}^{(2)}-\varphi_{2,2}^{(1)}\varphi_{2,1}^{(2)})\frac{d\varphi_{1,2}^{(2)}}{dy}$$
$$-\varepsilon_2^{(1)}\left(\varphi_{2,2}^{(2)}\frac{d\varphi_{2,1}^{(1)}}{dy}-\varphi_{2,1}^{(2)}\frac{d\varphi_{2,2}^{(1)}}{dy}\right)\varphi_{1,2}^{(2)}\Bigg],$$
$$\Delta_2=\Bigg[\varepsilon_1^{(2)}\left(\varphi_{2,1}^{(1)}\frac{d\varphi_{2,2}^{(2)}}{dy}-\varphi_{2,2}^{(1)}\frac{d\varphi_{2,1}^{(2)}}{dy}\right)\frac{d\varphi_{1,2}^{(2)}}{dy}$$
$$-\varepsilon_2^{(1)}\left(\frac{d\varphi_{2,2}^{(2)}}{dy}\frac{d\varphi_{2,1}^{(1)}}{dy}-\frac{d\varphi_{2,1}^{(2)}}{dy}\frac{d\varphi_{2,2}^{(1)}}{dy}\right)\varphi_{1,2}^{(2)}\Bigg]. \qquad (24)$$

When the coefficients ε_1 and ε_2 have constant values, formulas (24) become

$$\Delta_1=-y_1\left[\varepsilon_1\left(\frac{1}{y_1^3}-1\right)+\varepsilon_2\left(\frac{2}{y_1^3}+1\right)\right],$$
$$\Delta_2=-y_1\left[\varepsilon_1\left(\frac{1}{y_1^3}+2\right)+2\varepsilon_2\left(\frac{1}{y_1^3}-1\right)\right].$$

In the case of a one-component gas, the expression for f_T (20) takes the following, considerably simpler form:

$$f_T=2\{[C_T^{(0)}+\mathrm{Kn}((C_T'+C_T^{(B)})-(1+6C_\Theta\mathrm{Kn})C_r^{(T)})]$$
$$\times(\kappa\Delta_1+\varepsilon_N^{(2)}C_T^{(T)}\mathrm{Kn}\Delta_2)+\mathrm{Kn}(C_T^{(R)}-C_T^{(B)})$$
$$\times(\varepsilon_N^{(2)}\Delta_2-2C_q^{(T)}\mathrm{Kn}\kappa\Delta_1)\}\frac{1}{(1+2C_\Theta\mathrm{Kn})d_1},$$

where $d_1=2(1-C_q^{(T)}\mathrm{Kn})\kappa\Delta_1+(1+2C_T^{(T)}\mathrm{Kn})\varepsilon_N^{(2)}\Delta_2$.

The problem of the thermophoretic motion of a moderately large multilayer cylindrical particle oriented perpendicularly to $\nabla T_{e\infty}$ is solved in analogy to the foregoing. An expression for the projection of the thermophoretic velocity

u_T onto the polar $0Z$ axis of the cylindrical coordinate system is found during the solution of the boundary-value problem (1)–(9), in which r and Θ are the cylindrical coordinates, and $\mathrm{div}_\Theta = \partial^2/\partial\Theta^2$. The solutions of the system (1) in the case of a cylindrical particle are

$$v_r = \left(\frac{E}{y^2} - u_{TZ}\right)\cos\Theta, \quad v_\Theta = \left(\frac{E}{y^2} + u_{TZ}\right)\sin\Theta, \tag{25}$$

$$T_e = T_{e\infty} + R_N\left(y + \frac{A}{y}\right)|\nabla T_{e\infty}|\cos\Theta,$$

$$T_k = T_{e\infty} + R_N(B_{k,1}\varphi_{k,1} + B_{k,2}\varphi_{k,2})|\nabla T_{e\infty}|\cos\Theta, \tag{26}$$

$$c_1 = c_{1\infty} + R_N\frac{M}{y}\frac{1}{T_{e\infty}}|\nabla T_{e\infty}|\cos\Theta, \tag{27}$$

where E, A, M, $B_{k,1}$, and $B_{k,2}$ are integration constants; $y = r/R_N$; and $B_{1,1} = 0$.

The functions $\varphi_{k,1}$ and $\varphi_{k,2}$ are linearly independent particular solutions of the equation

$$\varepsilon_k y^2\frac{d^2\varphi_k}{dy^2} + y\frac{d}{dy}(\varepsilon_k y)\frac{d\varphi_k}{dy} - \varepsilon_k\varphi_k = 0, \tag{28}$$

which are divergent at $y = 0$ ($\varphi_{k,1}$) and nondivergent at $y = 0$ ($\varphi_{k,2}$).

The integration of (28) can be performed numerically or in quadratures. If the expression for ε_k can be represented in the form of a convergent power series (14), the particular solutions of (28) become

$$\varphi_{k,1} = y\left\{\sum_{n=-2}^{\infty}\beta_{n,k}^{(1)}y^n + \Delta^{(k)}(\ln y)\sum_{n=0}^{\infty}\beta_{n,k}^{(2)}y^n\right\},$$

$$\varphi_{k,2} = y\sum_{n=0}^{\infty}\beta_{n,k}^{(2)}y^n, \quad \beta_{0,k}^{(20)} = 1. \tag{29}$$

The values of the coefficients $\beta_{n,k}^{(1)}$ and $\beta_{n,k}^{(2)}$ appearing in (29) are found from the following formulas:

$$\beta_{\substack{n\geqslant -1,\\ n\neq 0}}^{(1)} = -\frac{1}{n(n+2)}\left\{\sum_{m=1}^{n+2}[(n-m)(n+1)+n]\times\beta_{n-m,k}^{(1)} + \Delta^{(k)}(2n-m+2)\beta_{n-m,k}^{(2)}]\alpha_m^{(k)} + 2\Delta^{(k)}(n+1)\beta_{n,m}^{(2)}\right\},$$

$$\beta_{\substack{n\geqslant 1}}^{(2)} = -\frac{1}{n(n+2)}\sum_{m=1}^{n}[(n-m)(n+1) + n]\alpha_m^{(k)}\beta_{n-m,k}^{(2)}, \quad \beta_{0,k}^{(2)} = 1. \tag{30}$$

In determining $\beta_{n,k}^{(1)}$, it must be taken into account that

$$\beta_{-2,k}^{(1)} = 1, \quad \beta_{0,k}^{(1)} = 0, \quad \Delta^{(k)} = \frac{1}{2}\sum_{m=1}^{2} m\beta_{-m,k}^{(1)}\alpha_m^{(k)}.$$

When $\varepsilon_k = \mathrm{const}$, we have $\varphi_{k,1} = 1/y$ and $\varphi_{k,2} = y$, respectively. If $\varepsilon_k = \varepsilon_k^{(0)}\exp(\alpha^{(k)}y)$, then $\varphi_{k,1} = (1/y)\exp(-\alpha^{(k)}y)$, and

$$\varphi_{k,2} = 2\left[\left(\frac{1}{\alpha_{(k)}} - \frac{1}{y\alpha^{(k)2}}\right) + \frac{1}{y\alpha^{(k)2}}\exp(-\alpha^{(k)}y)\right].$$

In the case of $\varepsilon_k = \varepsilon_k^{(0)}(1+\alpha^{(k)}y)^{\gamma^{(k)}}$ and $|\alpha^{(k)}| < 1$, we have

$$\varphi_{k,1} = y\left\{\sum_{n=-2}^{\infty}\beta_{n,k}^{(1)}y^n + \Delta^{(k)}(\ln y)\sum_{n=0}^{\infty}\beta_{n,k}^{(2)}y^n\right\},$$

$$\varphi_{k,2} = y\sum_{n=0}^{\infty}\beta_{n,k}^{(2)}y^n, \quad \beta_{0,k}^{(2)} = 1.$$

The recurrence relations for the coefficients $\beta_{n,k}^{(1)}$ and $\beta_{n,k}^{(2)}$ (30) have the following forms

$$\beta_{\substack{n\geqslant -1,\\ n\neq 0}}^{(1)} = -\frac{1}{n(n+2)}\{\alpha^{(k)}(n^2 + n\gamma^{(k)} - 1)\beta_{n-1,k}^{(1)} + 2\Delta^{(k)}(n+1)\beta_{n,k}^{(2)} + \alpha^{(k)}\Delta^{(k)}(2n + \gamma^{(k)})\beta_{n-1,k}^{(2)}\},$$

$$\beta_{\substack{n\geqslant 1}}^{(2)} = -\alpha^{(k)}\frac{[n^2 + n\gamma^{(k)} - 1]}{n(n+2)}\beta_{n-1,k}^{(2)}, \quad \beta_{0,k}^{(2)} = 1.$$

In finding the values of $\beta_{n,k}^{(1)}$, it must be taken into account that

$$\beta_{-2,k}^{(1)} = 1, \quad \beta_{0,k}^{(1)} = 0, \quad \Delta^{(k)} = \frac{1}{2}\alpha^{(k)}\beta_{-1,k}^{(1)}.$$

The values of the coefficients $\beta_{n,k}^{(2)}$ (30) can be found from the formula

$$\beta_{\substack{n\geqslant 1}}^{(2)} = (-\alpha^{(k)})^n\prod_{m=1}^{n}\frac{[m^2 + m\gamma^{(k)} - 1]}{m(m+2)}, \quad \beta_{0,k}^{(2)} = 1.$$

Also, in the case of cylindrical particles the expression for the thermophoretic velocity is equivalent to (19). The formula found during the solution of the boundary-value problem (1)–(9) for the coefficient f_T of multilayer cylindrical particles has the following form:

$$\begin{aligned} f_T = {} & \frac{1}{(1+2C_\Theta\mathrm{Kn})(1-C_v^{(c)}\mathrm{Kn})d_1} \\ & \times[d_2\{[(1-C_v^{(c)}\mathrm{Kn}) - (C_q^{(k)} - C_q^{(c)}\mathrm{Kn})K_T]\kappa\Delta_1 \\ & + [(1-C_v^{(c)}\mathrm{Kn})C_T^{(T)} - K_T C_T^{(n)}]\mathrm{Kn}\varepsilon_N^{(2)}\Delta_2\} \\ & + d_3\{[1 - \mathrm{Kn}(C_v^{(c)} - C_v^{(T)}C_T^{(n)}\mathrm{Kn})]\varepsilon_N^{(2)}\Delta_2 \\ & - [C_q^{(T)}(1 - C_v^{(c)}\mathrm{Kn}) - C_v^{(T)}(C_q^{(k)} - C_q^{(c)}\mathrm{Kn})]\mathrm{Kn}\kappa\Delta_1\}], \end{aligned} \tag{31}$$

where

$$\begin{aligned} d_1 = {} & \{[(1 - C_q^{(T)}\mathrm{Kn})(1 - C_v^{(c)}\mathrm{Kn}) + (C_q^{(k)} - C_q^{(c)}\mathrm{Kn}) \\ & \times(C_v^{(T)}\mathrm{Kn} - K_T)]\kappa\Delta_1 + [(1 + C_T^{(T)}\mathrm{Kn})(1 - C_v^{(c)}\mathrm{Kn}) \\ & + (C_v^{(T)}\mathrm{Kn} - K_T)C_T^{(n)}\mathrm{Kn}]\varepsilon_N^{(2)}\Delta_2\}, \end{aligned}$$

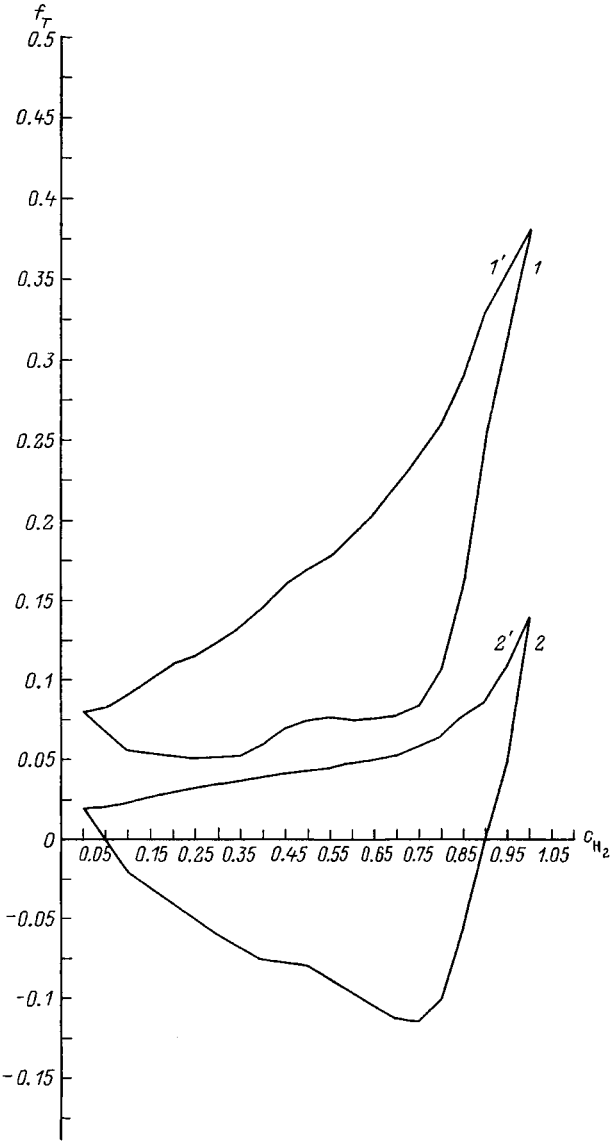

FIG. 1. Dependence of the coefficient f_T for large spherical particles on the relative concentration of hydrogen molecules.

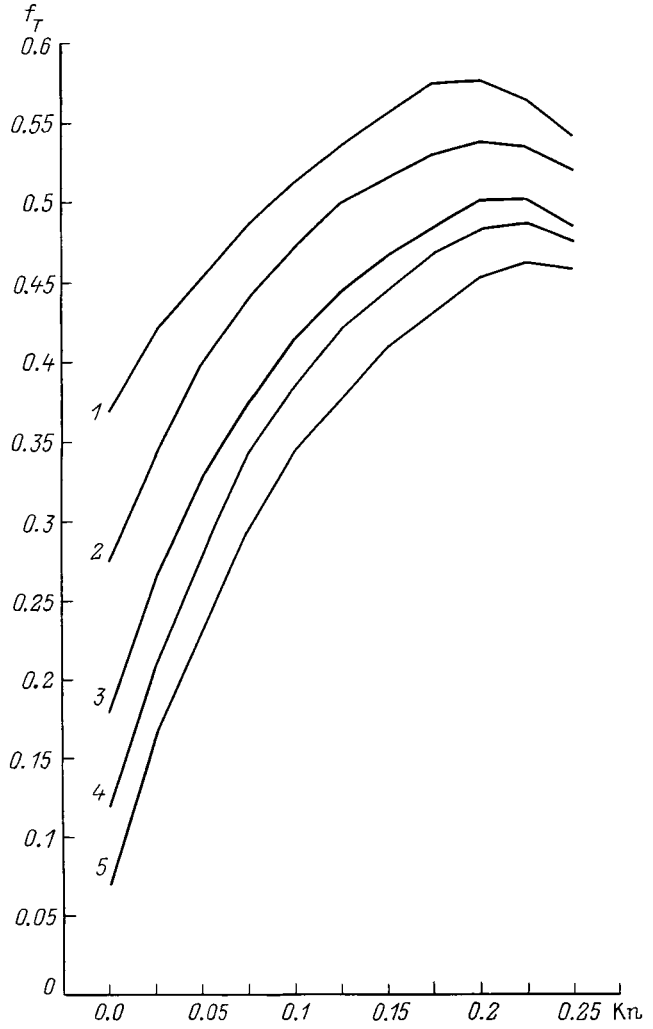

FIG. 2. Dependence of the coefficient f_T on the Knudsen number: $\alpha=4$ (*1*), 2 (*2,4*), 0 (*3*), 4 (*5*).

$$d_2=\Big\{[C_T^{(0)}+\mathrm{Kn}((C_T'+C_T^{(B)})-(1+4C_\Theta\mathrm{Kn})C_r^{(T)})]\times(1-C_v^{(c)}\mathrm{Kn})+\frac{D}{\nu_e}[C_D^{(0)}+\mathrm{Kn}((C_D'-2C_D^{(B)}-C_D^{(R)})-(1+4C_\Theta\mathrm{Kn})C_r^{(c)})]C_v^{(T)}\mathrm{Kn}\Big\},$$

$$d_3=\{\mathrm{Kn}(C_T^{(R)})-C_T^{(B)})(1-C_v^{(c)}\mathrm{Kn})+\frac{D}{\nu_e}[C_D^{(0)}+\mathrm{Kn}((C_D'-2C_D^{(B)}-C_D^{(R)})-(1+4C_\Theta\mathrm{Kn})C_r^{(c)})]K_T\}.$$

The values of the coefficients Δ_1 and Δ_2 are found from formulas (21). The solutions $\varphi_{k,1}$ and $\varphi_{k,2}$ of Eq. (28) must be used to determined the values of the determinant elements a_{im} and b_{im}. In the case of single-layer and two-layer particles, the expressions for Δ_1 and Δ_2 are equivalent to (23) and (24). If the particle moves in a one-component gas, the expression (31) for f_T takes the form

$$f_T=\frac{1}{(1+2C_\Theta\mathrm{Kn})d_1}\{[C_T^{(0)}+\mathrm{Kn}((C_T'+C_T^{(B)})-(1+4C_\Theta\mathrm{Kn})C_r^{(T)})](\kappa\Delta_1+\varepsilon_N^{(2)}C_T^{(T)}\mathrm{Kn}\Delta_2)+\mathrm{Kn}(C_T^{(R)}-C_T^{(B)})(\varepsilon_N^{(2)}\Delta_2-C_q^{(T)}\mathrm{Kn}\kappa\Delta_1)\},$$

where $d_1=(1-C_q^{(T)}\mathrm{Kn})\kappa\Delta_1+(1+C_T^{(T)}\mathrm{Kn})\varepsilon_N^{(2)}\Delta_2$.

The expressions (20) and (31) for f_T take their simplest forms in the case of large particles, for which $\mathrm{Kn}=0$:

$$f_T^{(1)}=\frac{2[C_T^{(0)}\kappa\Delta_1+C_D^{(0)}K_T(D/2\nu_e)\varepsilon_N^{(2)}\Delta_2]}{(2\kappa\Delta_1+\varepsilon_N^{(2)}\Delta_2)},$$

$$f_T^{(2)}=\frac{[C_T^{(0)}\kappa\Delta_1+C_D^{(0)}K_T(D/\nu_e)\varepsilon_N^{(2)}\Delta_2]}{(\kappa\Delta_1+\varepsilon_N^{(2)}\Delta_2)}.$$

The expression for $f_T^{(1)}$ can be used to find the values of f_T for spherical particles, and the expression for $f_T^{(2)}$ can be used for cylindrical particles. If the molecular weights are

not excessively large, the product appearing in $f_T^{(1)}$ and $f_T^{(2)}$ falls in the range $C_D^{(0)}K_T<0$. This means that diffusive slip causes slowing of the thermophoretic motion of large particles. The estimates made showed that in gas mixtures with strongly differing molecular weights the influence of diffusive slip on the thermophoretic velocity of large particles can be significant, to the point of changing the direction of motion. This is demonstrated quite well by the course of the plots of f_T in Fig. 1 for large spherical particles of glass (curves *1*) and granite (curves *2*) versus the relative concentration of hydrogen molecules. When the calculations were performed, it was assumed that the particles are in a gas mixture consisting of H_2 and N_2 molecules. Their motion occurs at $T_{e\infty}=293$ K and $p_\infty=101\,325$ Pa. Curves *1′* and *2′* were constructed without taking the influence of diffusive slip into account. The values of the gas-kinetic coefficients were taken with accommodation coefficients equal to unity.[3,17]

The dependence of the thermal conductivity of the particles on the radial coordinate can have a significant influence on the magnitude of the thermophoretic velocity. Plots of the dependence of the coefficient f_T for spherical particles on the Knudsen number for $\kappa/\varepsilon^{(S)}=0.1$ are shown in Fig. 2. The curves in Fig. 2 were constructed for the case of an exponential dependence of the coefficient ε for single-layer particles on $y=r/R$.

It was assumed in obtaining the estimates that the particle is immersed in air with $T_{e\infty}=293$ K and $p_\infty=101\,325$ Pa. The values of the gas-kinetic coefficients were taken from Ref. 2.

In conclusion, we thank the Soros Foundation for its material support of this scientific study.

[1] E. Z. Shchukin, Zh. Tekh. Fiz. **50**, 1332 (1980) [Sov. Phys. Tech. Phys. **25**, 767 (1980)].
[2] A. V. Poddoskin, A. A. Yushkanov, and Yu. I. Yalamov, Zh. Tekh. Fiz. **52**, 2253 (1982) [Sov. Phys. Tech. Phys. **27**, 1383 (1982)].
[3] E. G. Mayasov, A. A. Yushkanov, and Yu. I. Yalamov, Pis'ma Zh. Tekh. Fiz. **14**(6), 498 (1988) [Sov. Tech. Phys. Lett. **14**(3), 220 (1988)].
[4] M. Moo-Young and K. Yamaguchi, J. Chem. Eng. Sci. **30**, 1291 (1975).
[5] Yu. A. Vyskubenko and É. A. Tsalko, Teplofiz. Vys. Temp. **16**, 349 (1978) [High Temp. **16**, 287 (1978)].
[6] Yu. A. Vyskubenko, V. M. Maslenikov, and É. A. Tsalko, Teplofiz. Vys. Temp. **17**, 1037 (1979) [High Temp. **17**, 854 (1980)].
[7] E. R. Shchukin, Z. L. Shulimanova, and V. A. Zagainov, J. Aerosol Sci. **21**, 189 (1990).
[8] E. V. Zuev, A. A. Zemlyanov, and Yu. D. Kopytin, *High-Power Laser Radiation in Atmospheric Aerosols* [in Russian], Novosibirsk (1984), 223 pp.
[9] P. K. Wang and E. R. Proppucher, J. Atmos. Sci. **34**, 1664 (1977).
[10] R. A. Safiullin and Yu. I. Yalamov, Teplofiz. Vys. Temp. **32**, 271 (1994) [High Temp. **32**, 255 (1994)].
[11] H. L. Green and W. R. Lane, *Particulate Clouds: Dusts, Smokes and Mists*, 2nd ed. [Van Nostrand, Princeton (1964); Khimiya, Leningrad (1969), 428 pp.].
[12] E. R. Shchukin, A. N. Kabanov, and E. Yankovska, J. Aerosol Sci. **21**, 397 (1990).
[13] N. A. Yavorskiĭ, A. N. Terebnin, and A. P. Bykov, *Catching Aerosols in the Tin Industry* [in Russian], Novosibirsk (1974), 86 pp.
[14] A. G. Amelin, *Theory of Fog Condensation*, 2nd ed. [Israel Program for Scientific Translations, Jerusalem (1967); Khimiya, Moscow (1986), 294 pp.].
[15] N. A. Fuchs, *The Mechanics of Aerosols* [Pergamon Press, Oxford–New York (1964); Moscow (1955), 352 pp.].
[16] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. [Pergamon Press, Oxford (1987); Nauka, Moscow (1986), 733 pp.].
[17] Yu. I. Yalamov, A. A. Yushkanov, and S. A. Savkov, ''Diffusiophoresis of moderately large nonvolatile aerosol particles,'' Dokl. Akad. Nauk SSSR **301**, 1111 (1988) [Sov. Phys. Dokl. **33**, 615 (1988)].

Translated by P. Shelnitz

Influence of charge relaxation on the capillary oscillations of a charged viscous spheroidal drop

S. O. Shiryaeva

Yaroslavl State University, 150000 Yaroslavl, Russia
(Submitted February 9, 1998)
Zh. Tekh. Fiz. **69**, 28–36 (August 1999)

A dispersion relation is derived for the spectrum of capillary modes of a charged spheroidal drop of a viscous liquid with allowance for charge relaxation. It is shown that the finite charge transport rate leads to lowering of the instability growth rates for various capillary modes of a spheroidal drop of a low-viscosity liquid. As the degree of deformation of the drop increases, the magnitude of the absolute change in the growth rate caused by the finite rate of charge redistribution decreases. © *1999 American Institute of Physics.*
[S1063-7842(99)00608-X]

The problem of the capillary oscillations and stability of a charged viscous drop having the shape of a prolate spheroid with consideration of the relaxation of the electric charge is of interest in connection with investigations of the mechanism underlying the instability of a highly charged spherical drop with respect to its self-charge. The same physical object, viz., a charged drop, is encountered in a very long list of problems in technical physics, geophysics, scientific-instrument design, and chemical technology. For this reason, the investigation of the capillary oscillations and stability of a charged drop having the shape of a prolate spheroid has been the subject of a considerable number of studies.[1–8] Nevertheless, the question of the influence of the finite rate of charge redistribution in the liquid on the laws governing the development of the instability of a spheroidal drop has not heretofore been explored, although such an influence can be significant.[9–11]

1. We shall solve the problem of the axisymmetric capillary oscillations of a charged prolate spheroidal drop of a viscous liquid of finite conductivity in a vacuum in a spherical coordinate system with its origin at the center of the drop, assuming that the spheroidal shape of the drop is imposed by the action of some outside forces of nonelectrical nature. We presume that as the drop oscillates, redistribution of the electric charge occurs only on its surface, while the bulk charge density equals zero. In addition, we assume that the drop material and the external medium are characterized by the constant dielectric constants ε_1 and ε_2.

The solution is found in dimensionless variables, in which the radius of the original spherical drop R, the density of the liquid in the drop ρ, and the surface tension α are equal to unity: $R=1$, $\rho=1$, and $\alpha=1$.

The equation of the surface of a prolate spheroid perturbed by capillary wave motion in spherical coordinates in the linear approximation with respect to e^2, i.e., the square of the eccentricity of the spheroid, has the form

$$r=r(\Theta)+\xi(\Theta,t)\approx 1+e^2h(\Theta)+\xi(\Theta,t),$$

where

$$r(\Theta)=(1-e^2)^{1/6}(1-e^2\cos^2\Theta)^{-1/2},$$

$$h(\Theta)=\frac{1}{6}(3\cos^2\Theta-1),$$

e is the eccentricity of the spheroid, and $\xi(\Theta,t)$ is the axisymmetric perturbation of the equilibrium spheroidal drop surface caused by capillary oscillations, which occur as a result of the thermal motion of the molecules and have an amplitude $\sim\sqrt{kT/\alpha}$ (here k is Boltzmann's constant, and T is the absolute temperature).

We also note that for most liquids the amplitude of such thermal capillary oscillations is of the order of tenths of a nanometer.

The ensuing analysis is carried out within perturbation theory by expansion in the small parameters e^2 and ξ to within terms $\sim\xi$ and $e^2\xi$, i.e., in the linear approximation with respect to ξ. We note that the small parameters e^2 and ξ are independent, and it is assumed that $e^2\ll\xi$. It would thus appear that if we retain the terms $\sim e^2\xi$, we should also take into account the terms $\sim e^4$. However, as will be seen below, only the terms $\sim\xi$ and $e^2\xi$ make contributions to the dispersion relation sought, and the terms $\sim e^2$ and e^4 vanish when the kinematic boundary condition (which contains a partial derivative with respect to time) is taken into account. For this reason, retention of the terms $\sim e^4$ in the calculations would lead only to an unjustified complication of the mathematical expressions describing the arguments made.

We assume that the dependences of the velocity field $\mathbf{U}(\mathbf{r},t)$, the pressure field $p_1(\mathbf{U},t)$ within the liquid, and the perturbation of the free surface $\xi(\Theta,t)$ on the time t are exponential, i.e., $\sim\exp(st)$, where s is the complex frequency.

We write out the system of hydrodynamic equations describing the motion of the viscous liquid in a drop that is caused by a small perturbation of the shape of its equilibrium surface $\xi(\Theta,t)$ and can therefore be characterized by the velocity field $\mathbf{U}(\mathbf{r},t)$, which is of the same order of smallness as ξ. The system consists of the Navier–Stokes equation and the incompressibility condition for a liquid,[12]

$$\frac{d\mathbf{U}}{dt}\equiv\frac{\partial\mathbf{U}}{\partial t}+(\mathbf{U}\cdot\boldsymbol{\nabla})\mathbf{U}=-\nabla p_1+\nu\,\Delta\mathbf{U};\quad \boldsymbol{\nabla}\cdot\mathbf{U}=0, \tag{1}$$

where ν is the kinematic viscosity coefficient and Δ is the Laplacian operator.

The system of equations describing the electrostatics of the problem being solved has the form[13]

$$\Delta\Phi_j=0,\quad j=1,2, \tag{2}$$

where the index 1 refers to the liquid, the index 2 refers to the external medium, and Φ_j are the electric potentials.

On the perturbed, weakly spheroidal drop surface

$$F(\mathbf{r},t)\equiv r-[1+e^2\,h(\Theta)+\xi(\Theta,t)]=0 \tag{3}$$

the following boundary conditions should be satisfied:[10–13]

$$\frac{\partial F(\mathbf{r},t)}{\partial t}+\mathbf{U}\cdot\boldsymbol{\nabla}F(\mathbf{r},t)=0; \tag{4}$$

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

$$\Pi_\tau^{(j)}=\frac{\varepsilon_j}{4\pi}E_{jn}E_{j\tau}; \tag{5}$$

$$-(p_1-p_2)+2\nu\,\mathbf{n}(\mathbf{n}\cdot\boldsymbol{\nabla})\mathbf{U}-p_E+p_\alpha=0; \tag{6}$$

$$\frac{\partial\mu}{\partial t}-\sigma\,\mathbf{n}\cdot\mathbf{E}_1+\mathrm{div}_\Sigma(\mu\mathbf{U}_\tau+\mu b\mathbf{E}_{1\tau})=0;\quad \Phi_1=\Phi_2; \tag{7}$$

$$r\to\infty:\quad \Phi_2\to0;\quad r\to0:\quad \Phi_1\to\text{const}. \tag{8}$$

Here

$$\mu(\Theta,t)=\frac{1}{4\pi}(\varepsilon_2E_{2n}-\varepsilon_1E_{1n})$$

is the surface density of the electric charge, $\mathbf{E}_j=-\boldsymbol{\nabla}\Phi_j$, $\mathbf{U}_\tau$ is the component of the velocity vector tangent to the surface (3), $\mathrm{div}_\Sigma\mathbf{A}$ is the surface divergence, σ is the electrical conductivity of the liquid drop, b is the mobility of the charge carriers in the liquid, E_n and E_τ are the normal and tangential components of the electric field intensity vector, $\mathbf{n}$ and $\boldsymbol{\tau}$ are unit vectors of a normal and a tangent to the free surface, p_2 is the pressure of the external medium, and p_E and p_α are the pressures created by the electric forces and the forces of surface tension. We also require satisfaction of the following conditions: constancy of the drop volume

$$\int_V dV=\frac{4}{3}\pi, \tag{9}$$

immobility of its center of mass

$$\int_V\mathbf{r}\,dV=0, \tag{10}$$

and conservation of the total charge Q of the drop during oscillations of its surface

$$\int_S\mu(\Theta,t)\,dS=Q. \tag{11}$$

In (9) and (10) the integration is carried out over the entire drop volume, and in (11) it is carried out over the surface of the drop.

2. In the linear approximation with respect to $\mathbf{U}$ and ξ Eqs. (1) take the form

$$\frac{\partial\mathbf{U}}{\partial t}=-\nabla(p_1(\mathbf{U},t))+\nu\,\Delta\mathbf{U};\quad \boldsymbol{\nabla}\cdot\mathbf{U}=0, \tag{12}$$

where $p_1(\mathbf{U},t)$ is the additional pressure within the liquid, which is of first order with respect to U (i.e., with respect to ξ).

When the problem is solved in the first-order approximation with respect to U and ξ, it is sufficient to take the linearized boundary conditions (4)–(7) on the unperturbed surface of the spheroidal drop $r=r(\Theta)\approx1+e^2h(\Theta)$.

We rewrite the conditions (4)–(6) in terms of the projections of the velocity vector $\mathbf{U}(\mathbf{r},t)$ onto the unit vectors of the spherical coordinate system (U_r, U_Θ, and U_φ). As a result, the kinematic boundary condition (4) takes the form

$$r=r(\Theta):\quad -\frac{\partial\xi(\Theta,t)}{\partial t}+U_r+U_\Theta e^2\sin\Theta\cos\Theta=0. \tag{13}$$

The dynamic boundary condition for the tangential component of the stress tensor (5) can be separated into two conditions, since two mutually perpendicular unit vectors, viz., $\boldsymbol{\tau}_\varphi$ and $\boldsymbol{\tau}_\Theta$, can be chosen at a given point on the plane tangent to the drop surface. As a result, from (5) we obtain the following conditions: 1)

$$r=r(\Theta):\quad (\pi_{\tau_\varphi}^{(2)}-\Pi_{\tau_\varphi}^{(1)})-\nu\left\{\frac{\partial U_\varphi}{\partial r}-\frac{1}{r}U_\varphi+e^2\sin\Theta\cos\Theta\left[\frac{1}{r}\frac{\partial U_\varphi}{\partial\Theta}-\frac{1}{r}\frac{\cos\Theta}{\sin\Theta}U_\varphi\right]\right\}=0, \tag{14}$$

when $\boldsymbol{\tau}_\varphi$ is chosen as the unit vector of the tangent; 2)

$$r=r(\Theta):\quad (\pi_{\tau_\Theta}^{(2)}-\Pi_{\tau_\Theta}^{(1)})-\nu\left\{\frac{\partial U_\Theta}{\partial r}+\frac{1}{r}\frac{\partial U_r}{\partial\Theta}-\frac{1}{r}U_\Theta+e^2 2\sin\Theta\cos\Theta\left[\frac{1}{r}U_r+\frac{1}{r}\frac{\partial U_\Theta}{\partial\Theta}-\frac{\partial U_r}{\partial r}\right]\right\}=0, \tag{15}$$

when $\boldsymbol{\tau}_\Theta$ is chosen as the unit vector of the tangent.

The boundary condition (16) for the normal component of the stress tensor takes the form

$$r=r(\Theta):\quad \left\{-p_1(\mathbf{U},t)+2\nu\times\left[\frac{\partial U_r}{\partial r}+e^2\sin\Theta\cos\Theta\left(\frac{\partial U_\Theta}{\partial r}+\frac{1}{r}\frac{\partial U_r}{\partial\Theta}\right)\right]-p_E(\xi)+p_\alpha(\xi)\right\}=0, \tag{16}$$

where $p_E(\xi)$ and $p_\alpha(\xi)$ are additions to the corresponding pressures, which are of first order with respect to ξ.

Since the integration in (9) and (10) is carried out over the entire drop volume, in the linear approximation with respect to ξ these conditions can be written in the form

$$\int_0^{\pi} \xi(\Theta,t)\sin\Theta\, d\Theta = 0; \tag{17}$$

$$\int_0^{2\pi}\int_0^{\pi} \xi(\Theta,t)\mathbf{e}_r \sin\Theta\, d\Theta\, d\varphi = 0. \tag{18}$$

Thus, the system of vector hydrodynamic equations (12) with the boundary conditions (13)–(16) and the additional conditions (17) and (18), as well as the system of equations (2) for the electric potentials with the boundary conditions (7) and (9) and the additional condition (11) comprise a complete mathematical formulation of the problem posed.

3. We shall solve the stated problem by the scalarization method described in detail in Ref. 14, generalizing it in the linear approximation with respect to the square of the eccentricity to the case of the simplest spheroidal region for a spherical coordinate system.[6] For this purpose we represent the velocity field $\mathbf{U}(\mathbf{r},t)$ in the form of an expansion

$$\mathbf{U}(\mathbf{r},t) = \hat{\mathbf{N}}_1\Phi_1(\mathbf{r},t) + \hat{\mathbf{N}}_2\Phi_2(\mathbf{r},t) + \hat{\mathbf{N}}_3\Phi_3(\mathbf{r},t), \tag{19}$$

where $\Phi_k(\mathbf{r},t)$ are scalar functions, and $\hat{\mathbf{N}}_k$ are differential vector operators, which take the following form in a spherical coordinate system

$$\hat{\mathbf{N}}_1 \equiv \nabla; \quad \hat{\mathbf{N}}_2 \equiv \nabla\cdot\mathbf{r}; \quad \hat{\mathbf{N}}_3 \equiv \nabla(\nabla\cdot\mathbf{r}).$$

Taking into account the expansion (19), we easily transform the system of vector equations (12) for $\mathbf{U}(\mathbf{r},t)$ and $p_1(\mathbf{U},t)$ into a system of scalar equations for the scalar functions $\Phi_k(\mathbf{r},t)$

$$\Delta\Phi_k(\mathbf{r},t) - \frac{s}{\nu}(1-\delta_{k1})\Phi_k(r,t) = 0 \quad (k=1,2,3);$$

$$p_1(\mathbf{U},t) = -s\Phi_1(\mathbf{r},t), \tag{20}$$

where δ_{kj} is the Kronecker delta.

We also rewrite the boundary conditions (13)–(16) in terms of the scalar functions $\Phi_k(\mathbf{r},t)$: the kinematic boundary condition is

$$r = r(\Theta): \quad s\xi - \left[\frac{\partial\Phi_1}{\partial r} - \frac{1}{r}\Delta_\Omega\Phi_3\right] - e^2\sin\Theta\cos\Theta\frac{\partial}{\partial\Theta}\left[\frac{\Phi_1}{r} + \frac{1}{r}\frac{\partial}{\partial r}(r\Phi_3)\right] = 0; \tag{21}$$

the dynamic boundary conditions for the tangential components of the stress tensor are

$$\frac{1}{\sin\Theta}\frac{\partial}{\partial\Theta}\sin\Theta(\Pi^{(2)}_{\tau_\varphi} - \Pi^{(1)}_{\tau_\varphi}) + \nu\left\{\Delta_\Omega\left[\frac{\partial\Phi_2}{\partial r} - \Phi_2\right] + e^2\left[h(\Theta)\Delta_\Omega\left(\frac{\partial^2\Phi_2}{\partial r^2} - \frac{\partial\Phi_2}{\partial r} + 3\Phi_2\right) - \frac{2}{3}\Delta_\Omega(\Phi_2) - \sin\Theta\cos\Theta\frac{\partial}{\partial\Theta}\left(\frac{\partial^2\Phi_2}{\partial r^2} - \frac{\partial\Phi_2}{\partial r} - (3+\Delta_\Omega)\Phi_2\right)\right]\right\} = 0; \tag{22}$$

$$\frac{1}{\sin\Theta}\frac{\partial}{\partial\Theta}\sin\Theta(\Pi^{(2)}_{\tau_\Theta} - \Pi^{(1)}_{\tau_\Theta}) + \nu\left\{\Delta_\Omega\left[2\left(\frac{\partial\Phi_1}{\partial r} - \Phi_1\right) + \frac{\partial^2\Phi_3}{\partial r^2} - (2+\Delta_\Omega)\Phi_3\right] - e^2\left(h(\Theta)\left[2(6-\Delta_\Omega)\times\frac{\partial^2\Phi_1}{\partial r^2} - 4(3-\Delta_\Omega)\frac{\partial\Phi_1}{\partial r} - 16\Delta_\Omega\Phi_1 - \Delta_\Omega\left(\frac{\partial^3\Phi_3}{\partial r^3} + (22-\Delta_\Omega)\frac{\partial\Phi_3}{\partial r} - 2(4-\Delta_\Omega)\Phi_3\right)\right] + \sin\Theta\cos\Theta\frac{\partial}{\partial\Theta}\left[2\left(2\frac{\partial^2\Phi_1}{\partial r^2} - 3\frac{\partial\Phi_1}{\partial r} - \Delta_\Omega\Phi_1\right) + \frac{\partial^3\Phi_3}{\partial r^3} - (6+5\Delta_\Omega)\frac{\partial\Phi_3}{\partial r} + 4\Delta_\Omega\Phi_3\right] + 2\cos^2\Theta\Delta_\Omega\left[\Phi_1 + \frac{\partial\Phi_3}{\partial r} + \Phi_3\right]\right)\right\} = 0; \tag{23}$$

the dynamic boundary condition for the normal component of the stress tensor is

$$r = r(\Theta): \quad s\Phi_1 + 2\nu\left\{\left[\frac{\partial^2\Phi_1}{\partial r^2} - \Delta_\Omega\left(\frac{\partial}{\partial r}\left(\frac{\Phi_3}{r}\right)\right)\right] + e^2\sin\Theta\cos\Theta\frac{\partial}{\partial\Theta}\left[\frac{2}{r}\frac{\partial\Phi_1}{\partial r} - \frac{\Phi_1}{r^2} + \frac{\partial^2\Phi_3}{\partial r^2} + \frac{\partial}{\partial r}\left(\frac{\Phi_3}{r}\right) - \Delta_\Omega\left(\frac{\Phi_3}{r^2}\right)\right]\right\} - p_E(\xi) + p_\alpha(\xi) = 0;$$

$$\Delta_\Omega \equiv \frac{1}{\sin\theta}\frac{\partial}{\partial\Theta}\left(\sin\Theta\frac{\partial}{\partial\Theta}\right). \tag{24}$$

It is clear from the form of Eqs. (20) that in a spherical coordinate system the dependence of the functions $\Phi_k(\mathbf{r},t)$ on the angle Θ is specified by the spherical harmonics $Y^0_m(\Theta)$. Therefore, the solutions of the system (20) which are regular at $r=0$ are written in the form

$$\Phi_1(\mathbf{r},t) = \sum_{m=0}^{\infty} C^1_m r^m Y^0_m(\Theta)\exp(st);$$

$$\Phi_{2,3}(\mathbf{r},t) = \sum_{m=0}^{\infty} C^{2,3}_m i_m\left(\sqrt{\frac{s}{\nu}}r\right)Y^0_m(\Theta)\exp(st), \tag{25}$$

where $i_m(x)$ are modified spherical Bessel functions, and C_m^k are constants.

Since the function $\xi(\Theta,t)$ is related to the functions $\Phi_k(\mathbf{r},t)$ by the kinematic boundary condition, it is logical to represent it in the form of the following expansion in the spherical harmonics:

$$\xi(\Theta,t)=\sum_{m=0}^{\infty} Z_m Y_m^0(\Theta)\exp(st), \tag{26}$$

where C_m^1, $C_m^{2,3}$, and Z_m are constants.

The additional conditions (17) and (18) permit refinement of the range of variation of m in the expansions (25) and (26). Projecting (18) onto the axes of the Cartesian coordinate system, from (17) and (18) we can easily find that $Z_0=Z_1=0$. Therefore, in the expansions (25) and (26) the smallest value of m should be $m=2$.

4. The relationship of the unknown constants C_m^k $(k=1,2,3)$ in the expansions (25) to the expansion coefficients Z_m of the perturbation function $\xi(\Theta,t)$ is determined from the boundary conditions (21)–(24), as are the expression for the complex frequency of the capillary oscillations s.

We first of all note that in the case of axisymmetric oscillations the boundary condition (22) for $\Phi_2(\mathbf{r},t)$ is completely autonomous, i.e., it does not contain some other unknown function, i.e., $\Phi_2(\mathbf{r},t)$ does not influence the oscillations of the drop surface and, therefore, will not be considered below. We shall focus our attention on the boundary conditions (21), (23), and (24) for $\Phi_1(\mathbf{r},t)$, $\Phi_3(\mathbf{r},t)$, and $\xi(\Theta,t)$.

We substitute the solutions (25) and (26) into the condition (21), utilizing the recurrence relations for $i_m(x)$ and $Y_m^0(\Theta)$ (Ref. 15) and the orthonormality of the spherical harmonics and neglecting the interaction of the capillary modes. In the linear approximation with respect to e^2 we obtain an equation of the form

$$C_m^1[m+e^2\varkappa_m\gamma_1]+D_m^3\Bigg(m(m+1)+e^2\varkappa_m \times\Bigg[(m+1)\gamma_1+\sqrt{\frac{s}{\nu}}\,f_m\Bigg(\sqrt{\frac{s}{\nu}}\Bigg)\gamma_2\Bigg]\Bigg)-sZ_m=0. \tag{27}$$

Before transforming the boundary conditions (23) and (24), we must consider the electrical boundary-value problem (2), (7), (8), and (11). The solution of this problem was sought by expanding the potentials Φ_j, the boundary conditions (7) and (8), and the additional condition (11) in small quantities $\sim e^2$, ξ, and $e^2\xi$. The condition for redistribution of the charge over the drop surface (7) was explored using the solutions for the hydrodynamic potentials $\Phi_j(\mathbf{r},t)$ $(j=1,2,3)$ in the form (25). As a result of some very cumbersome calculations we were able to obtain a solution for the electric potentials and, using them, to write out the electrical part of the tangential component of the stress tensor

$$\begin{aligned}\frac{1}{\sin\Theta}\frac{\partial}{\partial\Theta}\sin\Theta(\Pi^{(2)}_{\tau_\Theta}-\Pi^{(1)}_{\tau_\Theta})&\approx 4W\frac{m}{L_m}\Bigg\{\varepsilon_2(m+1)[m(m+1)+e^2\varkappa_m g_m^2(\lambda)]C_m^1\\&+\varepsilon_2(m+1)\Bigg[m(m+1)^2+e^2\varkappa_m\Bigg(\gamma_2\frac{s}{\nu}+g_m^3(\lambda)\Bigg)\\&+\sqrt{\frac{s}{\nu}}\,f_m\Bigg(\sqrt{\frac{s}{\nu}}\Bigg)m(m+1)(1+e^2\varkappa_m\zeta_m(\lambda))\Bigg]D_m^3\\&-[(m^2-1)\lambda_2+e^2\varkappa_m(m+1)g_m^1(\lambda)]Z_m\Bigg\}\\&\times Y_m^0(\Theta)\exp(st)\end{aligned} \tag{28}$$

and the expression for the addition to the pressure created by the electric forces

$$\begin{aligned}p_E(\xi)|_{r=r(\Theta)}\approx\frac{Q^2}{4\pi\varepsilon_2}\sum_{m=2}^{\infty}\frac{1}{L_m}\Bigg\{&[m(m-1)(\lambda_1+(m+1)\\&\times 4b\sqrt{\pi\varepsilon_2 W})+e^2\varkappa_m((m-4)L_m-3\\&\times(m-1)\lambda_2-(m+1)g_m^{(1)}(\lambda))]Z_m+\varepsilon^2\\&\times(m+1)[m(m+1)+e^2\varkappa_m(3m+g_m^{(2)}(\lambda))]\\&\times C_m^1+\varepsilon_2(m+1)\Bigg[m(m+1)\Bigg((m+1)\\&+\sqrt{\frac{s}{\nu}}\,f_m\Bigg(\sqrt{\frac{s}{\nu}}\Bigg)\Bigg)+e^2\varkappa_m\Bigg(\gamma_2\frac{s}{\nu}+3m\\&\times(m+1)+g_m^{(3)}(\lambda)+\sqrt{\frac{s}{\nu}}\,f_m\Bigg(\sqrt{\frac{s}{\nu}}\Bigg)\\&\times(3m+m(m+1)\zeta_m(\lambda))\Bigg)\Bigg]D_m^3\Bigg\}\\&\times Y_m^0(\Theta)\exp(st).\end{aligned} \tag{29}$$

Using the expression (28) and substituting the solutions (25) and (26) into the dynamic boundary condition for the tangential component of the stress tensor (23), we obtain

$$\begin{aligned}C_m^1\Bigg\{&2(m-1)\nu+4W\varepsilon_2\frac{m(m+1)}{L_m}+e^2\varkappa_m\Bigg[\beta_1\nu\\&+4W\varepsilon_2\frac{1}{L_m}g_m^{(2)}(\lambda)\Bigg]\Bigg\}+D_m^3\Bigg\{\Bigg[s+2(m^2-1)\nu\\&+4W\varepsilon_2\frac{m(m+1)^2}{L_m}\Bigg]-\sqrt{\frac{s}{\nu}}\,f_m\Bigg(\sqrt{\frac{s}{\nu}}\Bigg)\\&\times\Bigg[2\nu-4W\varepsilon_2\frac{m(m+1)}{L_m}\Bigg]+e^2\varkappa_m\Bigg[\beta_2 s+\beta_3\nu\\&+4W\varepsilon_2\frac{1}{L_m}\Bigg(\gamma_2\frac{s}{\nu}+g_m^{(3)}(\lambda)\Bigg)+\sqrt{\frac{s}{\nu}}\,f_m\Bigg(\sqrt{\frac{s}{\nu}}\Bigg)\end{aligned}$$

$$\times\left[\beta_4 s+\beta_3\nu+4W\varepsilon_2\frac{1}{L_m}m(m+1)\zeta_m(\lambda)\right]\Bigg]\Bigg\}$$

$$-\frac{4W}{L_m}\{(m-1)\lambda_2+e^2\varkappa_m g_m^{(1)}(\lambda)\}Z_m=0. \tag{30}$$

In order to write out the dynamic boundary condition for the normal component of the stress tensor, we must utilize the expressions for the addition to the pressure created by the electric forces $p_E(\xi)$ (29) and the addition to the pressure created by the surface tension forces $p_\alpha(\xi)$, which, in the axisymmetric case, has the form[7]

$$p_\alpha(\xi)\approx\sum_{m=2}^{\infty}[(m-1)(m+2)+e^2 2(m^2+m+4)\varkappa_m]$$
$$\times Z_m Y_m^0(\Theta)\exp(st). \tag{31}$$

Substituting (25), (29), and (31) into (24), we obtain

$$C_m^4\Bigg\{\left[s+\gamma_3\nu-4W\varepsilon_2\frac{m(m+1)^2}{L_m}\right]$$
$$+e^2\varkappa_m\left[ms+\gamma_4\nu-4W\varepsilon_2\frac{(m+1)}{L_m}(3m+g_m^{(2)}(\lambda))\right]\Bigg\}$$
$$+D_m^3\Bigg\{m(m+1)\left[2(m-1)\nu-4W\varepsilon_2\frac{(m+1)^2}{L_m}\right]$$
$$+\sqrt{\frac{s}{\nu}}f_m\left(\sqrt{\frac{s}{\nu}}\right)m(m+1)\left[2\nu-4W\varepsilon_2\frac{(m+1)}{L_m}\right]$$
$$+e^2\varkappa_m\Bigg[2s\gamma_2+(m+1)\gamma_4\nu-4W\varepsilon_2\frac{(m+1)}{L_m}$$
$$\times\left(\gamma_2\frac{s}{\nu}+3m(m+1)+g_m^{(3)}(\lambda)\right)-\sqrt{\frac{s}{\nu}}f_m\left(\sqrt{\frac{s}{\nu}}\right)$$
$$\times\left(2\gamma_5\nu+4W\varepsilon_2\frac{m(m+1)}{L_m}(3+(m+1)\zeta_m(\lambda))\right)\Bigg]\Bigg\}$$
$$+Z_m\Bigg\{(m-1)(m+2)\Bigg[1-\frac{4W}{(m+2)}\frac{m}{L_m}$$
$$\times(\lambda_1+(m+1)4b\sqrt{\pi\varepsilon_2 W})\Bigg]-e^2\varkappa_m\Bigg[\gamma_6+4W(m-4)$$
$$-4W\frac{1}{L_m}(3(m-1)\lambda_2+(m+1)g_m^{(1)}(\lambda))\Bigg]\Bigg\}=0. \tag{32}$$

The following notation was used in (27)–(32):

$$f_m\left(\sqrt{\frac{s}{\nu}}\right)\equiv\frac{i_{m+1}\left(\sqrt{\frac{s}{\nu}}\right)}{i_m\left(\sqrt{\frac{s}{\nu}}\right)};\quad D_m^3\equiv C_m^3 i_m\left(\sqrt{\frac{s}{\nu}}\right);$$

$$\lambda_1\equiv s\varepsilon_1+4\pi\sigma;\quad \lambda_2\equiv s\varepsilon_2;\quad W\equiv\frac{Q^2}{16\pi\varepsilon_2};$$

$$L_m\equiv m\lambda_1+(m+1)\lambda_2+m(m+1)4\sqrt{\pi\varepsilon_2 W}b;$$

$$\varkappa_m\equiv\frac{m(m+1)}{3(2m-1)(2m+3)};$$

$$\beta_1=2\left(m^2-3m-1+\frac{3}{m}\right);$$

$$\beta_2=(m-2)\left(1-\frac{3}{m(m+1)}\right);$$

$$\beta_3=2\left(m^3-2m^2-4m+2+\frac{3}{m}\right);$$

$$\beta_4=\left(1-\frac{3}{m(m+1)}\right);$$

$$\beta_5=2(m^2+m-1)+\frac{6}{m(m+1)};$$

$$\gamma_1=m(m-1)-3;\quad \gamma_2=m(m+1)-3;\quad \gamma_3=2m(m-1);$$
$$\gamma_4=2(m^3-3m^2-4m+3);$$
$$\gamma_5=4m(m+1)-3;\quad \gamma_6=2(m(m+1)+4);$$
$$g_m^{(1)}(\lambda)=\lambda_2[(m-4)+(m-1)\,\zeta_m(\lambda)];$$
$$g_m^{(2)}(\lambda)=m(m+1)[(m+2)+\zeta_m(\lambda)]-6;$$
$$g_m^{(3)}(\lambda)=m(m+1)^2[m+\zeta_m(\lambda)];$$
$$\zeta_m(\lambda)=\frac{3}{L_m}(\lambda_1-\lambda_2).$$

Setting the determinant of the system of equations (27), (30), and (32) equal to zero, we obtain the dispersion relation of the problem in the linear approximation with respect to e^2:

$$s\{s^2+2(m-1)(2m+1)\nu s+m(m-1)(m+2)\alpha_m\}$$
$$-\sqrt{\frac{s}{\nu}}f_m\left(\sqrt{\frac{s}{\nu}}\right)\Bigg\{\left(2\nu-m(m+1)\frac{4W}{L_m}\right)$$
$$\times(s^2+2m(m-1)(m+2)\nu s+m(m-1)(m+2)\alpha_m)$$
$$+\frac{4W}{L_m}2\nu m(m^4-1)s\Bigg\}+e^2\varkappa_m\Bigg\{s\Bigg([s^2+m(m-1)$$
$$\times(m+2)\alpha_m]\left[2(m-1)-\frac{3(m-2)}{m(m+1)}+\left(1-\frac{3}{m(m+1)}\right)\right.$$
$$\left.\times\sqrt{\frac{s}{\nu}}f_m\left(\sqrt{\frac{s}{\nu}}\right)\right]-3[(2m-1)(m+2)\alpha_m+m^3]\Bigg)$$
$$+\frac{4W}{L_m}\Bigg\{\frac{s}{\nu}(m(m+1)-3)[s^2+m(m-1)(m+2)\alpha_m]$$
$$+2(m(m+1)-3)[2m(m-2)s-m(m-1)(m+1)$$
$$\times(m+2)\alpha_m]-\sqrt{\frac{s}{\nu}}f_m\left(\sqrt{\frac{s}{\nu}}\right)(s^2(m^2-2m-3)+4m^2$$
$$\times(m+1)(m^2+m+1)-4Wm^2(m+1)(m+2)$$
$$-\zeta_m(\lambda)m(m+1)(m(m-1)(m+2)\alpha_m+s^2))\Bigg\}$$

$$+2\nu\left[s^2\left(2m^3-8m^2+m-1+\frac{3}{m}-\frac{3(m-2)(m-1)}{(m+1)}\right)\right.$$
$$+\sqrt{\frac{s}{\nu}}\,f_m\left(\sqrt{\frac{s}{\nu}}\right)\left[s^2\left(2m^2-m-4+\frac{3(2m+1)}{m(m+1)}\right)\right.$$
$$\left.\left.+3\left(m^2+\frac{(m-1)}{(m+1)}\right)(m+2)\alpha_m+3m^3\right]\right]-2\nu s\frac{4W}{L_m}$$
$$\times\left\{2(m+1)[m(m+1)-3][2m^2-m-1]\right.$$
$$+\sqrt{\frac{s}{\nu}}\,f_m\left(\sqrt{\frac{s}{\nu}}\right)(m(m+1)(m-1)(8m+5)$$
$$\left.-9(m-1)-\zeta_m(\lambda)m(m+1)(m-1)(2m-1))\right\}$$
$$\left.+4\nu^2 s\sqrt{\frac{s}{\nu}}\,f_m\left(\sqrt{\frac{s}{\nu}}\right)\left[5m^3+5m^2-4m-\frac{6}{(m+1)}\right]\right\}=0. \tag{33}$$

Here

$$\alpha_m\equiv 1-\frac{4}{(m+2)}W.$$

Setting $e^2=0$ in (33), i.e., discarding the entire corresponding expression in the large curly brackets, we obtain the dispersion relation for the oscillation frequencies of a charged spherical drop of a viscous liquid of finite conductivity.[10]

In the ensuing numerical analysis of Eq. (33) we shall take $\varepsilon_2=1$ and set $\varepsilon_1\equiv\varepsilon$.

The qualitative form of the motions of the liquid (branches of the dispersion relation) obtained by a numerical calculation using (33) for nonzero eccentricity differ only slightly from those obtained for a spherical drop[10] and are illustrated in Fig. 1, where the general form of the branches of the dispersion relation (33) for the fundamental model of a drop of a low-conductivity liquid ($m=2$, $e^2=0.1$, $\nu=0.03$, $\sigma=0.1$, $\varepsilon=10$, $b=0.1$) is presented in the form of plots of $\mathrm{Re}\,s(W)$ and $\mathrm{Im}\,s(W)$. Curves *1–3* specify the capillary motions of the liquid in the drop, and curves *1*—R, *2*—R, and *3*—R characterize the damped relaxation–convection motions caused by the transport of charge due to equalization of the electric potential and the transport of charge by streams of liquid. Curve *1*—CP characterizes the damped capillary–poloidal periodic motion of the liquid appearing as a result of the interaction of branch *3* of the damped harmonic capillary motion with branch *4* of the damped poloidal vortex motion. It is noteworthy that the interaction of the harmonic capillary motion with the poloidal motion was manifested only when the charge relaxation effect was taken into account and that these two types of motion do not interact in an ideally conductive liquid.[16] Curves *4* and *5* characterize aperiodically damped poloidal motions [actually there is an endless set of poloidal vortex motions,[10,16] but only the ones whose damping decrements fall on the $(\mathrm{Re}\,s, W)$ plane in the scale adopted are shown in the figure]. The real part of curve *2* in the region $\mathrm{Re}\,s>0$ characterizes the growth rate of the instability of the drop with respect to its self-charge.

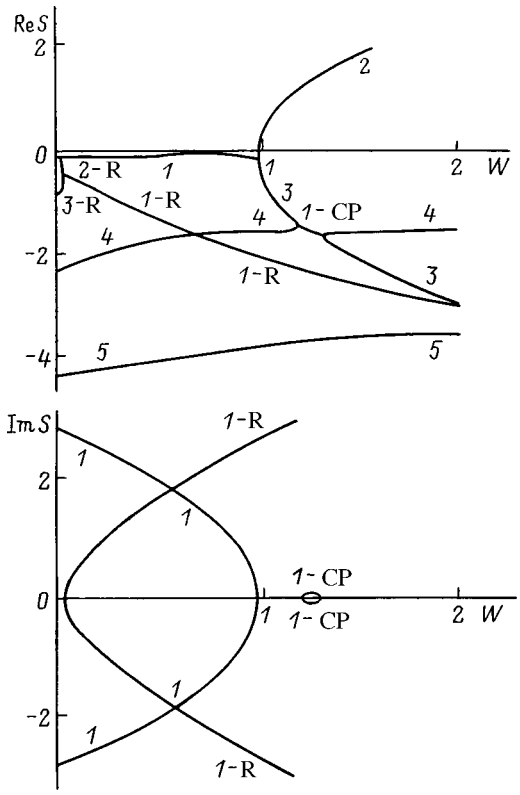

FIG. 1. Dependence of the real and imaginary components of the complex frequency on the dimensionless parameter W, which characterizes the stability of a drop with respect to its self-charge.

As the numerical calculations showed, consideration of the finite electrical conductivity of the liquid and finite mobility of the charge carriers leads to a decrease in the instability growth rate in comparison to an ideally conductive liquid, for which these parameters are infinitely large. In low-viscosity liquids with $\nu\ll 1$ the qualitative dependences of the instability growth rate on conductivity and carrier mobility for nonzero eccentricity are similar to the analogous dependences for a spherical drop.[10] Figure 2 presents plots of the instability growth rate of the tenth mode ($m=10$) as a function of the conductivity σ for a liquid such as dibutyl phthalate ($b=0.1$, $\varepsilon=10$, $\nu=0.1$, $W=4$) and various values of the square of the eccentricity (curve *1* was obtained for $e^2=0$, curve *2* for $e^2=0.1$, and curve *3* for $e^2=0.5$). It is seen that an increase in σ leads to an increase in the instability growth rate. It is also not difficult to see that the absolute change in the instability growth rate decreases with increasing values of the square of the eccentricity. Similar tendencies are also noted as the carrier mobility varies, as can be seen from Fig. 3 ($m=10$, $\sigma=0.1$, $\varepsilon=10$, $\nu=0.1$, $W=4$; curve *1* corresponds to $e^2=0$, and curve *2* corresponds to $e^2=0.5$): as the eccentricity increases, the absolute change in the instability growth rate decreases, i.e., the effect of the finite potential equalization rate, which is manifested as an increase in the time for the development of instability, abates as the degree of deformation rises. Such a phenom-

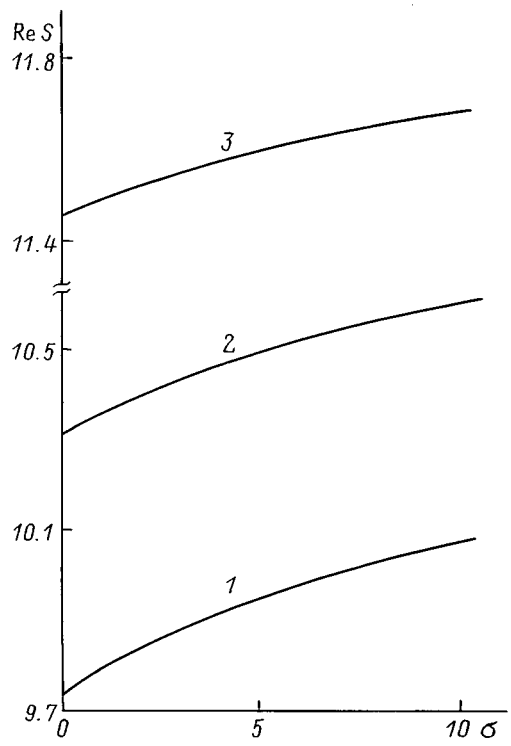

FIG. 2. Dependence of the instability growth rate of the tenth mode on the dimensionless electrical conductivity of the liquid.

enon can be explained on the qualitative level by a decrease in the mean surface charge density as the drop is deformed and by alteration of the characteristic distances over which charge must be transported to equalize the electric potential.

Numerical calculations for a high viscosity ($\nu=0.8$) also show that consideration of the finite electrical conductivity and finite carrier mobility for a deformed drop leads to dependences which are the opposites of those presented in Figs. 2 and 3: as σ and b increase, the instability growth rate

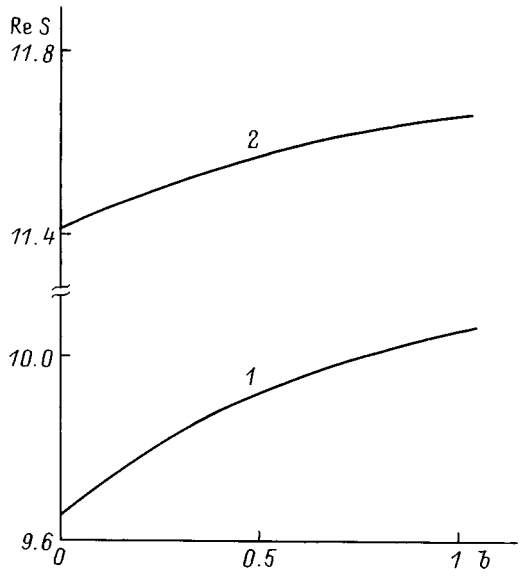

FIG. 3. Dependence of the instability growth rate of the tenth mode on the dimensionless carrier mobility in the liquid.

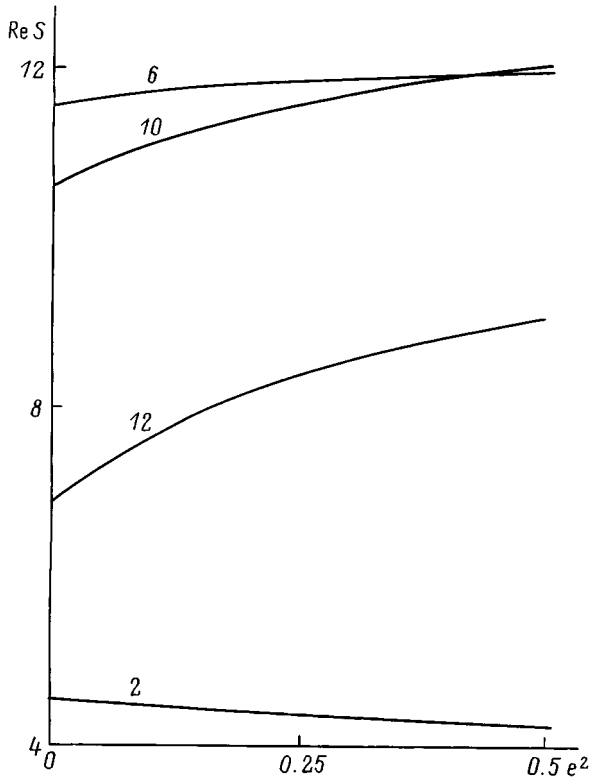

FIG. 4. Dependence of the instability growth rates of various modes on the square of the eccentricity.

of the same tenth mode decreases. This is probably because the mechanical stresses appearing upon redistribution of the charge over the surface of a high-conductivity liquid are characterized by very short relaxation times and are effectively quenched by the viscosity in a high-viscosity liquid without being manifested as an increase in the instability growth rate of the respective mode, but with additional dissipation of the energy of the capillary motions of the liquid.

When the eccentricity is nonzero, an increase in the dielectric constant of the liquid also leads to lowering of the instability growth rate of a spheroidal drop.

It would also be interesting to ascertain how the instability growth rates of the first few capillary modes of spheroidal drops in various liquids depend on eccentricity. Figure 4 shows plots of the instability growth rate versus of the square of the eccentricity for the second, sixth, tenth, and twelfth modes (the numbers of the curves in the figure correspond to the numbers of the modes) for a high-conductivity liquid such as seawater, which corresponds to the dimensionless values $\sigma=10^5$, $b=1$, $\varepsilon=80$, and $\nu=0.1$, with $W=4$ (the critical value of the Rayleigh parameter for realizing instability of a drop, i.e., for a loss of stability of the fundamental mode, is $W=1$). The plots in Fig. 5, which are similar to the curves in Fig. 4, were calculated for a low-conductivity liquid like dibutyl phthalate ($\sigma=0.1$, $b=0.1$, $\varepsilon=10$, and $\nu=0.1$ with $W=4$). It can be seen from the plots presented that the instability growth rate of the fundamental mode decreases as e^2 increases (generally speaking, this is consistent with Le Chatelier's principle). For higher modes an increase in e^2 leads to an increase in the instability growth

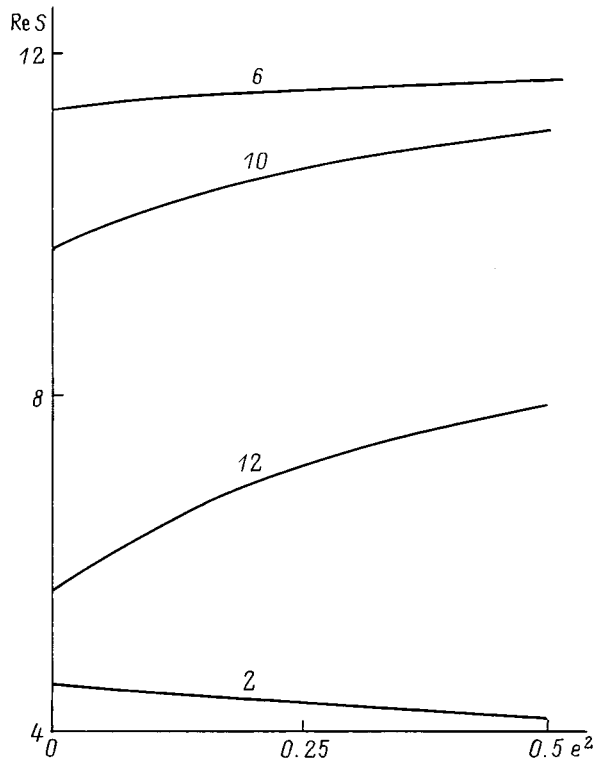

FIG. 5. Plots similar to the curves in Fig. 4, but for other values of σ, b, and ε.

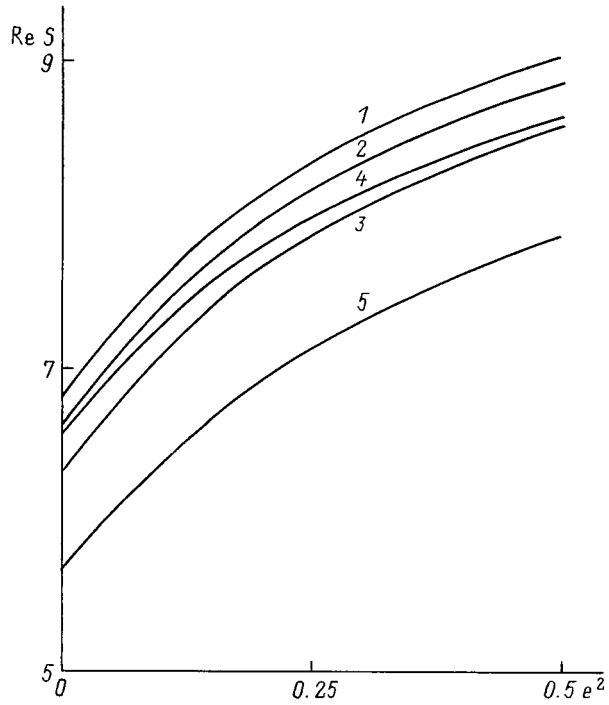

FIG. 6. Dependence of the instability growth rate of the twelfth mode on the square of the eccentricity for various liquids.

rates due to an increase in the surface charge density on the tips of the spheroid.[4] Numerical calculations performed for various small values of the dimensionless viscosity show that the influence of the finite charge transport rate on the instability growth rate of various modes is weaker for low-viscosity liquids than for high-viscosity liquids.

At high viscosity values the viscosity exhibits its strongest damping influence on rapid motions of the liquid (caused by the high instability growth rates of the higher modes). This causes the time for development of the instability of lower modes to become shorter than the time for development of the instability of higher modes. Consequently, unlike the breakup mechanism with the emission of hundreds of highly dispersed daughter droplets described in Ref. 4, the drop acquires a tendency to break up into several parts of comparable size.[17] At small eccentricity values in the range $e^2<0.1$ this tendency is more pronounced in a low-conductivity liquid than in a high-conductivity liquid. Conversely, when the eccentricity is high, the tendency for division of the drop into parts of comparable size is more pronounced in a high-conductivity liquid. This can be seen from Fig. 6, which presents the results of the calculations for the twelfth mode ($m=12$), $W=4$, and $\nu=0.1$ in the various liquids used in the experiments on the electrospraying of liquids in Ref. 18: curve *1* is for a liquid like seawater ($\sigma=10^5$, $b=1$, $\varepsilon=80$); curve *2* is for a liquid like glycerol ($\sigma=50$, $b=0.1$, $\varepsilon=50$); curve *3* is for a liquid like ethyl alcohol ($\sigma=10$, $b=0.1$, $\varepsilon=25$); curve *4* is for a liquid like distilled water ($\sigma=1$, $b=1$, $\varepsilon=80$); and curve *5* is for a liquid like dibutyl phthalate ($\sigma=0.1$, $b=0.1$, $\varepsilon=10$).

It is noteworthy that the range of variation of e^2 from 0 to 0.5 used to construct the plots in Figs. 2–6 was taken only for graphic purposes and for qualitative predictions of the possible behavior of the instability growth rates as the spheroidal deformation increases, since Eq. (33) was obtained for the range $e^2\ll 1$; therefore, the plots corresponding to large eccentricity values have a qualitative character.

CONCLUSION

Consideration of the finite equalization rate of the electric potential for a highly charged prolate spheroidal drop of a low-viscosity liquid leads to lowering of the instability growth rates. As the degree of deformation of the drop increases, the absolute change in the instability growth rate caused by the finite charge redistribution rate decreases. Drops of high-viscosity liquids display a different tendency. The mechanical stresses appearing as a result of redistribution of the charge over the surface of high-conductivity liquids are characterized by very short characteristic relaxation times and are effectively quenched in high-viscosity liquids, being manifested as a change in the instability growth rate of the respective mode to only a small extent.

The author thanks V. A. Koromyslov and M. I. Munichev for their assistance in the numerical calculations.

[1] A. I. Grigor'ev, Zh. Tekh. Fiz. **55**, 1272 (1985) [Sov. Phys. Tech. Phys. **30**, 736 (1985)].
[2] K. J. Cheng, Phys. Lett. A **112**, 392 (1985).
[3] A. I. Grigor'ev, S. O. Shiryaeva, and E. I. Belavina, Zh. Tekh. Fiz. **59**(6), 27 (1989) [Sov. Phys. Tech. Phys. **34**, 602 (1989)].
[4] A. I. Grigor'ev and S. O. Shiryaeva, Zh. Tekh. Fiz. **61**(3), 19 (1991) [Sov. Phys. Tech. Phys. **36**, 258 (1991)].
[5] S. O. Shiryaeva and A. I. Grigor'ev, Pis'ma Zh. Tekh. Fiz. **21**(16), 17 (1995) [Tech. Phys. Lett. **21**(8), 639 (1995)].

[6] S. O. Shiryaeva, Pis'ma Zh. Tekh. Fiz. **22**(4), 84 (1996) [Tech. Phys. Lett. **22**(2), 171 (1996)].
[7] S. O. Shiryaeva and A. I. Grigor'ev, Zh. Tekh. Fiz. **66**(9), 12 (1996) [Tech. Phys. **41**, 865 (1996)].
[8] V. A. Koromyslov, Yu. D. Rakhmanova, and S. O. Shiryaeva, Pis'ma Zh. Tekh. Fiz. **23**(14), 40 (1997) [Tech. Phys. Lett. **23**(7), 553 (1997)].
[9] J. R. Melcher and C. V. Smith, Phys. Fluids **12**, 778 (1969).
[10] S. O. Shiryaeva and A. I. Grigor'ev, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 5, 107 (1997).
[11] S. O. Shiryaeva and A. I. Grigor'ev, Zh. Tekh. Fiz. **67**(8), 34 (1997) [Tech. Phys. **42**, 884 (1997)].
[12] V. G. Levich, *Physicochemical Hydrodynamics* [Prentice–Hall, Englewood Cliffs, NJ (1962); Fizmatgiz, Moscow (1959), 699 pp.].
[13] J. A. Stratton, *Electromagnetic Theory* [McGraw–Hill, New York (1941); Gostekhizdat, Moscow (1948), 539 pp.].
[14] S. O. Shiryaeva, A. É. Lazaryants *et al.*, Preprint of the Institute of Microelectronics of the Russian Academy of Sciences, Yaroslavl, No. 27 (1994), 126 pp.
[15] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum: Irreducible Tensors, Spherical Harmonics, Vector Coupling Coefficients, 3nj Symbols* [World Scientific, Singapore–Teaneck, NJ (1988); Nauka, Leningrad (1975), 439 pp.].
[16] S. O. Shiryaeva, M. I. Munichev, and A. I. Grigor'ev, Zh. Tekh. Fiz. **66**(7), 1 (1996) [Tech. Phys. **41**, 635 (1996)].
[17] S. O. Shiryaeva, A. I. Grigor'ev, and I. D. Grigor'eva, Zh. Tekh. Fiz. **65**(2), 1 (1995) [Tech. Phys. **40**, 117 (1995)].
[18] S. O. Shiryaeva and A. I. Grigor'ev, Zh. Tekh. Fiz. **64**(3), 1 (1994) [Tech. Phys. **39**, 229 (1994)].

Translated by P. Shelnitz

Resonant oscillations of an elastic membrane on the bottom of a tank containing a heavy liquid

V. V. Alekseev, D. A. Indeĭtsev, and Yu. A. Mochalova

Institute of Mechanical-Engineering Problems, Russian Academy of Sciences, 199178 St. Petersburg, Russia
(Submitted April 8, 1988)
Zh. Tekh. Fiz. **69**, 37–42 (August 1999)

A study is made of the resonant oscillation modes in a three-dimensional channel filled by a heavy incompressible liquid, with a free surface at the top and an elastic membrane on the bottom. It is shown that for definite relations between the parameters of the channel and the inclusion there is a discrete spectrum, which extends only up to the waveguide cutoff frequency, in addition to the continuous spectrum of oscillation frequencies. The oscillation modes of the liquid have a localized character in the region of the inclusion. © *1999 American Institute of Physics.* [S1063-7842(99)00708-4]

INTRODUCTION

One of the major questions regarding the dynamic strength of elastic structures is the existence of a spectrum of natural oscillation frequencies. When the excitation frequency coincides with one of the natural frequencies, resonant oscillations appear in a structure, which lead, on the one hand, to acoustic emission into the surrounding medium and, on the other hand, to enhancement of the stressed state of the structure. Unlike elastic bodies of finite dimensions, which always have such a frequency spectrum, bodies which have at least one boundary of infinite length can have a mixed (discrete and continuous) spectrum when the parameters of the mechanical inclusions present have certain values.[1] One special feature of contact problems (oscillations of elastic bodies in liquids) is the indefiniteness of the response of the liquid to an oscillating body. Depending on the frequency range, the response can be defined as either inertial or dissipative.[2] In the latter case dissipation is defined as the removal of energy from the oscillating body to infinity in the form of waves which propagate in the liquid. In the present paper it is shown in the case of free oscillations of an elastic membrane fastened to the bottom of a channel of finite depth, which is filled with a heavy incompressible liquid, that localized oscillation modes of the liquid (trapping modes), which correspond to the discrete spectrum of natural oscillation frequencies of the particular mechanical system, can appear when the parameters of the waveguide and the membrane have certain values. The continuous part of the spectrum is located above the waveguide cutoff frequency and corresponds to diverging waves on the surface of the liquid. We note that the discrete spectrum is confined to the region below the cutoff frequency, in contrast to the familiar problem of the oscillations of an absolutely rigid die on the bottom of a channel.[2]

OSCILLATIONS OF A HEAVY LIQUID IN A CHANNEL WITH A MEMBRANE ON ITS BOTTOM

Let us consider small oscillations of a heavy incompressible liquid with a free surface in a three-dimensional channel of depth H that has an infinitely long elastic membrane of width $2a$ fastened to its bottom. The coordinate system is chosen so that the x,y plane coincides with the bottom of the channel, the y axis is directed along the membrane, and the z axis is directed upward and passes through the midline of the membrane (see Fig. 1). The bottom of the channel, except for the region $-a \leqslant x \leqslant a$, $-\infty < y < \infty$, is absolutely rigid. The disturbances acting in the liquid cause oscillations of the membrane, which are described by the function $W(x,y,t)$. The combined oscillations of the liquid and the membrane can lead to the formation of a system of surface waves propagating along the x and y axes.

The motion of the liquid is assumed to be governed by a velocity potential $\Phi(x,y,z,t)$ and is described by the linearized system of equations

$$\Delta \Phi = 0, \tag{1}$$

$$\frac{\partial \Phi}{\partial z} + \frac{1}{q}\frac{\partial^2 \Phi}{\partial t^2} = 0, \quad z = H, \tag{2}$$

$$\frac{\partial \Phi}{\partial z} = \begin{cases} \partial W/\partial t, & |x| \leqslant a, \\ 0, & |x| > a, \quad z = 0, \end{cases} \tag{3}$$

where g is the acceleration of gravity and Δ is the Laplacian operator.

The bottom of the channel is subjected to the pressure

$$p|_{z=0} = -\rho \frac{\partial \Phi}{\partial t} - \rho g \begin{cases} W, & |x| \leqslant a, \\ 0, & |x| > a, \end{cases} \tag{4}$$

where ρ is the density of the liquid.

The motion of the membrane is described with consideration of (4) by the equation

$$M\frac{\partial^2 W}{\partial t^2} - T\left(\frac{\partial^2 W}{\partial x^2} + \frac{\partial^2 W}{\partial y^2}\right) + \sigma W = \rho \frac{\partial \Phi}{\partial t} + \rho g W, \tag{5}$$

where M is the mass of the membrane per unit area, T is the tension, and σ is the stiffness of the elastic base.

The familiar radiation condition[3] is used to formulate the boundary condition at $|x| \to \infty$.

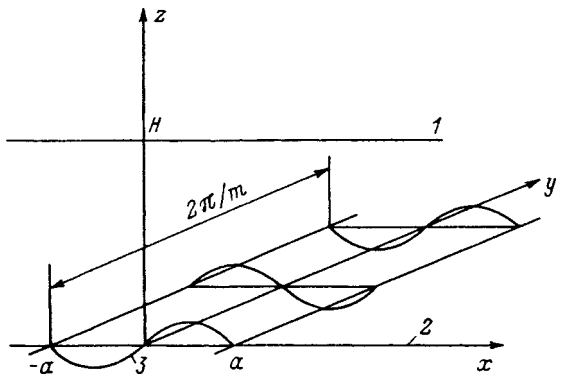

FIG. 1. Tank that has an elastic membrane on its bottom and is filled with a heavy liquid: *1* — free surface; *2* — tank bottom, *3* — membrane.

The oscillations of the liquid and the membrane are assumed to be periodic along the y axis with a wave number m and to occur with a frequency ω; therefore, the oscillation modes of the liquid and the membrane are represented in the form

$$W(x,y,t)=w(x)\exp(imy-i\omega t),$$

$$\Phi(x,y,z,t)=\varphi(x,z)\exp(imy-i\omega t).$$

The purpose of the present investigation is to find conditions under which there can exist a solution in the form of waves which do not propagate along the x axis, are localized in the region over the membrane, and move along the y axis. After separating the exponential factor and performing Fourier transformation in the system of equations (1)–(3), we obtain the following problem for determining $\varphi(x,z)$:

$$\frac{\partial^2\hat{\varphi}}{\partial t^2}-(k^2+m^2)\hat{\varphi}=0,$$

$$\frac{\partial\hat{\varphi}}{\partial z}-\frac{\omega^2}{g}\hat{\varphi}=0, \quad z=H,$$

$$\frac{\partial\hat{\varphi}}{\partial z}=-i\omega\hat{w}, \quad z=0, \tag{6}$$

where $\hat{\varphi}(k,z)=\int_{-\infty}^{\infty}\varphi(x,z)\exp(-ikx)\,dx$ and $\hat{w}(k)=\int_{-\infty}^{\infty}w(x)\exp(-ikx)\,dx$.

The solution of the problem (6) has the form

$$\hat{\varphi}(k,z)=i\hat{w}\frac{\omega}{r}\left[\frac{gr-\omega^2\tanh rH}{gr\tanh rH-\omega^2}\cosh rz-\sinh rz\right],$$

where $r^2=k^2+m^2$.

Performing the inverse Fourier transformation of $\hat{\varphi}(k,z)$ at $z=0$ using the properties of the convolution, we obtain

$$\varphi(x,z)|_{z=0}=i\omega\int_{-a}^{a}w(\xi)G(x-\xi)\,d\xi, \tag{7}$$

where $G(x)$ is represented in the form

$$G(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{gr-\omega^2\tanh rH}{r(gr\tanh rH-\omega^2)}\exp(ikx)\,dk. \tag{8}$$

The expression for $G(x)$ yields the dispersion relation

$$\omega^2=g\sqrt{k^2+m^2}\tanh H\sqrt{k^2+m^2}. \tag{9}$$

It follows from Eq. (9) that the dispersion curve $\omega^2(k)$, which specifies the surface waves in a layer, should not begin from zero, but from $\omega_*^2=gm\tanh Hm$, which we call the waveguide cutoff frequency. At $\omega<\omega_*$ Eq. (9) does not have real roots $k(\omega)$; therefore, the system of surface waves will not include waves which propagate along the x axis. Such waves can exist only at $\omega>\omega_*$. We now determine the form of $G(x)$ in these two cases.

a) In the case of $\omega<\omega_*$ the dispersion relation (9) has the purely imaginary roots $\pm ik_0$ for $k_0<m$ and $\pm ik_n$ ($n=1,2,\ldots$) for $k_n>m$, which can be determined from the relations

$$\omega^2=g\sqrt{m^2-k_0^2}\tanh H\sqrt{m^2-k_0^2},$$

$$\omega^2=-g\sqrt{k_n^2-m^2}\tan H\sqrt{k_n^2-m^2}, \tag{10}$$

respectively, and are first-order poles of the integrand in (8). We perform the integration in (8) on the complex plane using the residue theorem. As a result of this integration, we obtain the following expression for $G(x)$:

$$G(x)=\frac{2\sqrt{m^2-k_0^2}\exp(-ak_0|x_1|)}{k_0(\sinh 2H\sqrt{m^2-k_0^2}+2H\sqrt{m^2-k_0^2})}+\sum_{n=1}^{\infty}\frac{2\sqrt{k_n^2-m^2}\exp(-ak_n|x_1|)}{k_n(\sin 2H\sqrt{k_n^2-m^2}+2H\sqrt{k_n^2-m^2})}. \tag{11}$$

The expression obtained contains only modes which decay along x and does not contain waves which propagate along the x axis. The dimensionless variable $x_1=x/a$ is used in Eq. (11) and below. The subscript 1 is henceforth omitted.

b) In the case of $\omega>\omega_*$ the dispersion relation (9) has the real roots $\pm k_0$ and the purely imaginary roots $\pm ik_n$ ($n=1,2,\ldots$) for $k_n>m$, which can be determined from the second relation in (10). The integral (8) is calculated using a radiation condition stating that at $x\to\pm\infty$ there are only waves which go to infinity. This is accomplished by a standard procedure involving the limiting absorption principle, whereby the features on the real axis produced by the roots $\pm k_0$ are eliminated. As a result of the integration of (8) we obtain

$$G(x)=i\frac{2\sqrt{m^2+k_0^2}\exp(iak_0|x|)}{k_0(\sinh 2H\sqrt{m^2+k_0^2}+2H\sqrt{m^2+k_0^2})}+\sum_{n=1}^{\infty}\frac{2\sqrt{k_n^2-m^2}\exp(-ak_n|x|)}{k_n(\sin 2H\sqrt{k_n^2-m^2}+2H\sqrt{k_n^2-m^2})}. \tag{12}$$

The presence of an exponential function of imaginary argument in (12) indicates the presence of waves propagating along the x axis.

In accordance with (7), Eq. (5), which describes the oscillations of the membrane under the layer of liquid, takes the form

$$\frac{\partial^2 w}{\partial x^2}+\frac{a^2}{T}(M\omega^2-\sigma-Tm^2+\rho g)w$$

$$=-\rho\frac{\omega^2}{T}a^3\int_{-1}^{1}w(\xi)G(x-\xi)\,d\xi \tag{13}$$

under the boundary conditions $w(\pm 1)=0$.

The problem of determining the steady oscillations of the liquid and the membrane was reduced to finding the real spectrum of eigenvalues of the integrodifferential operator (13), where the spectral parameter is the frequency ω^2. It was shown in Ref. 4 that an equation of the form (13) can be reduced to a homogeneous Fredholm integral equation of the second kind. The existence of a discrete spectrum below the cutoff frequency follows from known theorems. Above the cutoff frequency it is impossible to prove the existence of a real discrete spectrum of the problem (13) in the general case because of the complex-valued nature of $G(x)$ and the absence of a general theory for nonself-conjugate operators.

DETERMINATION OF THE REAL SPECTRUM OF EIGENVALUES BELOW THE CUTOFF FREQUENCY

Obtaining an analytical solution of Eq. (13) in general form is difficult because of the fairly cumbersome expression for $G(x)$. Therefore, we shall henceforth adopt the long-wavelength approximation, which is often used in practice. We shall study waves which propagate along the y axis and have a wavelength $\lambda=2\pi/m$ that is much greater than the channel depth H. We shall also assume that the relation $a/H\gg 1$ is satisfied. Then, because of the small value of their wave numbers m in the first expression in (10), we obtain the dispersion relation

$$\omega^2=g(m^2-k_0^2)H. \tag{14}$$

The first term in the expression (11) for $G(x)$ can be represented in an approximation in the form

$$g_1(x)=\frac{1}{2k_0H}\exp(-ak_0|x|).$$

The second term in (11) can be simplified using the following arguments. The values of k_n are determined from the second relation in (10), which yields the inequality $-\pi/2+\pi n<H\sqrt{k_n^2-m^2}<\pi n$ $(n=1,2,\ldots)$. Then, since $a/H\gg 1$, the ak_n appearing in the exponents are much greater than unity. In addition, the inequality $|\sin 2H\sqrt{k_n^2-m^2}|\ll 2H\sqrt{k_n^2-m^2}$ holds for any n. Then, applying the familiar relation $\lim_{z\to\infty} z\exp(-z|x|)=2\delta(x)$ (Ref. 5) to each term in the sum in (11) and neglecting the sine in the denominator, we obtain the following approximate expression:

$$g_2(x)=\delta(x)\frac{2H}{a}\sum_{n=1}^{\infty}\frac{1}{(k_nH)^2}.$$

Setting $k_nH=\pi n$ in the last expression, we can sum the series obtained as a result. This gives

$$g_2(x)=\delta(x)\frac{H}{3a}. \tag{15}$$

Then, in the long-wavelength approximation we have the following expression for $G(x)$:

$$G(x)=\frac{1}{2k_0H}\exp(-ak_0|x|)+\delta(x)\frac{H}{3a}. \tag{16}$$

Substituting (16) into Eq. (13), we obtain

$$\frac{\partial^2 w}{\partial x^2}+\gamma^2 w=-A_0\int_{-1}^{1}w(\xi)\exp(-ak_0|x-\xi|)\,d\xi, \tag{17}$$

where we have introduced the following notation:

$$\gamma^2=\frac{a^2}{T}\left(M\omega^2-\sigma-Tm^2+\rho g+\frac{1}{3}\rho H\omega^2\right),$$

$$A_0=\frac{1}{2}\rho a^3\frac{\omega^2}{k_0TH}.$$

Performing the double differentiation in Eq. (17) with consideration of the conditions $w(\pm 1)=0$, we arrive at the boundary-value problem

$$\frac{\partial^4 w}{\partial x^4}+B_1\frac{\partial^2 w}{\partial x^2}-B_2 w=0 \tag{18}$$

with the boundary conditions

$$w(\pm 1)=0,$$

$$\left.\frac{\partial^2 w}{\partial x^2}\right|_{x=\pm 1}=-A_0\int_{-1}^{1}w(\xi)\exp(-ak_0|\pm 1-\xi|)\,d\xi. \tag{19}$$

Here we have introduced the following notation:

$$B_1=\gamma^2-(ak_0)^2,\quad B_2=(ak_0)^2\gamma^2+\rho\omega^2\frac{a^4}{TH}. \tag{20}$$

Equation (18) is equivalent to the equation describing the oscillations of a beam with a force B_1 pressing on both of its ends and a mass per unit length B_2. The general solution of Eq. (18) has the form

$$w(x)=C_1\sinh S_1x+C_2\cosh S_1x+C_3\sin S_2x$$
$$+C_4\cos S_2x,$$

where

$$S_1=\sqrt{-B_1/2+\sqrt{B_1^2/4+B_2}},$$

$$S_2=\sqrt{B_1/2+\sqrt{B_1^2/4+B_2}}.$$

Substituting the general solution into the boundary conditions (19), we obtain an algebraic system for determining the coefficients C_1, C_2, C_3, and C_4. The explicit form of the system is not presented here because of its cumbersome nature. The determinant of this system, set equal to zero, gives a frequency equation for determining ω^2, which separates into two equations for the cases of symmetric oscillations ($C_1=C_3=0$) and asymmetric oscillations ($C_2=C_4=0$): a) symmetric oscillations

$$\frac{ak_0\tanh ak_0+S_2\tan S_2}{(ak_0)^2+S_2^2}-\frac{ak_0\tanh ak_0-S_1\tanh S_1}{(ak_0)^2-S_1^2}$$
$$=\frac{\exp(ak_0)}{2A_0\cosh ak_0}(S_1^2+S_2^2); \tag{21}$$

b) asymmetric oscillations

$$\frac{ak_0\coth ak_0-S_2\cot S_2}{(ak_0)^2+S_2^2}-\frac{ak_0\coth ak_0-S_1\coth S_1}{(ak_0)^2-S_1^2}$$
$$=\frac{\exp(ak_0)}{2A_0\sinh ak_0}(S_1^2+S_2^2). \tag{22}$$

The quantities S_1, S_2, and k_0 in Eqs. (21) and (22) are functions of the unknown parameter ω^2 through Eqs. (20) and (14). In calculating the natural frequencies we confine ourselves to the case of symmetric oscillations. We represent Eq. (21) in the form

$$\tan S_2=S_2^3\frac{\exp(ak_0)}{2A_0\cosh ak_0}+S_2\left[\frac{\exp(ak_0)}{2A_0\cosh ak_0}[S_1^2+(ak_0)^2]+\frac{ak_0\tanh ak_0-S_1\tanh S_1}{(ak_0)^2-S_1^2}\right]$$
$$+\frac{1}{S_2}\left[\frac{\exp(ak_0)}{2A_0\cosh ak_0}S_1^2(ak_0)^2+\frac{ak_0\tanh ak_0-S_1\tanh S_1}{(ak_0)^2-S_1^2}(ak_0)^2-ak_0\tanh ak_0\right]=F(S_2). \tag{23}$$

The solutions of Eq. (23) are determined graphically from the points of intersection of the plots of $\tan S_2$ and $F(S_2)$. The domain of $F(S_2)$ is bounded on the right by the values of S_2 at which the condition

$$\omega^2<\omega_*^2=gm^2H \tag{24}$$

is violated.

Calculations performed over a broad range of variation of the waveguide and membrane parameters show that the number of positive roots ω^2 depends on the relation between the parameters but is always restricted by the condition (24). For example, for $M=20$ kg/m^2, $H=1$ m, $\sigma=10^4$ N/m^3, and $T=10^4$ kg/s^2, Eq. (23) gives the following values for the natural frequencies. In the case of $a/H=5$, for $m=0.1$ m^{-1} there is one solution $\omega^2/\omega_*^2=0.34$, and for $m=0.2$ m^{-1} there is also one solution $\omega^2/\omega_*^2=0.27$. In the case of $a/H=10$, for $m=0.1$ m^{-1} there is one solution $\omega^2/\omega_*^2=0.1$, and for $m=0.2$ m^{-1} there are two solutions $\omega^2/\omega_*^2=0.11$ and 0.92. The case in which the inequality

$$\sigma+Tm^2-\rho g<0 \tag{25}$$

is satisfied should receive particular attention.

The calculations show that in this case Eq. (23) has only negative roots ω^2. Physically, Eq. (25) means that the membrane has a weak tension T and a low value for the stiffness of the elastic base σ. Because of this the sagging of the membrane begins to increase without bound due to the difference in hydrostatic pressure acting on the stiff bottom and the sagging surface of the membrane. This difference in hydrostatic pressure is equal to $\rho g w(x)$ and is represented by the last term on the right hand side of Eq. (4). The same phenomenon can be interpreted on the basis of Eq. (18), which describes the oscillations of a beam with a force B_1 pressing on its ends. When the condition (25) is satisfied, the force B_1 exceeds a certain critical value, at which unrestricted sagging of the beam (the familiar Euler instability) occurs.

On the basis of the calculations performed it can be concluded that in the case of $\omega^2<\omega_*^2$ the integrodifferential operator (17), which is the long-wavelength approximation of the operator of the general form (13), has a discrete spectrum of positive eigenvalues ω^2 when the condition $\sigma+Tm^2-\rho g>0$ is satisfied, and that their number is restricted. Thus, it can be stated that surface waves which are localized in the region above the membrane and move over it along the y axis can form in the layer of liquid. The existence of eigenfrequencies can lead to the appearance of resonant oscillations of underwater structures when they are loaded by forces which vary harmonically with time, as well as to a loss of buoyancy of a structure on the bottom of a reservoir, as practical experience has shown.

DETERMINATION OF THE REAL SPECTRUM OF EIGENFREQUENCIES ABOVE THE CUTOFF FREQUENCY

Let us consider the oscillations of a liquid and a membrane for $\omega^2>\omega_*^2$. The function $G(x)$ appearing in Eq. (13) is described by (12). We retain the assumption of long waves propagating along the y axis. The structure of the roots k_n is the same as in the case of $\omega^2<\omega_*^2$. Therefore, the approximate expression (15), which was obtained for the case of $\omega^2<\omega_*^2$, remains valid for the sum appearing in (12).

We represent the function $G(x)$ in the form

$$G(x)=iA_1\exp(iak_0|x|)+\delta(x)\frac{H}{3a}, \tag{26}$$

where

$$A_1=\frac{2\sqrt{k_0^2+m^2}}{k_0(\sinh 2H\sqrt{k_0^2+m^2}+2H\sqrt{k_0^2+m^2})}.$$

After substituting (26) into Eq. (13), we obtain

$$\frac{\partial^2 w}{\partial x^2}+\gamma^2 w=-i\rho\frac{\omega^2a^3}{T}A_1\int_{-1}^{1}w(\xi)\exp(iak_0|x-\xi|)\,d\xi. \tag{27}$$

Let us determine the pressure on the bottom of the channel at $x>a$. From (4) and (7) with consideration of (26) we obtain

$$p(x) = -i\rho\omega^2 a A_1 \exp(iak_0 x) \times \int_{-1}^{1} w(\xi)\exp(-iak_0\xi)\, d\xi - \rho\frac{\omega^2 H}{3} w(x).$$

The presence of a term containing $\exp(iak_0x)$ in the last expression signifies the presence of waves propagating along the x axis. The existence of a discrete spectrum of oscillations of the liquid and the membrane can be ensured only in the absence of waves carrying off energy to infinity. For this purpose, we must require satisfaction of the condition for the absence of radiation

$$\int_{-1}^{1} w(\xi)\exp(-iak_0\xi)\, d\xi = 0. \tag{28}$$

Let us examine the case of symmetric oscillations of the membrane. In this case the condition (28) has the following form:

$$\int_{-1}^{1} w(\xi)\cos(ak_0\xi)\, d\xi = 0. \tag{29}$$

In the case of symmetric oscillations, Eq. (27) with allowance for the condition (29) takes the form

$$\frac{\partial^2 w}{\partial x^2} + \gamma^2 w = 2\rho\frac{\omega^2 a^3}{T} A_1 \int_{-1}^{x} w(\xi)\sin[ak_0(x-\xi)]\, d\xi. \tag{30}$$

Differentiating Eq. (30) twice with consideration of the boundary condition $w(1)=0$, we obtain the problem

$$\frac{\partial^4 w}{\partial x^4} + B_3\frac{\partial^2 w}{\partial x^2} - B_4 w = 0, \tag{31}$$

$$w(1)=0, \quad w_{xx}(1)=0, \tag{32}$$

where

$$B_3 = \gamma^2 + (ak_0)^2, \quad B_4 = -(ak_0)^2\gamma^2 + \frac{2\rho a^4 k_0}{T} A_1\omega^2.$$

The presence of a Volterra integral operator on the right-hand side of Eq. (30) leads to the additional condition

$$\frac{\partial^3 w}{\partial x^3} + \gamma^2\frac{\partial w}{\partial x} = 0, \quad x = 1, \tag{33}$$

which ensures equivalence of the problem (31) and (32) and the original problem (30).

The solution of Eq. (31) for symmetric oscillations has the form

$$w(x) = C_2\cosh S_1 x + C_4\cos S_2 x, \tag{34}$$

where

$$S_1 = \sqrt{-B_3/2 + \sqrt{B_3^2/4 + B_4}},$$

$$S_2 = \sqrt{B_3/2 + \sqrt{B_3^2/4 + B_4}}.$$

Substitution of the solution (34) into the boundary conditions (32) leads to a system of equations for C_2 and C_4, which yields the frequency equation $(S_1^2+S_2^2)\cosh S_1\cos S_2 = 0$, as well as the equality $C_2 = 0$. It can be seen that the frequency equation can hold only if $\cos S_2 = 0$, i.e., $(S_2)_n = \pi/2 + \pi n$ $(n = 0,1,\ldots)$.

Let us determine the possibility that the condition of zero radiation (29) and also the condition (33) are satisfied for solutions of the type $w(x) = \cos[(S_2)_n x]$. A calculation of the integral (28) under the condition $\cos[(S_2)_n] = 0$ leads to the equality

$$\frac{(S_2)_n}{(S_2)_n^2 - (ak_0)^2}\sin[(S_2)_n]\cos ak_0 = 0, \tag{35}$$

which can hold only if $\cos ak_0 = 0$. The condition (33) leads to the opposite requirement $ak_0 = 0$.

The treatment of the case of asymmetric oscillations leads to a similar result. It can thus be concluded that the original operator (13) does not have a real discrete spectrum at $\omega^2 > \omega_*^2$.

This work was carried out with support from the Russian Fund for Fundamental Research (Project Code 96-01-01153a) and the Foundation ''Integratsiya'' (No. 589).

[1] A. K. Abramyan, V. L. Andreev, and D. A. Indeĭtsev, Model. Mekh. **6**, 34 (1992).
[2] D. A. Indeĭtsev and E. V. Osipova, Zh. Tekh. Fiz. **66**(8), 124 (1996) [Tech. Phys. **41**, 811 (1996)].
[3] A. N. Tikhonov and A. A. Samarskii, *Equations of Mathematical Physics* [Pergamon Press, Oxford (1964); Glavizdat, Moscow (1953), 736 pp.].
[4] D. A. Indeitchev, in *Proceedings of EUROMECH-316*, Manchester (1994), Vol. 3, p. 304.
[5] W. Kecs and P. P. Teodorescu, *Applications of the Theory of Distributions in Mechanics* [Abacus, Tunbridge Wells (1974); Mir, Moscow (1978)].

Translated by P. Shelnitz

Stability of a bubble in a dielectric liquid in an external electrostatic field

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

Yaroslavl State University, 150000 Yaroslavl, Russia
(Submitted April 29, 1998)
Zh. Tekh. Fiz. **69**, 43–48 (August 1999)

Critical instability conditions are found for a gas bubble in a liquid dielectric in a uniform external electrostatic field $\mathbf{E}_0$. It is shown that they depend both on the magnitude of $\mathbf{E}_0$ and on the properties of the liquid, as well as on the gas pressure in the bubble. In a linear approximation with respect to the square of the eccentricity of an equilibrium spheroidal form, the equilibrium eccentricity of the bubble exceeds the equilibrium eccentricity of a drop in the field $\mathbf{E}_0$. The gas pressure in the bubble lowers the critical electric field $\mathbf{E}_0$ for development of an instability in the bubble. [S1063-7842(99)00808-9]

1. The instability of gas bubbles in a uniform electrostatic field $\mathbf{E}_0$ is of interest in studies of the electrical breakdown of liquid dielectrics, since at the time of breakdown, a region of reduced density develops at the cathode which is treated as the appearance of a microscopic gas (vapor) bubble.[1,2] Although the bubble instability problem has an important technological application, only a few studies have been made of the behavior of bubbles in electric fields, while many papers[3] have dealt with the similar problem of the instability of liquid drops in an electric field $\mathbf{E}_0$.

The classic paper on the state of a liquid drop in an external electrostatic is Ref. 4 which contains results that can be used for studying bubbles with the conductivities of the medium and drop taken into account. In that paper[4] the shape of drops with a finite conductivity and dielectric constant in a conducting medium were studied under the assumption that the drop has the spheroidal equilibrium shape in $\mathbf{E}_0$ and the potential energy is to be minimized.[4] Several scenarios for the drop behavior were studied, depending on the relationship of the physical properties of the drop and the surrounding medium:[4] (a) drop and medium are both absolute dielectrics, (b) conducting drop in a dielectric medium, (c) dielectric drop in a conducting medium, (d) zero deformation of a spheroidal drop in $\mathbf{E}_0$, and (e) deformation of the drop to an oblate spheroid. This paper[4] was followed by others[5,6] in which the individual cases distinguished in it[4] were examined in more detail.

The case of a dielectric drop and bubble in a dielectric medium have been studied in detail[5] experimentally and theoretically by determining the deformation experimentally as a function of the magnitude of $\mathbf{E}_0$ for gas bubbles in hydrocarbon oil and water drops in silicone oil. It was found that the bubble stretches without limit, while a drop experiences an instability with the emission of large number of daughter droplets. In addition, it was shown[5] that the shape of a drop in $\mathbf{E}_0$ can be taken to be spheroidal with fairly good accuracy. Based on equating the pressures at the pole and equator of the drop and minimizing the energy of this state, it was shown that liquid dielectric drops can become unstable only if the dielectric constant of the drop exceeds that of the surrounding medium by a factor of 20. This condition is satisfied with ample margin for drops of a well conducting liquid (e.g., water) in a vacuum. This case has been studied in detail theoretically.[6] Based on the pressure balance at all points on the equilibrium surface (at the pole point and on the equator) of a spheroidal ideal conducting drop, the magnitude of the deformations of the drop was determined as a function of the external electric field $\mathbf{E}_0$. It turned out that an ideally conducting drop becomes unstable at a certain critical field $\mathbf{E}_0$.

Since a gas bubble placed in a liquid dielectric is stable for any field $\mathbf{E}_0$ (because the dielectric constant of gases is low) according to Ref. 5, an attempt has been made[7] to treat the possible instability of a bubble in $\mathbf{E}_0$ in terms of the compressibility of the gas. The behavior of vapor bubbles in a dielectric liquid was studied[7] on the basis of the assumption that the equilibrium bubble shape in $\mathbf{E}_0$ is spheroidal and using the requirement that the thermodynamic potentials of the system be equal, as well as on the basis of the principle of minimizing the energy of the equilibrium state. The results of these theoretical and experimental studies have been confirmed by numerical calculations.[8]

Numerical solutions of the equations of electrodynamics and hydrodynamics have been used[8] to obtain the shape of liquid drops for different ratios of the conductivities λ and dielectric constants α of the drop and the surrounding medium. The ranges of values of λ and α within which different types of instability are possible was determined. Two types of instability were distinguished:[8] the ''dripping off'' of small daughter droplets from sharp ends of the parent drop, and the division of a drop into two. The possible type of instability for each pair of λ and α was determined. An instability which ends up with the ejection of fine drops was observed for dielectric liquids when the dielectric constant of the drop liquid was 19–21 times that of the surrounding medium or, in the case of a conducting liquid, if the conductivity of the drop was 28–29 times (a value given for liquids with equal dielectric constants) that of the surrounding medium. It also turned out that the instability associated with breakup of a drop into two is possible only in a conducting liquid.

An analysis of the papers cited above shows that a

bubble, which differs from a drop mainly in having a gas pressure and in its compressibility, can become unstable only when the gas filling it or the bubble walls are good conductors. If the bubble is in an absolute dielectric, it can experience an instability only if the gas in the bubble is a good conductor, perhaps through ionization. If the gas bubble is in a liquid that has even a weak conductivity, then charge carriers from the liquid may settle on the bubble walls. In this case, the good conductivity may be the result of high surface mobility of the charge carriers compared to their bulk mobility. In the following discussion, in order to clarify the conditions for stability of a bubble in a liquid, we limit ourselves to considering an idealized model of a bubble with ideally conducting walls in an absolute dielectric.

2. Let us suppose that the liquid dielectric with a dielectric constant ε has a free surface at atmospheric pressure P_{at}. We consider an initially spherical bubble of radius R_0 with an ideally conducting surface in the dielectric. We shall assume that the bubble is motionless, i.e., the buoyancy force is balanced by the force of inertia (which may occur in a rotation device[5]) or by a force owing to a temperature difference between the upper and lower layers of liquid.[9] An equilibrium spherical bubble must obey the pressure balance conditions at its walls,

$$P_0 = P_{at} + 2\sigma/R_0, \tag{1}$$

where σ is the coefficient of surface tension at the interphase boundary and P_0 is the gas pressure in the bubble.

After creation of an electric field $\mathbf{E}_0$ in the dielectric, the bubble is elongated along $\mathbf{E}_0$ and acquires a shape similar to a spheroid of revolution,[5–7] whose equation in spherical coordinates with the origin at the center of the drop has the form

$$F(r,\Theta) = r - R\frac{(1-e^2)^{1/2}}{(1-e^2\cos^2\Theta)^{1/2}} \equiv r - R\,\eta(\Theta).$$

We set ourselves the question of whether the pressure, temperature, and volume of the gas in the bubble change as it elongates. We assume that the rate at which the gas filling the bubble dissolves in the dielectric is so low that its mass can be regarded as constant during the time of deformation in the field $\mathbf{E}_0$. From a molecular standpoint, the pressure of the gas at the bubble boundary is determined only by the concentration n of gas particles and the gas temperature T, i.e.,

$$P = nkT,$$

where k is Boltzmann's constant.

Let us examine the conditions for equilibrium at the equator. Since the pressure of the electrostatic field is zero for the points on the equator, the equilibrium condition is easily written as

$$P = P_{at} + 2\sigma H, \tag{2}$$

where H is the Gaussian curvature at this point, with

$$2H = \operatorname{div}\left(\frac{\operatorname{grad} F(r,\Theta)}{|\operatorname{grad} F(r,\Theta)|}\right). \tag{3}$$

For a point lying on the equator, the Gaussian curvature in the linear approximation with respect to e^2 is given by[10]

$$H = \frac{1}{R_0}\left(1 - \frac{e^2}{3}\right). \tag{3a}$$

We now assume that the electric field has increased by an infinitesimal amount, which causes an infinitesimal increase in the eccentricity of the bubble. Equation (3a) implies that for a spheroidal bubble $H < 1/R_0$; then, on comparing Eqs. (1) and (2) it is easy to see that $P < P_0$, i.e., the gas pressure in the bubble decreases. From a molecular standpoint, the pressure decrease can be caused by an increase in the volume or by a reduction in the temperature. Since the total heat capacity of the gas in the bubble is much less than that of the entire liquid dielectric, it is natural to assume that the deformation of the bubble in the field $\mathbf{E}_0$ is isothermal and that the gas temperature does not change. Therefore, in the external field, the bubble, which has well conducting walls, not only is elongated into a spheroid of rotation in the direction of $\mathbf{E}_0$, but also increases in volume.

As the shape and volume of the bubble change, there is a change in the energy of the system, which, under the conditions, should be regarded as a function of the radius R of an equally large spherical bubble and of the square of the eccentricity, e^2. The potential energy of the bubble is determined by the sum of the surface and electrical energy, plus the work done on the system by the external forces as the bubble expands. The surface tension energy is given by[10]

$$U_\sigma = 2\pi\sigma R^2(1-e^2)^{1/6}\left((1-e^2)^{1/2} + \frac{\sin^{-1}e}{e}\right). \tag{4}$$

The electrical energy for a bubble extended in the direction of an external field $\mathbf{E}_0$ is given by[10]

$$U_E = -\frac{1}{2}\frac{\varepsilon E_0^2 R^3}{3}\frac{e^3}{(1-e^2)(\tanh^{-1}e - e)}. \tag{5}$$

The change in the internal energy of the system owing to the work of isothermal expansion is[11,12]

$$U_g = -NkT\ln\frac{V}{V_0} + P_{at}(V - V_0), \tag{6}$$

where N is the number of gas molecules in the bubble and V_0 is the initial volume of the bubble.

The change in energy of the system during elongation including Eq. (6) and an expansion of Eqs. (4) and (5) in a series in the square of the eccentricity can be written as

$$\Delta U = 4\pi\sigma R^2\left(1 + \frac{2}{45}e^4\right) - \frac{\varepsilon E_0^2 R^3}{2}\left(1 + \frac{2}{5}e^2 + \frac{58}{175}e^4\right) - NkT\ln\frac{V}{V_0} + P_{at}(V - V_0) - 4\pi\sigma R_0^2 + \frac{\varepsilon E_0^2 R_0^3}{2}. \tag{7}$$

Setting the derivatives of the energy change (7) with respect to the square of the eccentricity e^2 and with respect to the bubble radius R equal to zero, we obtain a system of

two equations for the unknowns e^2 and R corresponding to the equilibrium spheroidal shape of the bubble,[13]

$$e^2=9\frac{\varepsilon E_0^2 R}{16\sigma\pi}\equiv\frac{9}{16\pi}w,$$

$$1-3\frac{\varepsilon E_0^2 R}{16\sigma\pi}-(P-P_{\text{at}})R/2\sigma\equiv\beta-1+\frac{3}{16\pi}w=0, \qquad (8)$$

where $w=(\varepsilon E_0^2 R)/\sigma$ and $\beta=\{(P-P_{\text{at}})R\}/2\sigma$.

Therefore, as opposed to a drop, whose equilibrium eccentricity depends only on w,[4,10] the stable state of a bubble is determined by two equations because the bubble volume does not remain constant as it is deformed in the field $\mathbf{E}_0$.

3. We shall analyze the system of Eqs. (8), assuming that during deformation the bubble volume increases in a way such that the radius of the equilibrium spherical bubble increases by a factor of $(K+1)$, i.e.,

$$R=R_0(1+K). \qquad (9)$$

Using the isothermal condition and Eq. (9), we express the final pressure of the gas in the bubble in terms of the initial pressure as

$$P=\frac{P_0}{(1+K)^3}. \qquad (10)$$

We consider the initial pressure in the bubble to exceed atmospheric by a factor of μ, i.e.,

$$P_0=\mu P_{\text{at}}. \qquad (11)$$

Substituting Eq. (11) in Eq. (1), we find μ

$$\mu=1+\frac{2\sigma}{P_{\text{at}}R_0}. \qquad (12)$$

Given Eqs. (9), (10), and (11), we reduce the second equation of the system (8) to

$$1-\frac{e^2}{3}-\left(\frac{\mu}{(1+K)^3}-1\right)\frac{P_{\text{at}}R_0}{2\sigma}(1+K)=0. \qquad (13)$$

Using Eq. (12), we rewrite Eq. (13) as

$$1-\frac{e^2}{3}-\left(\frac{\mu}{(1+K)^3}-1\right)\frac{1}{\mu-1}(1+K)=0. \qquad (14)$$

If it is assumed that the relative increase K in the radius is small, in a linear approximation with respect to K and e^2, Eq. (14) yields

$$K=\frac{\mu-1}{2\mu+1}\frac{e^2}{3}, \qquad (15)$$

i.e., K is of the same order of smallness as e^2. Substituting Eq. (9) into the first of Eqs. (8) and using Eq. (15), we obtain an expression for the square of the eccentricity of the bubble,

$$e^2=\frac{9w}{16\pi}\left(1+\frac{9w}{48\pi}\frac{\mu-1}{2\mu+1}\right), \qquad (16)$$

where $w=(\varepsilon E_0^2 R_0)/\sigma$ is the Taylor parameter for the bubble.

Equation (16) implies that accounting for the gas pressure in a bubble leads to a dependence of its equilibrium eccentricity on the initial gas pressure. In this regard, it is interesting to examine the effect of the gas pressure on the critical conditions for the instability of a gas bubble in an electric field $\mathbf{E}_0$.

4. We shall do the calculation using a method[6] based on equating the pressures at the pole and at the equator.

Equation (3) yields the capillary pressure at the pole and at the equator of a spheroidal bubble:

$$P_\sigma^{eq}=\frac{\sigma(1-e^2)^{5/6}}{R_0(1+K)}\left(1+\frac{1}{(1-e^2)}\right),$$

$$P_\sigma^{p}=\frac{2\sigma}{R_0(1+K)(1-e^2)^{2/3}}. \qquad (17)$$

Using the expression for the electrostatic field at the surface of a spheroidal bubble,[14]

$$\mathbf{E}=\mathbf{E}_0\frac{e^3}{(1-e^2)^{1/2}(\tanh^{-1}e-e)}\frac{\eta(\Theta)\cos\Theta}{(1-e^2\eta^2(\Theta)\cos^2\Theta)^{1/2}},$$

it is easy to find expressions for the electric field pressure at the same points on the bubble,

$$P_E^{eq}=0,\quad P_E^{p}=\frac{\varepsilon E_0^2 e^6}{8\pi(1-e^2)^2(\tanh^{-1}e-e)^2}. \qquad (18)$$

Given Eqs. (18), (17), and (20), we rewrite the equal pressure conditions at the equator and pole as

$$\frac{\mu P_{\text{at}}}{(1+K)^3}-\frac{\sigma(1-e^2)^{5/6}}{R_0(1+K)}\left(1+\frac{1}{(1-e^2)}\right)-P_{\text{at}}=0, \qquad (19)$$

$$\frac{\mu P_{\text{at}}}{(1+K)^3}+\frac{\varepsilon E_0^2 e^6}{8\pi(1-e^2)^2(\tanh^{-1}e-e)^2}-\frac{2\sigma}{R_0(1+K)(1-e^2)^{2/3}}-P_{\text{at}}=0. \qquad (20)$$

Subtracting Eq. (19) from Eq. (20) and using Eq. (12), we obtain a system of equations for the unknowns K and e^2 which can be used to find the critical instability conditions for an uncharged bubble with well conducting walls in a uniform external electrostatic field $\mathbf{E}_0$:

$$\frac{we^6}{16\pi(1-e^2)^2(\tanh^{-1}(e)-e)^2}=\frac{1}{(1+K)}\left\{\frac{1}{(1-e^2)^{2/3}}-\frac{(1-e^2)^{5/6}}{2}\left[1+\frac{1}{(1-e^2)}\right]\right\}, \qquad (21)$$

$$\mu-(1+K)^3-\frac{(\mu-1)(1+K)^2(1-e^2)^{5/6}}{2}\times[1+(1-e^2)^{-1}]=0. \qquad (22)$$

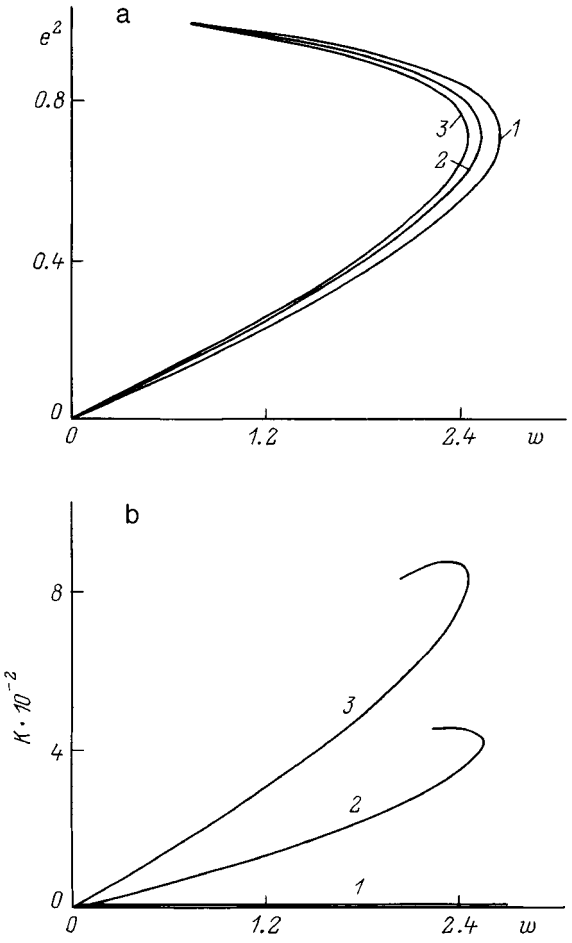

FIG. 1. The square of the eccentricity of a bubble (a) and the relative change in the bubble radius (b) as functions of the Taylor parameter for $\mu = 1.1$ (*1*), 2 (*2*), 5 (*3*).

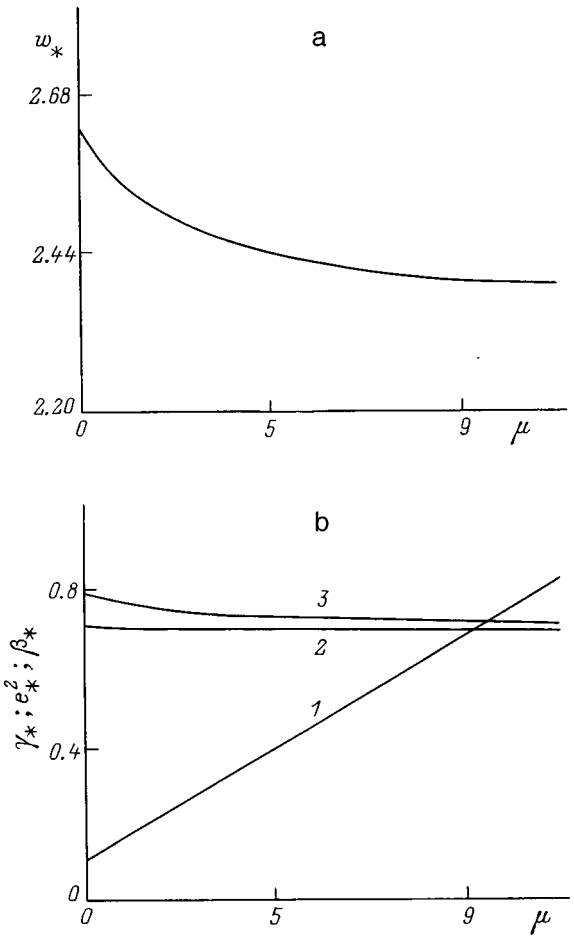

FIG. 2. The critical Taylor parameter (a) and the critical dimensionless gas pressure (*1*), the square of the eccentricity (*2*), and the parameter β_* (*3*) (b) as functions of the dimensionless initial pressure of the gas in a bubble.

Equations (21) and (22) determine the eccentricity e^2 of the bubble and the relative change in its radius, K, as functions of the Taylor parameter w for a given μ.

To analyze the system of Eqs. (21) and (22) we have done numerical calculations and plotted the square e^2 of the eccentricity and K as functions of the Taylor parameter w for different values of μ. These curves are shown in Figs. 1a and 1b. It is clear that an increase in the initial internal pressure in a bubble for fixed w leads to an increase in the equilibrium bubble eccentricity. It is also evident that at some point, corresponding to the maximum of w, the derivative $\partial e^2/\partial w$ goes to infinity. According to Refs. 5 and 6, when the Taylor parameter reaches this critical value $w = w_*$, the bubble becomes unstable. Thus for a given μ the system of equations (21) and (22) has a unique solution corresponding to the critical Taylor parameter w_* that separates the stable and unstable states of the bubble.

The critical Taylor parameter w_* and critical square of the eccentricity e_*^2 are plotted in Figs. 2a and 2b as functions of the parameter μ, which characterizes the pressure in the bubble prior to its elongation. Evidently, an increase in the initial gas pressure μ produces a significant drop in the critical Taylor parameter w_* but has very little effect on the critical eccentricity e_*^2.

Figure 2b also shows plots of $\gamma_* = P_*/P_{\text{at}}$ and $\beta_* = (P_* - P_{\text{at}})R_0/(2\sigma)$ at the time instability sets in for various values of the parameter μ characterizing the initial pressure in the bubble.

5. We now ascertain how the compressibility χ of the gas affects the critical Taylor parameter and the equilibrium eccentricity of the bubble. We shall assume that the gas pressure in the bubble is given by

$$P = P_0 - \frac{1}{\chi}((1+K)^3 - 1), \tag{23}$$

rather than by Eq. (10), which is valid only for an ideal gas.

By choosing different values of χ, it is possible to obtain different special cases of a liquid drop, $\chi = 10^{-4}\ \text{atm}^{-1}$, a bubble of real gas, $\chi = 1\ \text{atm}^{-1}$, and a bubble of ideal gas, $\chi = (1+K)^3/P_0$.

If we do calculations similar to those in Sec. 4 using Eq. (23), then we obtain two equations for determining K and e^2 as functions of the Taylor parameter w. The first equation again has the form of Eq. (21), while instead of Eq. (22), we obtain

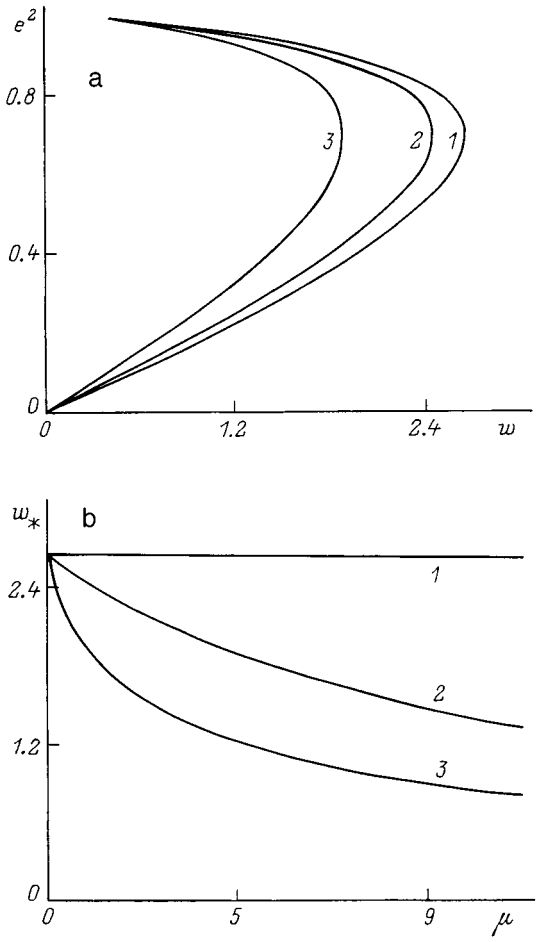

FIG. 3. The square of the eccentricity as a function of the Taylor parameter for $\mu=2$ (a) and the critical Taylor parameter as a function of the initial pressure of the gas in a bubble (b) for: *1* — liquid drop with $\chi P_{at}=10^{-4}$; *2, 3* — gas bubble with $\chi P_{at}=1$ and *4*, respectively.

$$\left[\mu-1+\frac{1}{\chi P_{at}}\right](1+K)-\frac{(1+K)^4}{\chi P_{at}}$$
$$-\frac{(\mu-1)(1-e^2)^{5/6}}{2}\left[1+\frac{1}{(1-e^2)}\right]=0. \qquad (24)$$

For an ideal gas, i.e., for $\chi=(1+K)^3/\mu P_{at}$, Eq. (24) transforms to Eq. (22).

Calculations according to Eqs. (24) and (21) are shown in Figs. 3a and 3b. It is evident from Fig. 3a that increasing the coefficient of isothermal compressibility χ for a fixed Taylor parameter w and initial gas pressure μ leads to a rise in the square e^2 of the eccentricity of the bubble. The plot of e^2 as a function of w has a point at which the derivative $\partial e^2/\partial w$ becomes infinite. This makes it possible to determine the critical Taylor parameter w_*. Plots of w_* as a function of the initial pressure in a bubble are shown in Fig. 3b. Evidently, increasing the coefficient of isothermal compressibility causes a drop in w_*.

Note that the compressibility and internal pressure have a significant effect on the equilibrium eccentricity of the bubble and on the critical Taylor parameter w_* when the initial bubble radius is less than a characteristic R_*, i.e., when

$$R_0 \leqslant R_* = 2\sigma/P_{at}.$$

CONCLUSION

Increasing the volume of a gas bubble as it is deformed in a uniform electrostatic field $\mathbf{E}_0$ leads to a dependence of the critical Taylor parameter w_* and the equilibrium eccentricity e^2 of the bubble on the gas pressure μ. Increasing the initial gas pressure μ leads to a reduction in w_* and an increase in e^2. In addition, the critical Taylor parameter w_* and the equilibrium eccentricity e^2 of the bubble depend on the coefficient of isothermal compressibility χ.

[1] V. V. Glazkov, O. A. Sinkevich, and P. V. Smirnov, Teplofiz. Vys. Temp. **29**, 1095 (1991).
[2] I. V. Pylaeva, O. A. Sinkevich, and P. V. Smirnov, Teplofiz. Vys. Temp. **30**, 367 (1992).
[3] A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk. Mekh. Zhidk. Gaza, No. 3, pp. 2–28 (1994).
[4] Ch. T. O'Konski and F. E. Harris, J. Chem. Phys. **61**, 1172 (1957).
[5] C. G. Garton and Z. Krasucki, Trans. Faraday Soc. **60**, 211 (1964).
[6] G. Taylor, Proc. R. Soc. London, Ser. A **280**, 383 (1964).
[7] K. J. Cheng and J. B. Chaddock, Phys. Lett. A **106**, 51 (1984).
[8] J. D. Sherwood, J. Fluid Mech. **188**, 133 (1988).
[9] Ya. E. Geguzin, *Bubbles* [in Russian], Nauka, Moscow (1985), 174 pp.
[10] A. I. Grigor'ev, S. O. Shiryaeva, and E. I. Belavina, Zh. Tekh. Fiz. **59**(6), 27 (1989) [Sov. Phys. Tech. Phys. **34**, 602 (1989)].
[11] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 2nd ed., Pergamon Press, Oxford (1969); Nauka, Moscow (1964), 567 pp.].
[12] A. I. Grigor'ev, V. A. Koromyslov, and A. N. Zharov, Pis'ma Zh. Tekh. Fiz. **23**(19), 60 (1997) [Tech. Phys. Lett. **23**(10), 760 (1997)].
[13] S. O. Shiryaeva and A. I. Grigor'ev, Zh. Tekh. Fiz. **65**(2), 11 (1995) [Tech. Phys. **40**, 122 (1995)].
[14] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, 2nd ed., rev. and enl., with L. P. Pitaevskii [Pergamon Press, Oxford (1984); Nauka, Moscow (1982), 620 pp.].

Translated by D. H. McNeill

Local enhancement of a uniform electrostatic field near the tip of a spheroidal drop

S. I. Shchukin and A. I. Grigor'ev

Yaroslavl State University, 150000 Yaroslavl, Russia
(Submitted April 29, 1998)
Zh. Tekh. Fiz. **69**, 49–54 (August 1999)

The enhancement in a uniform electrostatic field at the tip of a spheroidal drop is shown to depend on the dielectric constant of the drop material, its initial radius, and the external electric field and to become greater as these increase. The loss of stability of a drop in an external electrostatic field that is accompanied by a very rapid growth in the magnitude of the spheroidal deformation causes a rapid, transient enhancement of the field at its tip. © *1999 American Institute of Physics.* [S1063-7842(99)00908-3]

Calculating the local enhancement of a uniform external electrostatic field at the tip of a dielectric spheroid is of interest for analyzing physical aspects of the formation of fractal structures from natural smoke and dust aerosols,[1,2] in research on discharge initiation by streak lightning,[3,4] and in studies of the field evaporation of ions during mass spectrometry of nonvolatile and thermally unstable substances.[5–8] Despite the simplicity with which this problem can be stated, solving it involves certain computational difficulties; therefore, it has yet been solved, and the present paper is devoted to analyzing it.

ENHANCEMENT OF THE ELECTRIC FIELD NEAR THE TIP OF A DIELECTRIC SPHEROID IN A UNIFORM ELECTROSTATIC FIELD AS A FUNCTION OF THE SHAPE OF THE SPHEROID AND ITS DIELECTRIC CONSTANT

When a dielectric spheroid in a vacuum is extended along a uniform external electrostatic field $\mathbf{E}_0$, the electric field inside it is uniform and coincides in direction with $\mathbf{E}_0$, but its strength E_e is given by[9]

$$E_e=\frac{E_0}{1+(\varepsilon-1)n_x}, \tag{1}$$

where ε is the dielectric constant of the material and n_x is the depolarization coefficient, which depends on the shape (but not on the size) of the spheroid as

$$n_x=\frac{k\ln(k+\sqrt{k^2-1})-\sqrt{k^2-1}}{(k^2-1)^{3/2}}, \tag{2}$$

where k is the ratio of the semiaxes of the spheroid, with $k>1$.

Given the continuity of the normal component of the electric induction on passing through the interphase boundary, for the electric field E_x in an external medium at the tip of the spheroid we can write[9]

$$E_x=\frac{\varepsilon E_0}{1+(\varepsilon-1)n_x}. \tag{3}$$

If we assume in the zeroth approximation that the shape of the particle is independent of the external field, i. e., if we speak of a rigid spheroidal particle, rather than a drop, then E_x depends linearly on E_0 and the clearest representation of the dependence of interest to use will be a plot of the field enhancement E_x/E_0 as a function of the dielectric constant ε of the substance and the ratio k of the semiaxes of the spheroid. Equations (2) and (3) imply that

$$\frac{E_x}{E_0}=\frac{\varepsilon(k^2-1)^{3/2}}{(k^2-1)^{3/2}+(\varepsilon-1)[k\ln(k+\sqrt{k^2-1})-\sqrt{k^2-1}]}. \tag{4}$$

The graph in Fig. 1 shows that: 1) the field enhancement is greater for more elongated particles, and 2) the field enhancement is higher for particles with a higher dielectric constant.

For an ideally conducting particle, the limit of Eq. (3) as $\varepsilon\to\infty$ gives

$$E_c=\frac{E_0}{n_x}. \tag{5}$$

The plot of the field enhancement E_c/E_0 as a function of the ratio k of the semiaxes of a conducting spheroid in Fig. 2 shows that the field enhancement increases monotonically with increasing k.

Therefore, the increase of a uniform electrostatic field near the tip of a rigid spheroid depends on the dielectric constant ε of the material and on the ratio k of the semiaxes of the spheroid, but is independent of its size.

DEPENDENCE OF THE ELECTRIC FIELD NEAR THE TIP OF A SPHEROIDAL DROP ON THE STRENGTH OF THE UNIFORM EXTERNAL ELECTROSTATIC FIELD AND THE DROP SIZE

If we assume that n_x in Eq. (3) is the depolarization coefficient of a dielectric drop in an equilibrium state, then the parameter k can no longer be regarded as independent, since it depends on the Taylor parameter W for an equilibrium spheroidal drop shape. The parameter W characterizes the stability of a drop relative to the polarization charge and is a function of the external electrostatic field E_0, the radius R, and the coefficient of surface tension σ of the drop:[10]

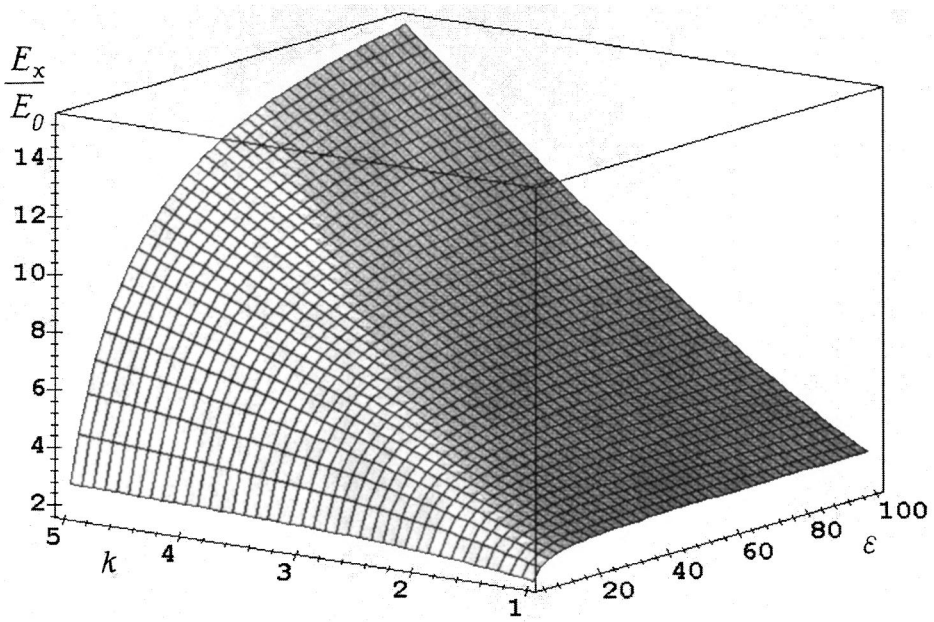

FIG. 1. Field enhancement in the neighborhood of the tip of a dielectric spheroid as a function of the ratio of the semiaxes of the spheroid and of the dielectric constant of the material.

$$W=\frac{4\pi\varepsilon_0 E^2 R}{\sigma}. \tag{6}$$

The function $k=k(W)$ for the equilibrium state of a drop in an electrostatic field is determined by the conditions for a minimum total potential energy of the drop,

$$\frac{\partial U}{\partial k}=0, \tag{7}$$

$$\frac{\partial^2 U}{\partial k^2}\geqslant 0. \tag{8}$$

The total potential energy of a drop in this case can be regarded as the sum of the energy of surface tension and the electrostatic energy of the drop in the field $\mathbf{E}_0$. Taking the potential energy of a spheroidal drop in dimensionless and referring it to the energy of a spherical drop, we obtain

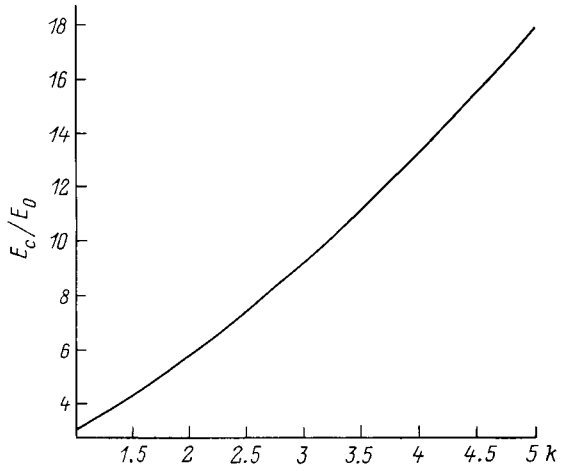

FIG. 2. Field enhancement for a conducting spheroidal drop as a function of the ratio of the semiaxes of the spheroid.

$$U\equiv U_\sigma+U_e=\left\{2\pi\left[\frac{1}{k^{2/3}}+\frac{k^{4/3}\tan^{-1}(\sqrt{k^2-1})}{\sqrt{k^2-1}}\right]-\frac{W(\varepsilon-1)}{6[1+(\varepsilon-1)n_x]}\right\}\left[4\pi-\frac{W(\varepsilon-1)}{2(2+\varepsilon)}\right]^{-1}. \tag{9}$$

Substituting Eq. (9) in Eq. (7) and solving the resulting equation analytically for W, we obtain

$$W=2\pi\left[\frac{2}{3}k^{-5/3}-\frac{4}{3}\frac{k^{1/3}C}{A}-\frac{k^{1/3}}{A^2}+\frac{k^{7/3}C}{A^3}\right]\times[3kA-(2k^2+1)B]^{-1}\frac{D^2}{A},$$

where

$$A=\sqrt{k^2-1},\quad B=\ln(k+\sqrt{k^2-1}),$$

$$C=\tan^{-1}(\sqrt{k^2-1}),$$

$$D=(k^2-\varepsilon)\sqrt{k^2-1}\,(\varepsilon-1)^{-1}+k\ln(k+\sqrt{k^2-1}). \tag{10}$$

After similar transformations of Eq. (8), we find

$$W=2\pi k^{1/3}\left(\frac{10}{9}k^{-3}+\frac{4}{9}\frac{C}{A}k^{-1}+\frac{5}{3}\frac{k^{-3}}{A^2}-\frac{11}{3}\frac{kC}{A^3}-3\frac{k}{A^4}+3\frac{k^3C}{A^5}\right)D^2\left\{\frac{(2k^2+1)Bk}{A}-3A^2+4kAB-4k^2+1+2A[3kA-(2k^2+1)B]\left(\frac{3kA}{\varepsilon-1}+B\right)D^{-1}\right\}^{-1}. \tag{11}$$

Equations (10) and (11) are transcendental with respect to k. In addition, as can be seen from plots of Eqs. (10) and (11) (surfaces *1* and *2* in Fig. 3, respectively), the inverse function $k=k(W)$ is multivalued and becomes single-valued only for $\varepsilon<20.8$, when Eq. (10) becomes monotonically in-

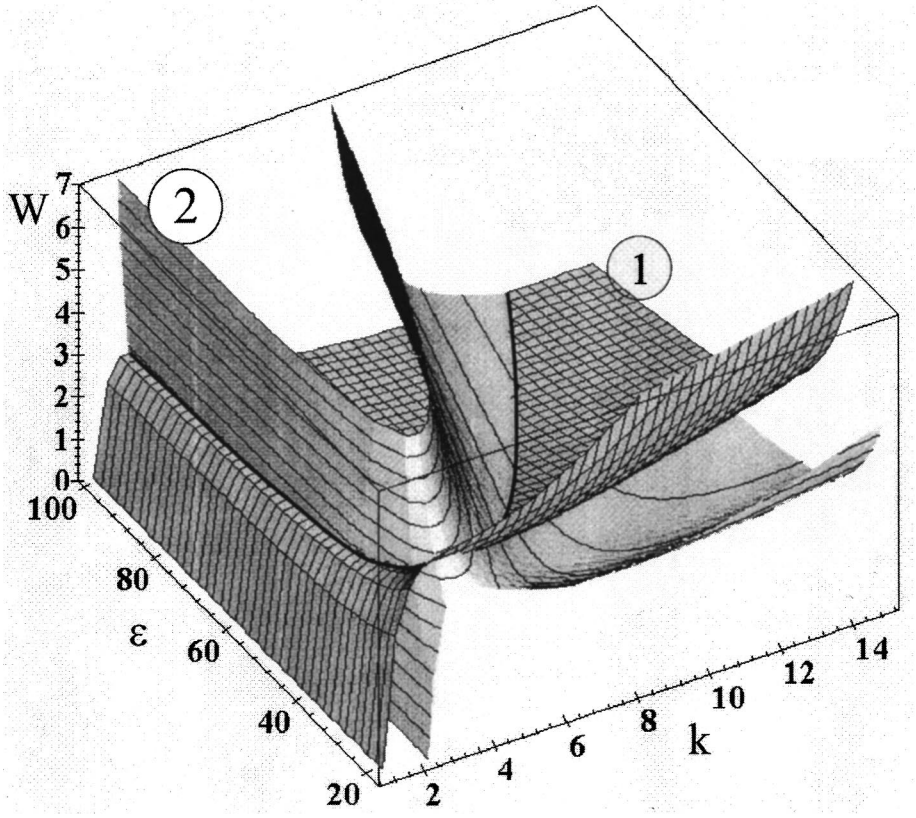

FIG. 3. The Taylor parameter as a function of the ratio of the semiaxes of the spheroid and of the dielectric constant.

creasing. Surface *1* corresponds to the extreme values of the energy of a spheroidally deformed drop and surface *2* determines the location of the zeros of the second derivative of the drop energy with respect to the magnitude of the spheroidal deformation. The equilibrium states of the drop correspond to the locus of points on surface *1* lying outside surface *2*.

Some calculations using Eqs. (10) and (11) for drops of water ($\varepsilon = 81$) are shown in Fig. 4. Curve AD in this figure corresponds to the solution of Eq. (7) and determines the set of points (k, W) at which the total potential energy of a spheroidally deformed drop takes its extreme values. The curves FG and HJ in this figure correspond to the solution of Eq. (8) and determine the boundary of the regions where the second derivative of the total potential energy U of the drop with respect to the spheroidal deformation k have a fixed sign. The set of points where $\partial^2 U/\partial k^2 > 0$ lies below the curve FG and above the curve HJ, and $\partial^2 U/\partial k^2 < 0$ in the region between the curves FG and HJ. Therefore, the segments AB and CD of curve AD correspond to state of the drop with a minimum potential energy, while segment BC corresponds to local maxima of the potential energy of the drop as it is deformed spheroidally. The presence of such a ''potential barrier'' inhibits spontaneous lengthening and destruction of the drop when the Taylor parameter $W < W_*$. For $W = W_*$ the height of the potential barrier is zero because of the discontinuity in the potential energy as a function of the spheroidal deformation, so that when the drop enters this state it begins to lengthen spontaneously; here the amplitude of the spheroidal deformation increases much more rapidly than exponentially.[11] Using the generally

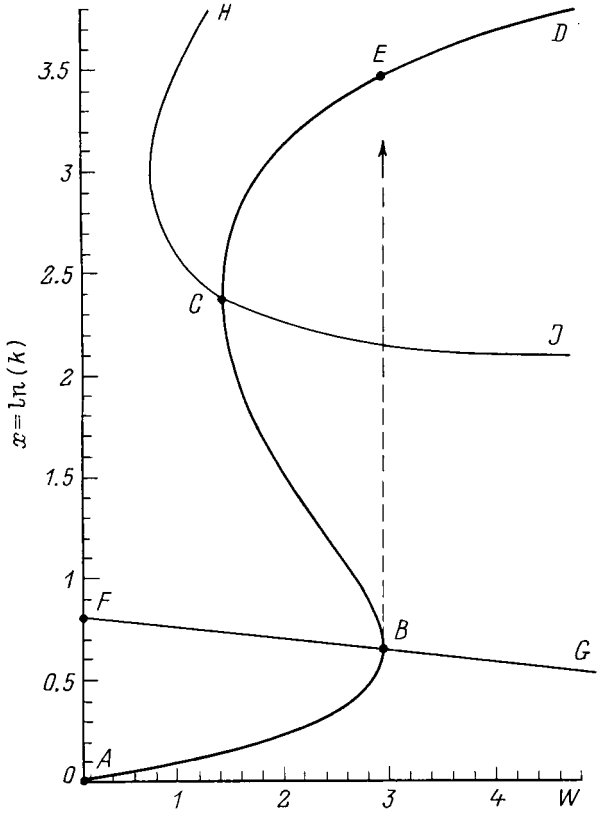

FIG. 4. The logarithm of the ratio of the semiaxes of a spheroidal drop, $x = \ln k$, as a function of the Taylor parameter W according to Eqs. (7) and (8).

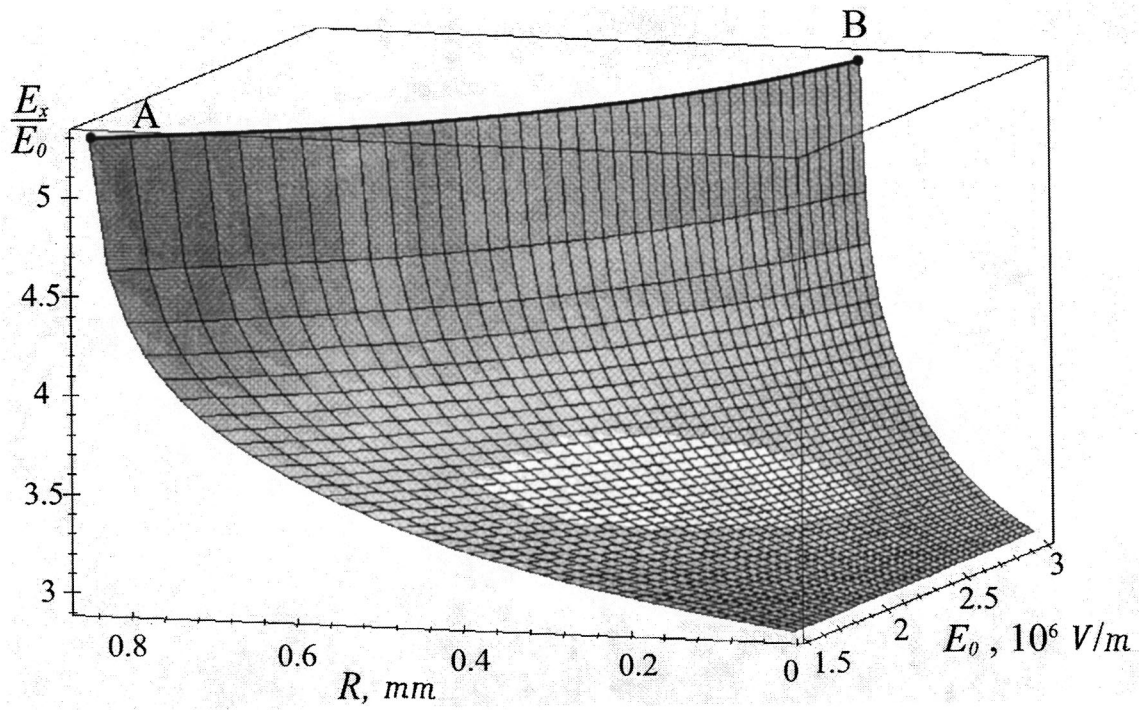

FIG. 5. Enhancement in the electrostatic field at the tip of a spheroidal drop in its subcritical equilibrium state as a function of the equilibrium spherical drop radius and of the external electric field.

accepted terminology, we refer to the states corresponding to segment AB of the curve as subcritical, to those with a Taylor parameter $W=W_*$ as critical, and to those corresponding to segment CD as supercritical.

For liquids with a dielectric constant $\varepsilon>20.8$ and a surface tension coefficient σ, we can propose the following scheme for calculating the electric field near the tip of a spheroidal drop of radius R in its equilibrium state in an external electrostatic field $\mathbf{E}_0$: 1) starting with a known value of ε it is necessary to find the critical Taylor parameter W_* by jointly solving Eqs. (7) and (8) numerically; 2) starting with the specific characteristics of the drop and field, σ, R, and $\mathbf{E}_0$, Eq. (6) is to be used to find the corresponding instantaneous value of W; 3) if the value of the Taylor parameter found in the preceding step is $W<W_*$, then the parameter k for the equilibrium state of the drop must be found by numerical solution of Eq. (7); 4) finally, E_x can be found using Eq. (4).

In order to reduce the volume of numerical calculations, it is appropriate to find analytical expressions which approximate $W_*=W_*(\varepsilon)$ and $k=k(\varepsilon,W)$ with sufficient accuracy.

For ε ranging from 20.8 to ∞, the expression

$$W_*=\frac{101+26.6\varepsilon+0.308\varepsilon^2}{-60.8+9.05\varepsilon+0.12\varepsilon^2} \tag{12}$$

is accurate to the third significant digit, which is sufficient for most practical calculations.

Since approximation of a function of two variables is an extremely complicated problem, an approximation for $k=k(W)$ was found for $\varepsilon=81$. For subcritical Taylor parameters, this approximation gives

$$k=\frac{4.08-1.85W+0.1658W^2}{4.08-2.167W+0.2693W^2}. \tag{13}$$

Equation (13) is also accurate to the third significant digit.

These approximations make it possible to determine the electrostatic field at the tip of an equilibrium spheroidal water drop in an external electrostatic field $\mathbf{E}_0$ as a function of the strength E_0 of this field and of the radius R of an equally large spherical drop. A plot of the resulting function is shown in Fig. 5. Curve AB in this graph corresponds to the critical states of the drop.

If we assume that, as the external field is increased, when the drop reaches a critical state, it begins to elongate and enters a supercritical equilibrium spheroidal state (Fig. 4); then, on interpolating the numerically calculated $k=k(W)$ for the supercritical states by the function

$$k=-0.676W^2+12.14W+2.33, \tag{14}$$

it is possible to obtain the curve shown in Fig. 6. Experiments show that the supercritical drop states are not stable and, furthermore, that drop breakup is observed before a supercritical equilibrium spheroidal state is reached. However, a sudden increase in the electric field at the tip of a spheroidal drop on passing from a subcritical to a supercritical state can precede its breakup. This also applies to emission proturberances formed on a plane liquid surface in the Tonks–Frenkel instability and may be the basis of the ''cathode spot'' phenomenon.[12–14]

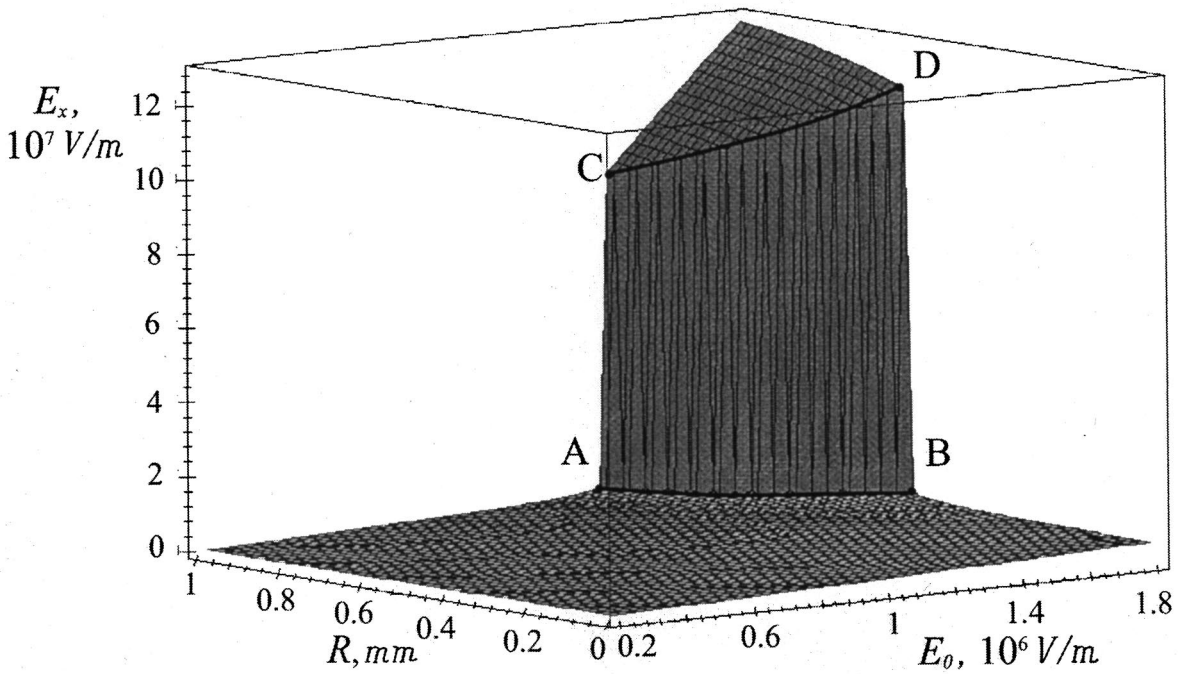

FIG. 6. Strength of the electrostatic field at the tip of a spheroidal drop as a function of equilibrium spherical drop radius and of the external electrostatic field. The discontinuous increase in the electric field (transition from curve AB to CD) corresponds to a transition from a subcritical to a supercritical equilibrium state.

CONCLUSION

The enhancement of a uniform external electrostatic field $\mathbf{E}_0$ in the neighborhood of the tip of a spheroidal rigid particle is determined by the eccentricity (degree of spheroidality) of the particle and dielectric constant of the particle material, but is independent of the geometric size of the particle. The enhancement for a uniform external electrostatic field in the neighborhood of the tip of a drop which takes a spheroidal shape in a field $\mathbf{E}_0$ depends on the characteristics of the drop material (dielectric constant, coefficient of surface tension) and on the drop size, as well as on the strength of the external field. When a drop loses stability with respect to the polarization charge, its eccentricity increases very rapidly in time (faster than exponentially) and this leads to such a rapid rise in the field enhancement at its tip that the drop breaks up or the polarization charge begins to be ejected.

[1] Ya. R. Akhvlediani, in *Physics of Clouds. Proceedings of the Institute of Geophysics of Academy of Sciences of the Georgian SSR* [in Russian], Metsniereba, Tbilisi (1975), Vol. 36, pp. 53–57.

[2] A. A. Lushnikov, A. E. Negin, A. V. Pakhomov, and B. M. Smirnov, Usp. Fiz. Nauk **161**, No. 2, 113 (1991) [Sov. Phys. Usp. **34**, 160 (1991)].

[3] V. A. Dyachuk and V. M. Muchnik, Dokl. Akad. Nauk **248**, No. 1, 60 (1979).

[4] A. I. Grigor'ev and S. O. Shiryaeva, Phys. Scr. **54**, 660 (1996).

[5] A. I. Grigor'ev and S. O. Shiryaeva, Inzh.-Fiz. Zh. **60**, 632 (1991).

[6] S. O. Shiryaeva and A. I. Grigor'ev, Zh. Tekh. Fiz. **63**(8), 162 (1993) [Tech. Phys. **38**, 715 (1993)].

[7] N. B. Zolotoĭ, Zh. Tekh. Fiz. **65**(11), 159 (1995) [Tech. Phys. **40**, 1175 (1995)].

[8] N. B. Zolotoĭ and G. V. Karpov, Dokl. Akad. Nauk. **348**, 336 (1996).

[9] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, 2nd ed., rev. and enl., with L. P. Pitaevskii [Pergamon Press, Oxford (1984); Nauka, Moscow (1982), 620 pp.].

[10] G. Taylor, Proc. R. Soc. London, Ser. A **280**, 383 (1964).

[11] S. O. Shiryaeva, A. I. Grigor'ev, and I. D. Girgor'eva, Zh. Tekh. Fiz. **65**(9), 39 (1995) [Tech. Phys. **40**, 885 (1995)].

[12] G. N. Furseĭ and P. N. Vorontsov-Vel'yaminov, Zh. Tekh. Fiz. **37**, 1880 (1967) [Tech. Phys. **12**, 1377 (1967)].

[13] I. P. Stakhanov, M. M. Fiks, and D. G. Fil'kin, Zh. Tekh. Fiz. **44**, 1373 (1974) [Sov. Phys. Tech. Phys. **19**, 861 (1974)].

[14] A. I. Grigor'ev and S. O. Shiryaeva, Izv. Akad. Nauk, Mekh. Zhidk. Gaza No. 3, pp. 3–22 (1994).

Translated by D. H. McNeill

Electrical charging of engines in the efflux of combustion products. Experimental results

Yu. A. Nagel'

M. V. Keldysh Research Center, 125438 Moscow, Russia
(Submitted July 13, 1998)
Zh. Tekh. Fiz. **69**, 55–59 (August 1999)

The electrical charging of jet engines during efflux of combustion products is studied. Measurements are made of the electrical current beyond the exit section of a nozzle as a function of such parameters as the characteristic length of the space charge region, the electrical potential of the engine, and the pressure of the surrounding medium. Possible practical applications of the results (estimating the charging current and the rate of rise and maximum value of the potential on a flight vehicle) are discussed. © *1999 American Institute of Physics.* [S1063-7842(99)01008-9]

INTRODUCTION

Jet engines are a major source of static electricity on flight vehicles, as they eject charged particles of combustion products during their operation.[1] This effect, which is known as engine charging, may cause some undesirable effects (electrical discharges on board the vehicle, interference with control and radio communication systems, disturbances of the ionospheric plasma, etc.). Practical work on improving the reliability of flight vehicles has drawn the attention of researchers to the problem of engine charging.

Of the factors which contribute to engine charging, two have been studied: violation of the quasineutrality condition for the combustion products, which form a weakly ionized multicomponent gas, in the boundary layer,[2,3] and the development of instabilities in the distribution of electrical charge over the entire volume of the outflowing combustion products.[4] A space charge develops near the channel wall because of a difference in the thermal speeds of the electrons and ions, and in a moving gas this charge gives rise to an electrical current J_w along the wall and, when it flows out the channel (nozzle), a current I_a in the surrounding space. In the one-dimensional approximation[2]

$$J_w = e\int_0^{\delta}[n_i(x)-n_e(x)]v(x)dx, \tag{1}$$

where J_w is the current per unit length of the transverse channel cross section, $n_i(n_e)$ is the ion (electron) density, δ is the thickness of the space charge layer, $v(x)$ is the outflow velocity, and the x axis directed perpendicular to the surface.

The current J_w was calculated for three hypothetical cases: a Maxwellian velocity distribution of the electrons and collisionless motion of the ions within the charged layer,[2] and surfaces that ideally reflect and absorb charged particles.[3] The current J_w is positive and, for other conditions fixed, it is higher for higher charged particle densities (for higher temperature, and for lower ionization potential of the combustion products).

The development of instabilities in the electrical charge distribution is related to an acoustic interaction that produces macroscopic fluctuations in the combustion product volume. Their characteristic spatial and time scales greatly exceed the corresponding Debye length and electron–neutral atom collision time. The sign of the unneutralized charge released into the air, and of the corresponding current I_a, changes from positive to negative as the initial charged particle density is lowered.[4]

Besides the contributions of these factors to engine charging, one can indirectly (by specifying the initial data and boundary conditions at the channel wall) include thermionic emission from the walls and from macroscopic particles contained in the combustion products, deposition of electrons on macroscopic particles, easily ionized alkali metals in fuel impurities, etc. The picture of engine charging based on these ideas is, on the whole, consistent with the available experimental data.[1–3,5–7] However, these data are limited in volume and have been obtained by different methods and under nonidentical conditions, so is hard to proceed with further scientific interpretation of engine charging or improvements in the physical and mathematical model for its mechanism. In this paper we present some generalized results from experimental studies of the electrical charging of liquid- and solid-fuel jet engines. As opposed to the earlier data,[1–3,5–7] here a larger number of parameters were determined (associated both with processes taking place inside the engine and with the external conditions), and these have been obtained on the basis of a unified methodological approach.

EXPERIMENTAL SETUP. MEASUREMENT TECHNIQUES

Our initial considerations regarding the setup of the experiment were based on solutions of a one-dimensional model of the flow of a charged gas beyond the exit section of a unipolar charge source[8] and reduced to the following. In general, the current I_a in the surrounding space should depend on the internal (applying purely to the engine) and external conditions. The external conditions are the set of parameters that determine the density of the space charge beyond the exit section of the nozzle: the characteristic

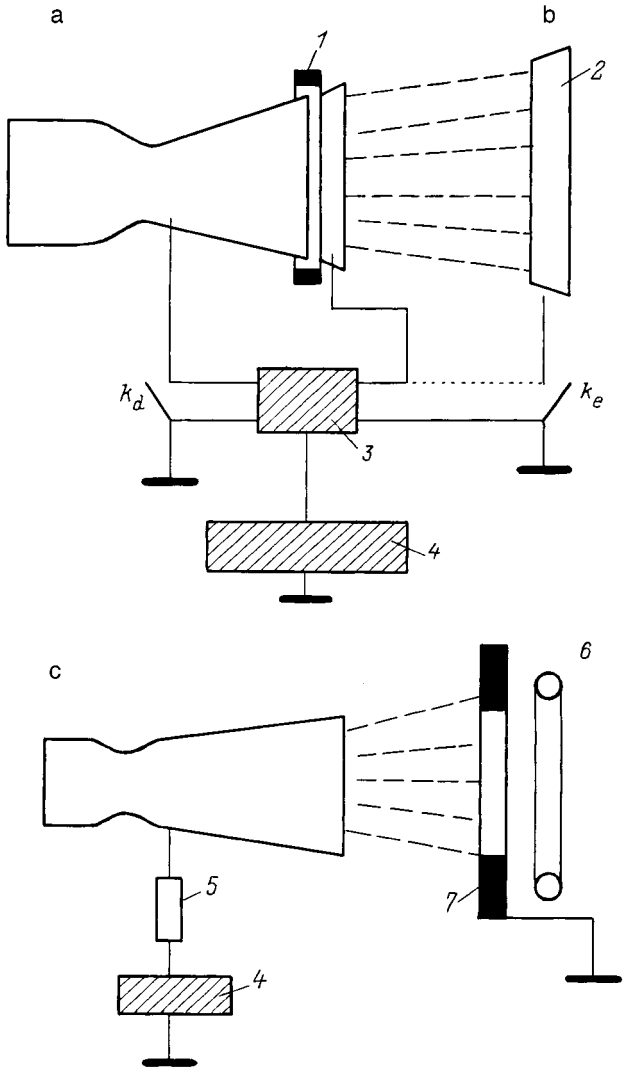

FIG. 1. Setup for measurements with ''hot'' (a) and ''cold'' (b) electrodes and a neutralizer (c): (*1*) insulator, (*2*) electrode, (*3*) current amplifier, (*4*) mirror–galvanometer oscillograph, (*5*) load resistance, (*6,7*) annular corona and shielding electrodes.

length l_a of the space charge region, the electrical potential U of the engine, and the electric field E_a at the nozzle. The electric field $|E_a| \leqslant \min\{E_p, E_s\}$, where $E_p(E_s)$ is the electrical breakdown strength of the combustion products (surrounding medium). If $l_a \to 0$, then the current $I_a \to I_m$, where I_m is the maximum current (saturation current), whose magnitude is determined only by the internal conditions: the composition, temperature, and pressure of the combustion products, etc.

The measurement setup chosen in accordance with these considerations is shown in Fig. 1. The parameter l_a was determined using a profiled electrode or neutralizer mounted beyond the nozzle exit. If the electrode was mounted so that its inner surface was washed by the cold, essentially unionized periphery of the stream (Fig. 1b), then it acted as a collector of the radially drifting unneutralized charges in the stream, and the current in the measurement circuit was $I \approx I_a$. When the electrode was placed immediately adjacent to the exit section and the profile of its inner surface matched the nozzle profile, the electrode was washed by the hot part (actually, the core) of the stream, so that the processes leading to charging could show up equally on the nozzle walls and on the electrode. The signal measured in this case equals the difference between the currents to the electrode and engine, $I = I_e - I_d$, while the current in the stream in the exit–electrode segment is $I_a = kI$, where k is a coefficient which is to be determined. If the gap between the nozzle and the electrode is small and the change (over the length of the gap and the electrode) in the parameters of the combustion products can be neglected, then for $I_e \propto D_e$ and $I_d \propto D_a$, where D_e (D_a) is the exit diameter of the electrode (nozzle), then the coefficient $k = 1/(D_e/D_a - 1)$. The use of a neutralizer made it possible to avoid direct contact between the electrodes and the stream. The unneutralized charge in the stream was neutralized by creating an electrical charge of the required density and polarity in a corona discharge between electrodes (*6* and *7* in Fig. 1c) connected to a high voltage source. The measured signal $I \approx I_a$. In order to avoid an electrical discharge on the engine and the leads to the measurement circuit, the neutralizer was mounted some distance from the nozzle exit. A scheme with a ''hot'' electrode was used when the outflow was fully expanded and $l_a/D_a < 0.2$. In this case, it was sufficient to use only the instrumentation in the electrode circuit (switch k_e in Fig. 1 open); this was important when it was difficult to isolate the engine from ''ground'' (switch k_d closed). In the case of underexpanded outflow, but also with $l_a/D_a \geqslant 0.2$, the measurements were done using a scheme with a ''cold'' electrode (grounded engine) or a neutralizer (isolated engine). Comparison tests of the three measurement schemes for $l_a/D_a \approx 0.2$ yielded satisfactory agreement. A variable load resistance R_n (*5* in Fig. 1) was used to measure the potential U. Under atmospheric conditions, it was not possible to vary the parameter E_a. This became possible when the smaller-sized engines were placed in a vacuum chamber.

TABLE I.

Engine No.	Fuel	P_0, MPa	T_0, K	D_C, mm	D_a, mm
1	Liquid	0.3	3100	3.5	10
2	''	0.6	3150	5.0	30
3	''	0.6	3150	5.0	30
4	''	0.6	3150	7.0	40
5	''	0.6–0.7	3150	10	13
6	''	0.6–0.7	3150	12.1	87
7	Solid	4.0	3300	18	40
8	''	5.6	2300	29	60
9	''	3.0–4.5	3480	36–39	90
10	''	5.7	3600	225	1575
11	''	3.2	3400	250	1125

MEASUREMENT RESULTS AND DISCUSSION

Measurements were made on a number of engines with different types of fuel, combustion chamber pressures P_0 and temperatures T_0, and diameters of the nozzle throat, D_C, and exit section, D_a (Table I).

TABLE II.

Engine No.	Experiment I_a, μA	Experiment $I_a/\pi D_a$, μA/m	J_w, μA/m absorbing wall	J_w, μA/m reflecting wall
1	0−0.6	0−20	…	…
5	0−5	0−120	…	…
6	0−9	0−50*	…	…
7	50−200	470−1880	590	2000
9	100−700	350−2450	590	2000
10	750−4500	127−765	120	410
11	500−750	175−260*	330	450

The measured currents I_a for $l_a/D_a \leqslant 0.2$ and $R_n = 0$ are listed in Table II.

An asterisk * denotes flows in a nozzle with separation at a cross section $D_{\text{eff}} < D_a$. In this case, D_a was replaced by the full-expansion value D_{eff}. Significant fluctuations in the current I_a were observed in all the engines. (The character and possible causes of the current fluctuations are discussed in Ref. 3.) This is reflected in Table II by indicating the typical range of the fluctuations. In the experiments on engines Nos. 7 and 9–11, including those with arbitrary values of l_a/D_a, the current I_a was positive (i.e., the engines were charged negatively) and polarity reversal of the current was not observed; this justified assuming that the wall current J_w makes a significant contribution to engine charging. The validity of this assumption is confirmed by estimates of J_w in accordance with Ref. 3 that are in qualitative agreement with the range of the experimental data on I_a. In the case of engines Nos. 1–6 and 8, the sign of I_a was not constant and could vary in successive experiments. Furthermore, of three engines (Nos. 2–4) with an identical design, one was charged predominantly positively, and the other two, negatively. This is illustrated by Fig. 2, which shows the electrical charge $q(t) = -\int_0^t I_a \cdot dt$ leaking through the measurement circuit from the engines as a function of time for several successive experiments under identical conditions for $l_a/D_a > 50$ and $R_n = 0$. One of the possible reasons for the change in sign is the observed dependence of the current on the excess oxidant coefficient α in liquid-fuel engines[5] (the specified value of α for engines Nos. 1–6 was 0.6–0.8). The loss of structural material from the engine had a significant effect on the current I_a but no change in sign was observed. Figure 3 shows averaged (over 5–10 experiments) plots of $q(t)$ for throat bushings made from three different materials (Fluoroplastic-4 (Teflon), molybdenum, and graphite) with different erosion rates: 105, 62, and ~0 g/s. The fact that I_a decreases as more material is removed has a simple physical explanation: the material that is carried away (in the gaseous and condensed phases) reduces the electron density in the combustion products because of a lowering of the temperature in the boundary layer on account of the screening effect

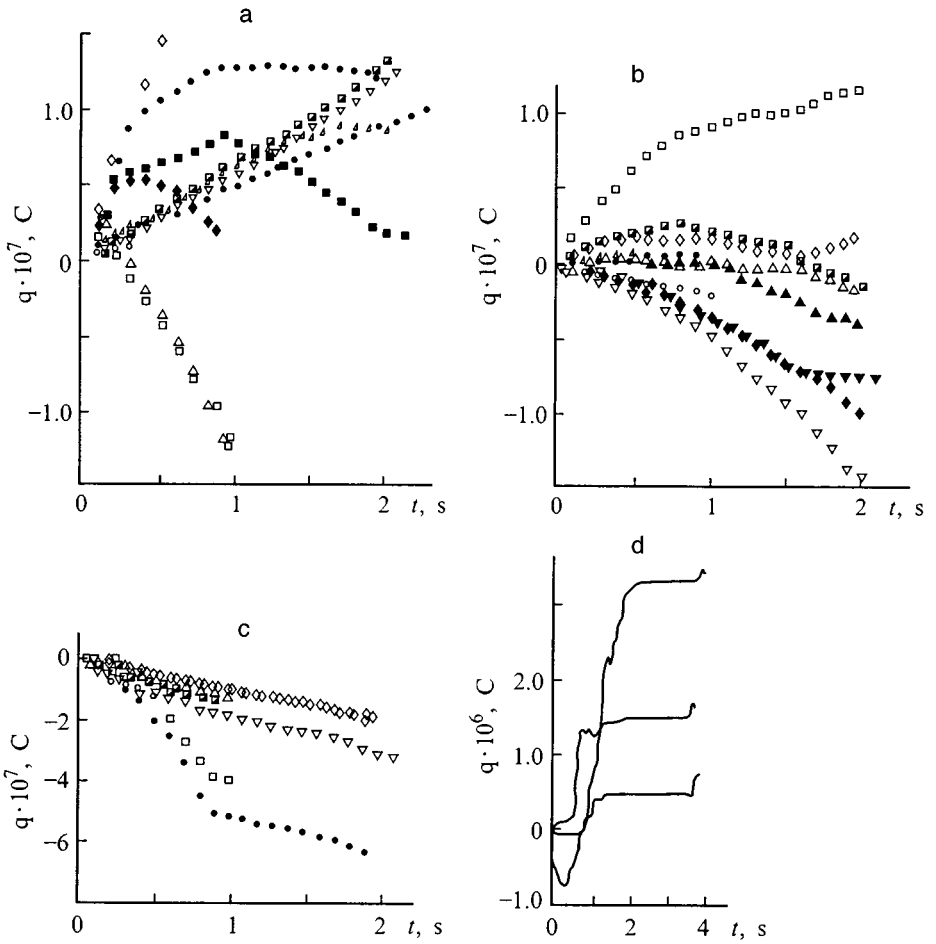

FIG. 2. $q(t)$ for engines No. 2 (a), 3 (b), 4 (c) and 8 (d).

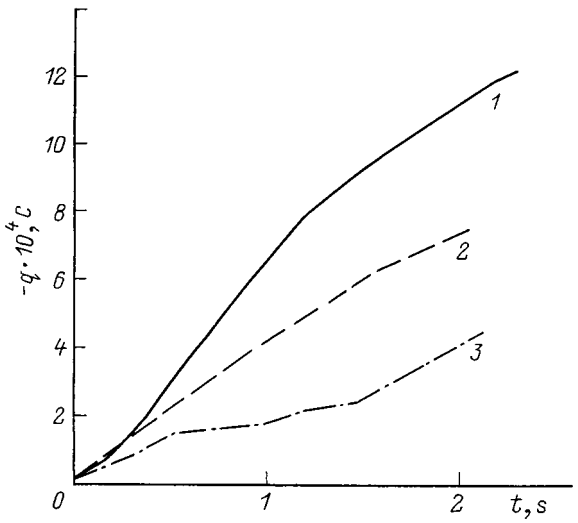

FIG. 3. $q(t)$ for $l_a/D_a \approx 0.1$ (engine No. 9): *1* — graphite, *2* — molybdenum, *3* — teflon.

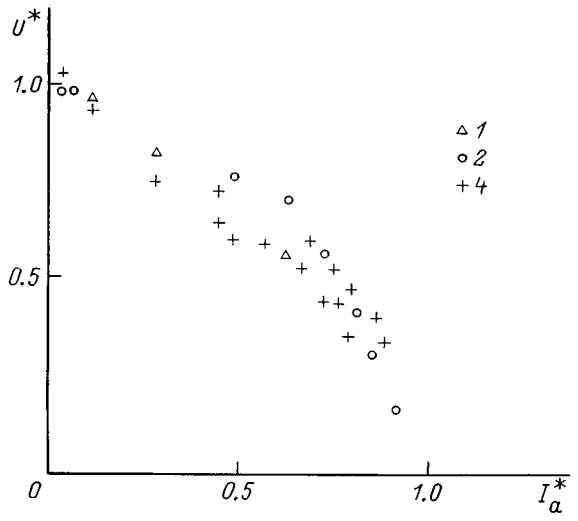

FIG. 5. The current I_a as a function of the electrical potential of the engine: $l_a/D_a \approx 50$ (*1*), 15 (*2*), 12.5–16.5 (*4*). All the other parameters correspond to Fig. 4.

and because of absorption of part of the electrons by macroscopic particles.[9]

The experimental data on I_a as a function of l_a for $R_n=0$ on engine No. 9 and in model charged streams (a portion of these data has been published previously by the author[3]) are generalized in Fig. 4. For dimensionless coordinates we chose $l'_a=l_a/l_n$, where l_n is the length of the initial segment of the stream,[10] and $I'_a=0.735I_a(l'_n)/I_a(2)$ (Ref. 3). Despite substantial differences in the experimental conditions, in these coordinates the data are grouped within a fairly narrow corridor, which suggests a universal character for the function $I'_a(l'_a)$. That is, the value of $I_a(l_a)$ required for particular conditions may be determined roughly if the value of I_a is known for some fixed value of l_a. The relationship between the current I_a and the engine potential (the current–voltage characteristic) was determined by varying the load resistance (Fig. 1) over $0 \leqslant R_n \leqslant \infty$. The resulting data are plotted in the space of coordinates $I^*_a=I_a/I_a(R_n=0)$, $U^*=U/U(R_n=\infty)$ in Fig. 5. Based on these data, it is possible to construct an approximate current–voltage characteristic in dimensional variables if two points are known, such as the current at $R_n=0$ and the potential at $R_n \to \infty$.

The effect of the pressure of the surrounding medium (the flight altitude) on E_a and, therefore, on I_a, has been studied in small-sized engines mounted in a pressure vessel. The results are shown in Fig. 6. At pressures of 10^2-10^3 Pa, the current is low, but beyond this range, the current increases. The pressure dependence of I_a resembles the well known Paschen curve of gas discharge theory.

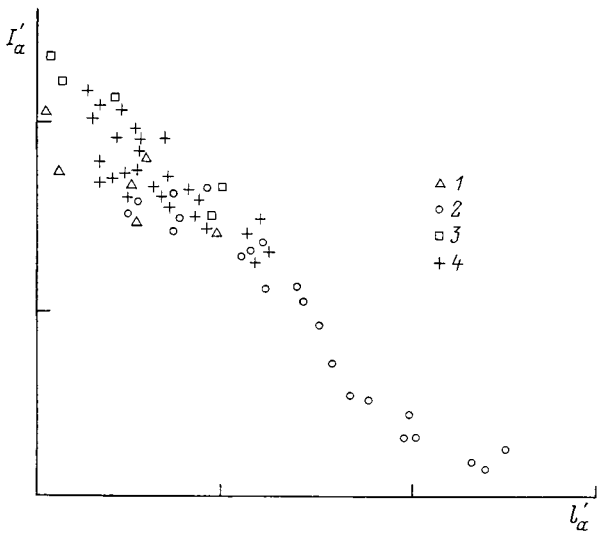

FIG. 4. Current in the gas I_a as a function of the length of the space-charge zone: *1* — engine No. 9, degree of underexpansion[10] $n=1-1.5$, degree of heating[10] $\Theta=7$; *2* — water vapor; $P_0=0.7$, 1.3 MPa; $P_a=0.1$ MPa; $D_c=2.0$, 1.1 mm, $D_a=2.35$, 1.3 mm, $n=1.3$, 2.45, $\Theta=1.35$, 1 . . . 45; *3* — silicone liquid vapor; $P_0=0.28$ MPa; $P_a=0.07$ MPa; $D_c=D_a=2$ mm; $n=2.3$; $\Theta=1.1$; *4* — organofluoric liquid vapor; $P_0=0.12$ MPa; $P_a=0.03$ MPa; $D_c=2.0$ mm; $D_a=2.0$, 3.0 mm; $n=2.4$, 0.5; $\Theta=1.05$, 0.95.

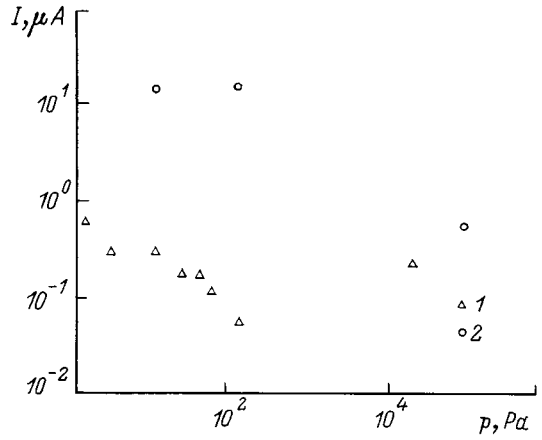

FIG. 6. The current I_a as a function of ambient pressure: *1* — engines Nos. 2, 3, $l_a/D_a \approx 50$; *2* — engine No. 6, $l_a/D_a \leqslant 0.2$.

CONCLUSION

These experimental studies have yielded measurements of the electrical current I_a as $l_a/D_a \to 0$, which represents the maximum possible current I_m at the nozzle exit of an engine and is a parameter that depends only on the engine characteristics, and a determination of the current I_a as a function of the three major parameters which affect it: the length l_a of the space charge region, the electrical potential U of the engine, and the pressure p of the surrounding medium. These results can be applied to practical problems (for estimating the charging current $I_d = -I_a$ and the rate of rise and maximum value of the potential on a flight vehicle) if I_m, l_a, U, and p are known for the particular conditions. The value of l_a will depend on a number of factors: the charged particle mobility, the degree of ionization of the combustion products, the parameter of the surrounding medium, etc., and finding it is an independent problem (according to estimates, at an altitude of 400–500 km, $4-6 < l_a/D_a < L/D_a$, where L is the characteristic geometric size of the flight vehicle). The rate of rise of the potential of a flight vehicle when the engine is turned on is

$$\left[\frac{dU}{dt}\right]_{U=U_0} = \frac{I_d}{C_1+C_2},$$

where C_1 and C_2 are the capacitance of the flight vehicle and of the vehicle–earth system.

The maximum potential can be estimated from the condition that the charging and discharge currents are balanced at the point where the corresponding current–voltage characteristics intersect.

The author thanks P. P. Andreev, V. A. Nikol'skiĭ, and S. G. Rebrov for assistance in doing the experiments.

[1] A. B. Vatazhin, V. I. Grabovskiĭ, V. A. Likhter, and V. I. Shul'gin, *Electrogasdynamic Flows* [in Russian], Nauka, Moscow (1983), 344 pp.
[2] L. Aronowitz, IEEE Trans. Electromagn. Compat. **EMC-10**, 341 (1968).
[3] Yu. A. Nagel', Teplofiz. Vys. Temp. **25**, No. 6 (1987); Deposited 8.06.87, No. 4115-V-87.
[4] V. A. Pinchuk, Zh. Tekh. Fiz. **67**(8), 21 (1997) [Tech. Phys. **42**, 872 (1997)].
[5] V. A. Shkalikov, in *Thermal Processes and Properties of the Working Medium of Aircraft Engines* [in Russian], edited by V. E. Alemasov, Kazan Aeronautical Institute, Kazan (1980), pp. 41–47.
[6] V. A. Shkalikov, A. A. Lebedkov, and Zh. Kul'zhanov, in *Thermal Processes and Properties of the Working Medium of Aircraft Engines*, edited by V. E. Alemasov, Kazan Aeronautical Institute, Kazan (1982), pp. 79–83.
[7] A. Brunet and F. Faure, La Recherhe Aerospatiale, No. 1, pp. 29–38 (1990).
[8] Yu. A. Nagel', Zh. Prikl. Mekh. Tekh. Fiz., pp. 24–32 (1971).
[9] S. Sou, *Hydrodynamics of Multiphase Systems*, Mir, Moscow (1971), 536 pp.
[10] G. N. Abramovich, *Theory of Turbulent Jets* [in Russian], GITTL, Moscow (1960), 715 pp.

Translated by D. H. McNeill

Transient spectroscopy of surface states in a constant subthreshold current mode for MIS transistors

M. N. Levin, E. N. Bormontov, A. V. Tatarintsev, and V. R. Gitlin

Voronezh State University, 394693 Voronezh, Russia
(Submitted May 27, 1998)
Zh. Tekh. Fiz. **69**, 60–64 (August 1999)

The spectrum of surface states at a semiconductor–insulator interface is studied through the relaxation of the gate voltage of an MIS transistor measured with a constant subthreshold current. The proposed method makes permits study of these states in both halves of the semiconductor gap and is convenient for testing integrated circuits. The possibilities of this method are illustrated using the example of radiation induced changes in the distribution of surface states. © *1999 American Institute of Physics.* [S1063-7842(99)01108-3]

The analysis of relaxation processes resulting from the refilling of nonequilibrium surface states is an effective means of studying insulator–semiconductor interfaces. Simmons and Wei[1] were the first to point out the possibility of directly determining the population of surface states from the transient currents in a metal–insulator—semiconductor (MIS) structure during switching from enrichment to deep depletion. They neglected the effect of the change in the depth of the space-charge region of the semiconductor and limited the applicability of the method by the condition $C_i \gg C_{sc}$, where C_i and C_{sc} are the capacitances of the insulator and semiconductor, respectively. This limitation was eliminated by Johnson's method of constant-capacitance transient spectroscopy of surface states (CC-DLTS).[2] This mode ensures a constant depth of the space-charge region during relaxation and eliminates the effect of the semiconductor on the measurements. The constant-capacitance method makes it possible to study surface states in the half of the gap adjacent to the majority carrier band. Difficulties arise in the middle of the gap because surface currents develop owing to the uncontrolled accumulation of minority carriers at the surface.[3] In the case of measurements on MIS transistors, it is also possible to attain extreme sensitivity in the transient spectroscopic methods by using their intrinsic amplifying effect. Note that, in studies of the transient refilling of surface states in the MIS elements of large integrated circuits with small dimensions, a constant depth of the space-charge region is most easily attained by keeping the drain current constant, rather than the gate capacitance. The relaxation of the gate voltage V_g of an open MIS transistor, whose magnitude exceeded that of the threshold voltage V_T, has been measured[4] in this mode (with a constant drain voltage). It was possible to detect the refilling of discrete surface states in an MIS transistor with dimensions of 1×1 μm.[4] However, by keeping the drain current constant in an open transistor, it was possible to study the surface states only near the minority carrier band.

In this paper we study the spectrum of surface states through the relaxation of the gate voltage of an MIS transistor in a constant subthreshold current mode. This variant makes it possible to study surface states in both halves of the semiconductor gap.

THEORETICAL PART

For concreteness, we shall consider a *p*-channel MIS transistor. In these transistors, the spectrum of surface states in the upper half of the gap (adjacent to the majority carrier band of the semiconducting substrate of the transistor) is studied through the relaxation of $V_g(t)$ as it is switched from enrichment into a state at the start of a weak inversion. In the state at the start of an inversion, the Fermi quasilevel of the minority carriers, E_{Fp}, at the surface lies near the middle of the gap and the surface band bending ψ_s equals the bulk potential φ_B of the semiconductor, while the corresponding voltage at the gate of the transistor, V_{mg}, can be referred to as the ''midgap'' voltage. Surface states in the lower half of the gap (adjacent to the minority carrier band) are studied using the relaxation of $V_g(t)$ during switching of an MIS transistor from strong inversion to a state in the ''midgap.'' The drain voltage should be high enough to ensure saturation of the drain current. When the current is saturated with respect to the drain voltage in the weak inversion region, the semiconductor surface is an equipotential over essentially the entire channel length.[5] This means that the charge relaxation process in the surface states can be examined in a one- dimensional approximation.

When an MIS transistor is switched from enrichment to inversion, two processes refill the surface states: emission of electrons from the surface states into the conduction band and capture (recombination) of free holes at the surface states. The kinetics of the charge variation at the surface states, $Q_{it}(t)$, obeys the equation

$$Q_{it}(t)=q\int_{E_{Fp}}^{E(t)} D_{it}(E)\,\exp[-(e_n+c_p)t]dE, \tag{1}$$

where q is the charge of an electron, $D_{it}(E)$ is the density of surface states, e_n and c_p are the coefficients of electron emission and hole capture (recombination), respectively, E_{Fp} is

the location of the Fermi quasilevel for the minority carriers in the semiconductor, and $E(t)$ is the energy level of the surface states at time t.

The coefficient of electron emission from surface states is determined by their energy,

$$e_n = \sigma_n v_{\text{th}} n_i \exp\left(\frac{E-E_i}{kT}\right), \tag{2}$$

while the capture coefficient is determined by the free hole concentration and is given in terms of the Fermi quasilevel E_{Fp} by

$$c_p = \sigma_p v_{\text{th}} n_i \exp\left(\frac{E_i-E_{Fp}}{kT}\right). \tag{3}$$

Here n_i is the intrinsic carrier concentration in the semiconductor, v_{th} is the thermal speed, σ_n and σ_p are the surface state cross sections for capture of electrons and holes, respectively, E_i is the Fermi level in the intrinsic semiconductor, k is Boltzmann's constant, and T is the absolute temperature. Equating the emission and recombination coefficients determines the location of the demarcation level E_D above which the surface states are mainly emptied by emission and below which they are mainly discharged by recombination,

$$E_D - E_i = E_i - E_{Fp} + kT \ln(\sigma_p/\sigma_n). \tag{4}$$

Electron are emitted into the conduction band successively as the depth of the energy of the surface states increases. Here the energy level of the surface states from which electron are emitted at time t is given by the well known expression[1]

$$E(t) = E_i - kT \ln(\sigma_n v_{\text{th}} n_i t). \tag{5}$$

The time t is reckoned from the time the MIS transistor is switched from enrichment to weak inversion. Equations (3)–(5) imply that the time t_D at which the electron emission reaches surface states at the demarcation energy E_D is inversely proportional to the hole capture coefficient, i.e.,

$$t_D = c_p^{-1}. \tag{6}$$

Given Eqs. (3)–(6), Eq. (1) for the variation in the charge in the surface states can be written in the form

$$Q_{it}(t) = q \exp\left(-\frac{t}{t_D}\right) \int_{E_{Fp}}^{E(t)} D_{it}(E)\,dE. \tag{7}$$

When a constant drain current is maintained during relaxation of an MIS transistor, the variation in the gate voltage is determined only by refilling of the surface states, i.e., by the kinetics of the charge $Q_{it}(t)$, i.e.,

$$V_g(t) = V_{g0} - \frac{Q_{it}(t)}{C_i}, \tag{8}$$

where V_{g0} is the gate voltage which at $t=0$ ensures flow of the chosen level of constant current under conditions of fixed enrichment of the transistor channel; $C_i = \varepsilon_i/d_i$ is the geometric volume of the subgate insulator layer, ε_i is the absoulte permittivity of the insulator, and d_i is its thickness.

Equations (7) and (8) can be used to express the density of surface states in terms of the gate voltage $V_g(t)$ and its time derivative $dV_g(t)/dt$ at time t, which are related by Eq. (5) to the energy of the surface states,

$$D_{it}(t) = \frac{C_i}{qkT} \exp\left(\frac{t}{t_D}\right)\left[\frac{V_{g0}-V_g(t)}{t_D} - \frac{dV_g(t)}{dt}\right] t. \tag{9}$$

The system of Eqs. (5) and (9) makes it possible to determine the energy distribution of the density of surface states in the upper half of the gap from the relaxation behavior of the gate voltage measured with a constant subthreshold current. The upper and lower limits of the energy interval that can be studied are determined according to Eq. (5) by the minimal measurement time and the demarcation energy E_D, respectively.

Practical realization of this method requires an initial determination of the demarcation time t_D and the capture cross section for the majority carriers, σ_n. Note that with Eq. (7), Eq. (8) can be rewritten in the form

$$V_g(t) = V_{g\infty} - \frac{q}{C_i} \exp\left(-\frac{t}{t_D}\right) \int_{E_i}^{E(t)} D_{it}(E)\,dE, \tag{10}$$

where $V_{g\infty}$ is the final steady state value of the gate voltage $V_g(t\to\infty)$, which is given by

$$V_{g\infty} = V_{g0} - \frac{q}{C_i} \int_{E_i}^{E_{Fp}} D_{it}(E)\,dE. \tag{11}$$

Taking the logarithm of Eq. (10), we obtain

$$\ln\left[\frac{C_i(V_{g\infty}-V_g(t))}{q}\right] = -\frac{t}{t_D} + \ln \int_{E_i}^{E(t)} D_{it}(E)\,dE. \tag{12}$$

On plotting the relaxation curve in the coordinates

$$\ln\left[\frac{C_i(V_{g\infty}-V_g(t))}{q}\right] = f(t),$$

we can determine t_D from the slope of the curve. Then, by numerically differentiating the relaxation curve $V_g(t)$, for each fixed temperature Eq. (9) yields D_{it} for times $t < t_D$.

In order to find the majority carrier capture cross section σ_n from these $D_{it}(t)$ curves, one determines the corresponding values of the temperature and time satisfying the condition $D_{it} = \text{const}$ (see the inset to Fig. 2), i.e., the points t_1 and t_2 for the curves taken at T_1 and T_2, respectively. Using the resulting values of t_i and T_i, one can construct the curve $\ln(v_{\text{th}} n_i t) = f(1/kT)$, which is a straight line, provided that $\sigma_n = \text{const}$, and from the slope of this line, determine the capture cross section σ_n. The lower portion of the gap is studied by switching the MIS transistor from a strong inversion region into a state in the ''midgap.'' The ''midgap'' voltage V_{mg} was determined from the theoretical value of the subthreshold current I_{mg} by a standard method.[6] Refilling of the surface states in this case is determined by a single mechanism — emission of holes from the surface states into the valence band. Equation (9) transforms to the simple form

$$D_{it}(t) = -\frac{C_i(t)}{qkT} \frac{dV_g(t)}{dt}. \tag{13}$$

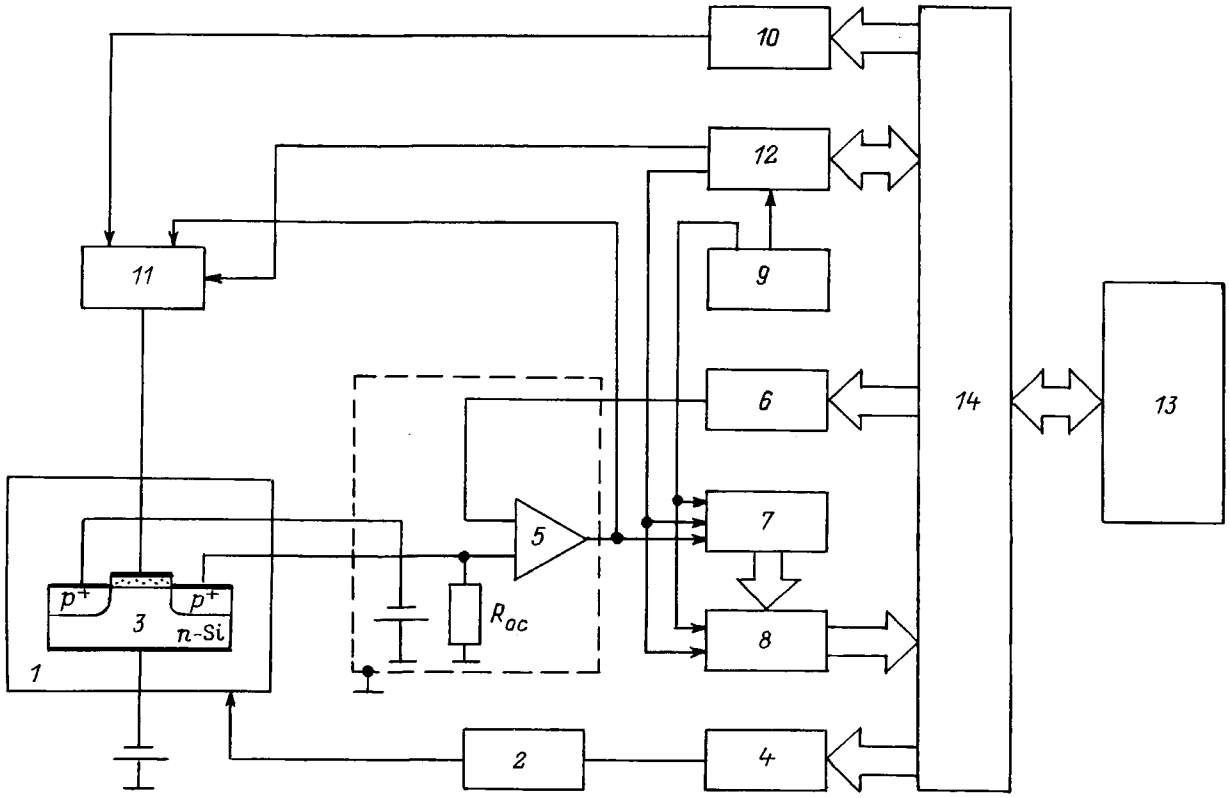

FIG. 1. Block diagram of the apparatus for transient spectroscopy of surface states in the subthreshold saturation current mode.

Here the energy and time are related by the formula[1]

$$E(t)=E_i+kT\ln(\sigma_p v_{\text{th}} n_i t). \tag{14}$$

Note also that, in order to determine the capture cross section σ_p for minority carriers, only one relaxation curve, measured at a fixed temperature, is needed. Thus, the system of Eqs. (13) and (14) yields a spectrum of the surface states in the lower half of the gap of the semiconductor.

MEASUREMENT APPARATUS

Figure 1 shows a block diagram of the apparatus for transient spectroscopy of the surface states with saturation of the subthreshold current. The apparatus includes the following functional units:

1) A vacuum cryostat *1* with a temperature control unit *2* that regulates the temperature of the sample *3* between 80 and 723 K within 0.5 K. The reference voltage for this unit is provided by a TsAP1 digital-to-analog converter *4* based on a KR594PA1 integrated circuit.

2) A current generator based on a K574UD2 fast operational amplifier *5* which provides dynamic maintenance of a constant subthreshold current. The level of the current is determined by the reference voltage provided by a TsAP2 digital-to-analog converter *6* based on a KR594PA1 integrated circuit and can reach 10^{-10} A.

3) A fast 10-channel analog-to-digital converter (ADC) *7* assembled from 8-channel K1107PV4 ADCs with an access time of 10^{-8} s and used to measure the relaxation voltage at the gate of the MIS transistor owing to refilling of the surface states.

4) A buffer (buffer RAM) *8* for preliminary recording of the data from the ADC and sending it on to the computer. It has a capacity of 10×1024 bits and is based on a K1500RU415 static, emitter-coupled logic RAM device with a minimum access time of 2×10^{-8} s.

5) A timing pulse generator *9* made up of discrete emitter-coupled logic devices and used to strobe the ADC and buffer.

6) A TsAP3 digital-to-analog converter *10* based on a KR594PA1 integrated circuit, used to specify the parameters of the bias pulses delivered to the gate of the MIS transistor.

7) A fast electronic switch *11* for switching the signals delivered to the gate of the MIS transistor.

8) A programmable timer *12* based on a KR580VI53 integrated circuit for producing the timing diagram for the measurement process (specifying the temporal parameters of the bias pulse, controlling the switching device, the ADCs and the buffer).

9) An Élektronika-60 microcomputer *13* for controlling the measurement process and processing the data.

10) A matching device *14* for coupling the measurement unit to the microcomputer.

The apparatus works as follows. When the specified temperature is reached, a voltage pulse is delivered to the gate of the transistor and converts the surface in the subgate region into an enriched or inverted state (depending on the half of the semiconductor gap being studied). Here the switching device disconnects the feedback circuit to ensure maintenance of a constant subthreshold current. After a steady state is reached (no more than 0.01 s), the bias voltage

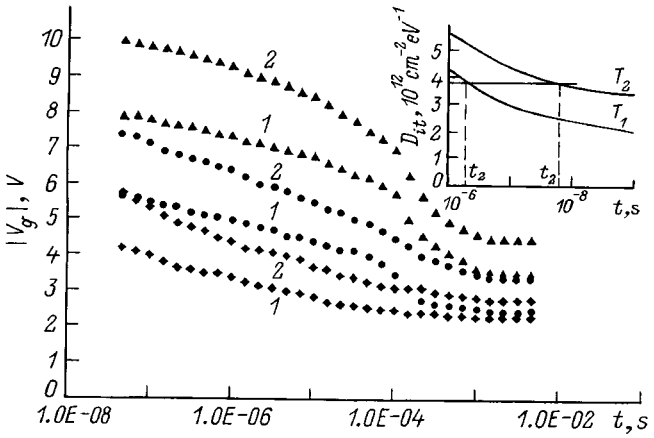

FIG. 2. Relaxation curves $V_g(t)$, measured during switching of an MIS transistor into a weak inversion from enrichment (● — T=300 K, ▲ — $T=250$ *K*) *from a strong inversion* (♦ — $T=300$) *after x-ray irradiation*: $D=10^5$ (*1*), 5×10^5 R (*2*).

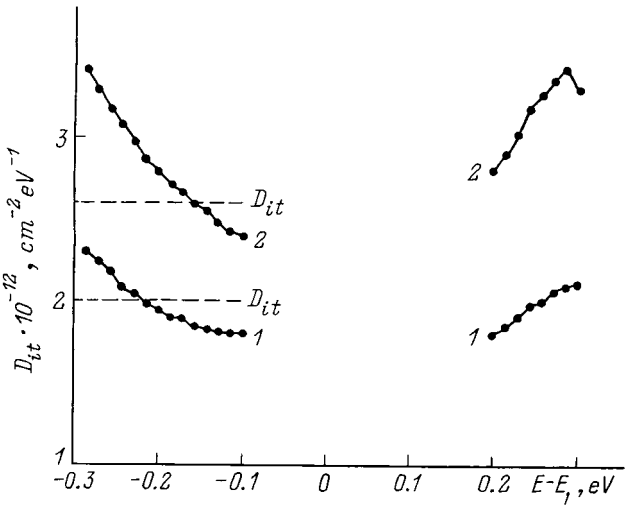

FIG. 3. Energy spectra of the density of surface states after x-ray irradiation (labels *1* and *2* are as in Fig. 2).

is removed and the feedback circuit is turned on to renew and maintain a specified level of the subthreshold current by varying the gate voltage on the MIS transistor. After the end of the transition process (no more than 5×10^{-8} s), which is determined by the speed of the operational amplifier in the current generator, the fast ADC begins to measure the relaxation of the gate voltage on the MIS transistor as the surface states are refilled and the data are then recorded in the buffer. The sampling time of the ADC can be varied from 2×10^{-8} to 10^{-3} s in various parts of the relaxation curve. When the buffer is full or the measurement of the curve is complete, the data are read out by the computer. Then the measurement is repeated at the same or a different temperature. The time required to measure a single relaxation curve is less than 0.02 s when the data are recorded in the computer memory and less than 0.1 s when the data are recorded on a disk.

EXPERIMENTAL RESULTS

The proposed method was used to determine the radiation induced changes in the spectrum of surface states. The measurements were made on the *p*-channel of an MIS transistor with a subgate oxide layer thickness $d_{0x}=98$ nm, channel length $L=3$ μm, channel width $W=100$ μm, and doping concentration in the substrate $n_0=10^{14}$ cm^{-3}. The energy spectrum of the surface states was changed by irradiation with soft x rays with an energy $E_x=20$ keV at doses of up to 10^6 R followed by low temperature annealing at 723 K. The samples were irradiated on an IRIS-M3 x-ray system.

Typical relaxation curves $V_g(t)$ for these MIS transistors, measured with a saturation subthreshold current after x-ray bombardment at doses of 10^5 R and 5×10^5 R, are shown in Fig. 2. The energy distributions of the surface states, $D_{it}(E)$, calculated from these experimental data are shown in Fig. 3.

The surface state spectra calculated in this method can be compared with data obtained from these same samples by the standard stationary subthreshold current method.[7] The steady-state current–voltage characteristics were measured on the apparatus described in Ref. 8. Recall that the stationary drain–gate current–voltage characteristic of a long-channel MIS transistor in the subthreshold region is given by the equation[7]

$$I_D=I_0\exp\left(\frac{V_g-V_T}{S}\right), \tag{15}$$

where

$$S=\frac{kT}{q}\,\frac{C_i+C_{sc}+C_{ss}}{C_i}$$

and C_{ss} is the capacitance due to refilling of the surface states.

The changes in the current–voltage characteristics caused by the x-ray irradiation are shown in Fig. 4.

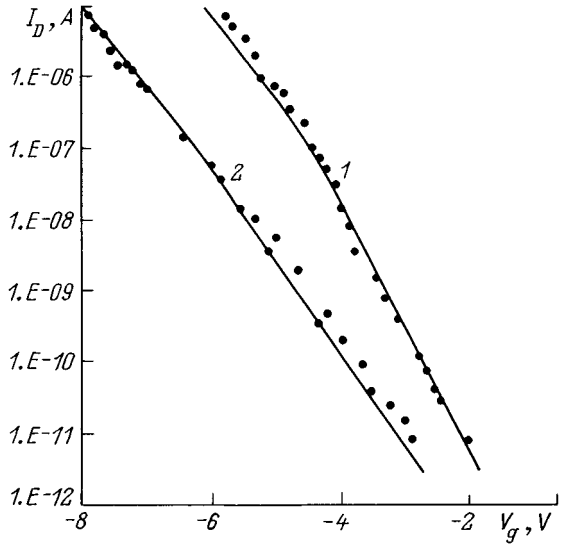

FIG. 4. Subthreshold portions of the drain–gate current–voltage characteristics of *p*-channel transistors after x-ray irradiation (labels *1* and *2* are as in Fig. 2).

The average density of surface states, D_{it}, was calculated from the parameter S, which is determined experimentally from the slope of a plot of $\ln(I_D/I_0)$ versus V_g:

$$D_{it}=\frac{1}{q}\left[C_i\left(\frac{qS}{kT}-1\right)-C_{sc}\right]. \tag{16}$$

The averaged values of the density D_{it} of surface states determined from the curves of Fig. 4 using Eq. (16) are shown in Fig. 3 by the dashed lines and equal 2×10^{12} and $2.6\times10^{12}\,\mathrm{cm}^{-2}\mathrm{eV}^{-1}$ for irradiation doses of 10^5 and 5×10^5 R, respectively. These are in good agreement with the results obtained by the method proposed here.

To summarize, using the relaxation of the gate voltage of an MIS transistor in the constant subthreshold current mode for studying the spectrum of the surface states at a insulator–semiconductor interface offers the possibility of studying the surface states in both halves of the semiconductor gap and is convenient for testing of integrated circuits.

[1] J. G. Simmons and L. S. Wei, Solid-State Electron. **17**, 117 (1974).
[2] N. M. Johnson, J. Vac. Sci. Technol. **21**, 303 (1982).
[3] K. L. Wang, IEEE Trans. Electron Devices **ED-27**, 22311 (1980).
[4] A. Karwath and M. Schulz, Appl. Phys. Lett. **52**, 634 (1988).
[5] R. J. Van Overstraeten, G. J. Declerck, and P. A. Muls, IEEE Trans. Electron Devices **ED-22**, 282 (1975).
[6] P. J. Ms Whorter and P. S. Winokur, Appl. Phys. Lett. **48**, 133 (1986).
[7] S. Sze, *Physics of Semiconductor Devices*, 2nd ed. [Wiley, New York (1981); Mir, Moscow (1984), Vol. 2; 456 pp.].
[8] M. N. Levin, V. L. Izraĭlit, A. V. Tatarintsev, and S. G. Kadmenskiĭ, Prib. Tekh. Eksp. No. **2**, 119 (1992).

Translated by D. H. McNeill

Optimization of a differential absorption and scattering lidar for sensing molecular hydrogen in the atmosphere

V. E. Privalov and V. G. Shemanin

Baltic State Technical University, 195008 St. Petersburg, Russia
(Submitted June 30, 1998)
Zh. Tekh. Fiz. **69**, 65–68 (August 1999)

The lidar equation for remote sensing of molecular hydrogen in the atmosphere by differential absorption and scattering is solved to optimize the lidar system. The background conditions are taken into account.
[S1063-7842(99)01208-8]

The widespread use of differential absorption and scattering lidars in the infrared to probe for atmospheric gas molecules suggests that they might be also used for remote sensing of molecular hydrogen in the atmosphere.[1] The differential absorption and scattering method is highly sensitive because, of all the known spectroscopic effects, resonant absorption has the largest cross section in the visible and IR.[2] In a differential absorption and scattering lidar of this type, probing is done at two wavelengths, one of which coincides with a maximum in the absorption band of the molecules being studied while the other lies outside this band. Information on the distribution of the concentration of the molecules is extracted by comparing the detected lidar signals at the two wavelengths within a sufficiently narrow spectral interval.

There is, therefore, some interest in a numerical solution of the lidar equation for differential absorption and scattering detection of H_2 molecules in order to choose an optimum variant of the lidar system.

We write the lidar equation for elastic backscatter in the form[3]

$$P(\lambda,R)=P_0K_1A_2T^2(\lambda_0)\rho/R^2, \tag{1}$$

where $P(\lambda,R)$ is the backscatter signal power at wavelength λ_0 arriving at the detector rom a distance R; $P_0(\lambda_0)$ is the laser power at its wavelength; K_1 is the lidar constant; A_2 is the area of the receiver telescope; $T(\lambda_0)$ is the transmission of the atmosphere at the wavelength of the laser light and of the backscatter signal; and, ρ is the reflection coefficient of the topographic target or the combined scattering coefficient for elastic Mie and molecular Rayleigh scattering.

All the information on the molecular hydrogen concentration is contained in the factor $T(\lambda_0)$, which, in general, can be written in the form[4]

$$T(\lambda,R)=\exp\left[-\int_0^R k(\lambda)dR\right], \tag{2}$$

where the attenuation coefficient $k(\lambda,R)$ is given by an equation of the form

$$k(\lambda,R)=k_A(\lambda,R)+N(R)\sigma_0(\lambda), \tag{3}$$

where the first term is the atmospheric attenuation at the laser wavelength without the hydrogen molecules and the second is the product of the H_2 concentration and the resonance absorption cross section of these molecules.

Then, following the idea of differential absorption and scattering,[4] we take two lidar equations of the type (1) for two laser wavelengths λ_0 and λ_1, where the second wavelength lies outside the absorption band of H_2, and divide one by the other. Dividing yields an equation for the general case of differential absorption and scattering, where it is assumed that all the wavelength dependent factors are different:

$$\frac{P(\lambda_0,R)}{P(\lambda_1,R)}=\frac{P_0K_{10}\rho_0}{P_1K_{11}\rho_1}\times\exp\left\{-2\int_0^R[k(\lambda_0,R)-k(\lambda_1,R)]dR\right\}. \tag{4}$$

Substituting an equation of the type (3) for the attenuation coefficients at both wavelengths in Eq. (4), we finally write the differential absorption and scattering equation in the form

$$\int_0^R N(R)dR=\frac{1}{2\sigma_0}\ln\left[\frac{P(\lambda_1,R)P_0K_{10}\rho_0}{P(\lambda_0,R)P_1K_{11}\rho_1}\right]-\int_0^R[k_A(\lambda_0,R)-k_A(\lambda_1,R)dR]. \tag{5}$$

We now determine the quantities in Eq. (5) for our case. The wavelength at the maximum of the IR band for H_2 molecules is taken to be 2.4 μm. According to Ref. 5, a purely vibrational transition is forbidden in hydrogen, while rotational transitions in which j changes by ± 1 are allowed. However, it has been pointed out[6] that, because of external perturbations, forbidden transitions can become allowed (induced) as a result of deformations of diatomic molecules by collisions in the atmosphere. Data[6] on pressure-induced absorption by nitrogen molecules under atmospheric conditions in the neighborhood of the fundamental vibration at 4 μm yield an estimate for the nitrogen absorption cross section at 1 atm on the order of $10^{-24}\,\mathrm{cm}^2$. This is a minimum value, which does not include the intensity of the laser radiation,

TABLE I. Calculated power ratios of the differential absorption and scattering signals from molecular H_2 for ranging distances of $0.01-10.0$ km and molecular concentrations of $10^{10}-10^{16}\,cm^{-3}$.

N, cm^{-3}	10^{10}	10^{11}	10^{12}	10^{13}	10^{14}	10^{15}	10^{16}
R, m	Ratio of the differential absorption and scattering signal powers, $P(\lambda_1,R)/P(\lambda_0,R)$						
10	1.00007	1.00009	1.00027	1.00207	1.02027	1.22149	7.38959
100	1.00074	1.00092	1.00272	1.02094	1.22228	7.39438	
1000	1.00743	1.00924	1.02757	1.23023	7.44245		
2000	1.01491	1.01857	1.05591	1.51346			
3000	1.02245	1.02798	1.08502	1.8619			
4000	1.03004	1.03749	1.11494	2.29057			
5000	1.03769	1.04707	1.14568	2.81792			
6000	1.0454	1.05675	1.17727	3.46669			
7000	1.05317	1.06652	1.20973	4.26482			
8000	1.06099	1.07638	1.24309	5.2467			
9000	1.06887	1.08632	1.27737	6.45464			
10000	1.07681	1.09636	1.31259	7.94069			

which in our case is at least $1\,MW/cm^2$. Studies[7] of the absorption by hydrogen molecules of IR laser radiation at 10.6 μm have revealed the existence of IR absorption by hydrogen molecules and yield an estimate for the absorption cross section (an increase by 5 orders of magnitude at a power of 1 MW) on the order of $10^{-19}\,cm^2$. The IR absorption spectrum of molecular hydrogen induced by an electric field (up to 12 kV/cm) has been studied[8] by intracavity spectroscopy. In this paper we consider the case of absorption induced by external perturbations (pressure and high power laser radiation) in the neighborhood of 2.4 μm.

In addition, it is also possible to create an experimental situation in which a sum of rotational bands is detected by suitable choice of the transmission passband for the lidar monochromator.

The wavelength outside this band was chosen to be 2.1 μm, in order to fit within the transparency region of the atmosphere.[9] A lidar of this sort can be built using pulsed solid state lasers with rods of YAG:Cr:Er and YAG:Cr:Ho.[10,11]

Furthermore, for the specific case of our lidar, we separate a factor $\xi_p(\lambda)$ that depends on the spectral sensitivity of the detector from the lidar constant K_1, i.e.,

$$K_1 = K_2 \xi_p(\lambda). \tag{6}$$

The remaining factors in Eq. (1) have the following values: $A_2 = 0.008\,m^2$; $K_2 = 0.4$ for a wavelength of 1.06 μm; peak laser powers of $P_0 = 1$, 10, and 100 kW with the ratio of the powers of the laser light at the two chosen wavelengths equal to the reciprocal of the spectral sensitivities of the detectors at the same wavelengths; ranging distances $R = 0.01$, 0.1, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0 and 10.0 km; concentrations of $10^{10}-10^{16}\,cm^{-3}$ for the molecules being studied; spectral sensitivities of the LFD-2 avalanche photodiodes at the chosen wavelengths[12] equal to 0.1 and 0.05 of the maximum at a wavelength of 1.4 μm; attenuation coefficients[9] k_A equal to 0.0314 and $0.035\,km^{-1}$, respectively, for the wavelengths of interest to us; resonant absorption cross section of the hydrogen molecule[4] $\sigma_0 = 0.8 \times 10^{-18}\,cm^2$; reflection coefficients of the topographic targets estimated from the data of Ref. 3 to be 0.3 for a corner reflector and 0.15 for a matte surface; and, a total atmospheric scattering coefficient[4] of 10^{-7}.

The data listed above have been used in numerical calculations of the ratio of the differential absorption and scattering signal powers according to Eq. (5) for the chosen concentrations of hydrogen molecules and wavelengths and for ranging distances of 0.01 to 10.0 km in order to search for the optimum variant of the lidar system. The results of the calculations for molecular H_2 are listed in Table I. The table shows that this ratio is a minimum for very low concentrations and distances and is excessive for high concentrations at any distance. The calculations were not done where the ratio exceeded 10, since the dynamic range of the detector system was taken to be 10^4, in accordance with Refs. 12 and 13. For a distance of 10 m, the range of accessible concentrations is $10^{12}-10^{16}\,cm^{-3}$ and for a distance of 10 km, $10^8-10^{13}\,cm^{-3}$ for the integrated value over the entire beam path. These results show that the differential absorption and scattering method, like ordinary IR absorption,[13] has lower and upper bounds on the range of possible products of the concentration and layer thickness, which are determined by the optical design and detector of the lidar. The first row of Table I lists signal ratios for a 10 m path, which are essentially the same as those for a similar path length of 10 m located at distances of from 100 m to 10 km from the lidar. This means that a lidar of this sort can determine molecular H_2 densities of $10^{12}-10^{16}\,cm^{-3}$ with a spatial resolution of 10 m along a path of up to 10 km in length.

Later, the power of the lidar signal was calculated using Eq. (1) for different experimental situations for a hydrogen concentration integrated over the entire path. The results are listed in Table II. This table implies that increasing the laser power leads to a proportionate increase in the signal power, and a reduction in the concentration of the molecules also leads to an increase in the power without a change in the spectral dependence of the factors in Eq. (1). As the distance is increased, the signal powers decrease, which makes it impossible to probe high concentrations at large distances. The signal power for a segment of the beam path 10 m long

TABLE II. Calculated power of the differential absorption and scattering signals from molecular H_2 for laser powers of $1-100$ kW, molecular densities of $10^{10}-10^{16}$ cm^{-3} integrated over the path, ranging distances of $0.01-10.0$ km and different scattering targets.

		N_a, cm^{-3}	10^{16}	10^{14}	10^{12}	10^{10}
R, m	ρ	P, kW	Backward differential absorption and scattering signal power, W			
10	0.15	100	1.24×10^{-8}	4.91	5.99	6.0
		10	1.24×10^{-9}	0.49	0.599	0.6
		1	1.24×10^{-10}	0.049	0.0599	0.06
	0.3	100	2.47×10^{-8}	9.82	11.98	12.0
		10	2.47×10^{-9}	0.98	1.198	1.2
		1	2.47×10^{-10}	0.098	0.01198	0.12
	10^{-7}	100		3.28×10^{-6}	3.99×10^{-6}	4×10^{-6}
		10		3.28×10^{-7}	3.99×10^{-7}	4×10^{-7}
		1		3.28×10^{-8}	3.99×10^{-8}	4×10^{-8}
100	0.15	100		8.1×10^{-3}	5.88×10^{-2}	6×10^{-2}
		10		8.1×10^{-4}	5.88×10^{-3}	6×10^{-3}
		1		8.1×10^{-5}	5.88×10^{-3}	6×10^{-3}
	0.3	100		1.62×10^{-2}	0.117	0.12
		10		1.62×10^{-3}	1.17×10^{-2}	1.2×10^{-2}
		1		1.62×10^{-4}	1.17×10^{-3}	1.2×10^{-3}
	10^{-7}	100		5.41×10^{-9}	3.92×10^{-8}	4×10^{-8}
		10		5.41×10^{-10}	3.92×10^{-9}	4×10^{-9}
		1		5.41×10^{-11}	3.92×10^{-10}	4×10^{-10}
1000	0.15	100		1.23×10^{-12}	4.88×10^{-4}	5.95×10^{-4}
		10		1.23×10^{-13}	4.88×10^{-5}	5.95×10^{-5}
		1		1.23×10^{-14}	4.88×10^{-6}	5.95×10^{-6}
	0.3	100		2.46×10^{-12}	9.77×10^{-4}	1.19×10^{-3}
		10		2.46×10^{-13}	9.77×10^{-5}	1.19×10^{-4}
		1		2.46×10^{-14}	9.77×10^{-6}	1.19×10^{-5}
	10^{-7}	100			3.26×10^{-10}	3.97×10^{-10}
		10			3.26×10^{-11}	3.97×10^{-11}
		1			3.26×10^{-12}	3.97×10^{-12}
10000	0.15	100			7.65×10^{-7}	5.54×10^{-6}
		10			7.65×10^{-8}	5.54×10^{-7}
		1			7.65×10^{-9}	5.54×10^{-8}
	0.3	100			1.53×10^{-6}	1.11×10^{-5}
		10			1.53×10^{-7}	1.11×10^{-6}
		1			1.53×10^{-8}	1.11×10^{-7}
	10^{-7}	100			5.1×10^{-13}	3.69×10^{-12}
		10			5.1×10^{-14}	3.69×10^{-13}
		1			5.1×10^{-15}	3.69×10^{-14}

TABLE III. Calculated powers of the differential absorption and scattering signals from molecular H_2 for a laser power of 100 kW, local molecular densities of $10^{10}-10^{16}$ m^{-3} of a 10 m path length, ranging distances of $0.1-10.0$ km, and different scattering targets.

Backward differential absorption and scattering signal power (W) for a laser power $P=100$ kW					
		N_a, cm^{-3}			
R, m	ρ	10^{16}	10^{14}	10^{12}	10^{10}
100	0.15	1.24×10^{-10}	4.91×10^{-2}	5.99×10^{-2}	6.0×10^{-2}
	0.3	2.47×10^{-10}	9.82×10^{-2}	0.1197	0.1199
	10^{-7}		3.27×10^{-8}	3.99×10^{-8}	3.998×10^{-8}
1000	0.15	1.23×10^{-12}	4.883×10^{-4}	5.94×10^{-4}	5.964×10^{-4}
	0.3	2.45×10^{-12}	9.766×10^{-4}	1.188×10^{-3}	1.193×10^{-3}
	10^{-7}		3.26×10^{-10}	3.968×10^{-10}	3.976×10^{-10}
10000	0.15	1.16×10^{-14}	4.626×10^{-6}	5.64×10^{-6}	5.651×10^{-4}
	0.3	2.33×10^{-14}	9.25×10^{-6}	1.128×10^{-5}	1.13×10^{-5}
	10^{-7}		3.08×10^{-12}	3.76×10^{-12}	3.767×10^{-12}

separated from the laser by a distance of 10 km was calculated for concentrations of $10^{10}-10^{16}\,\mathrm{cm}^{-3}$. Table III shows only the maximum powers for a peak laser power of 100 kW and distances of 100, 1000 and 10 000 m. An analysis of these results shows that the power varies by more than 8 orders of magnitude, so that not all of the data can be detected by a lidar of this type. The optimum is to use a laser in the lidar whose output power can be changed as the concentration of H_2 molecules varies, so that it is possible to obtain a maximum backscattered signal power for distances of 0.1 $-$ 10.0 km.

All these results, however, are for the case of zero background illumination. Since background thermal radiation has a strong effect on the power detected by a lidar, we have calculated the background power at the detector, $P_b(\lambda,R)$. The spectral brightness of the IR background, $S_b(\lambda)$, was taken from Ref. 4. Using this value of $S_b(\lambda)$, an equation of the type[14]

$$P_b(\lambda,R)=S_b(\lambda)T(\lambda,R)K_2\xi_p(\lambda)A_2\Omega(R)\Delta\lambda, \tag{7}$$

was used to calculate the background power $P_b(\lambda,R)$ for our case, where $\Omega(R)$ is the solid angle of the field of view of the receiver telescope and $\Delta\lambda$ is the spectral bandwidth of the receiver.

Following Ref. 4, we shall consider the minimum allowable signal to noise ratio to be 1.5 and define the minimum power P_m that can be detected by the lidar as

$$P_m=1.5P_b(\lambda,R). \tag{8}$$

Some calculations for the chosen experimental situation yield the following: 5×10^{-14} W for a distance of 100 m, 5×10^{-16} W for 1 km, and 5×10^{-18} W for 10 km. A comparison of these results with the data of Tables II and III (where the empty places mean that the calculated powers are less than P_m) shows that the power exceeds the background by the greatest amount for concentrations below $10^{16}\,\mathrm{cm}^{-3}$ for the entire range of distances and below $10^{14}\,\mathrm{cm}^3$ up to 1 km. A 100 kW laser makes it possible to detect a concentration of molecules below $10^{14}\,\mathrm{cm}^{-3}$ for any of these distances.

The above results, therefore, demonstrate the possibility of choosing optimum laser parameters for differential absorption and scattering sensing of molecular hydrogen in the atmosphere at a given distance. Taking the background conditions into account limits the prospects for such a laser only at high concentrations and large distances.

[1] V. V. Zuev, M. Yu. Kataev, M. M. Makogon, and A. A. Mitsel', Opt. Atmos. Okeana **8**, 1136 (1995).
[2] V. E. Zuev, *Propagation of Laser Radiation in the Atmosphere* [in Russian], Radio i Svyaz', Moscow (1981), 288 pp.
[3] T. M. Engoyan, V. I. Zhil'tsov, V. I. Kozintsev *et al.*, ''Ranging devices and methods for measuring atmospheric pollution and emissions,'' in Proceedings of the E. K. Fedorov Institute of Applied Geophysics [in Russian], Gidrometeoizdat, Moscow (1986), pp. 53–61.
[4] R. Measures, *Laser Remote Sensing* [Wiley, New York (1984); Mir, Moscow (1987), 550 pp.].
[5] M. V. Vol'kenshteĭn, L. A. Gribov, M. A. El'yashevich, and B. I. Stepanov, in *Molecular Vibrations* [in Russian], Nauka, Moscow (1972), p. 486.
[6] V. E. Zuev, Yu. S. Makushkin, and Yu. N. Ponomarev, *Spectroscopy of the Atmosphere* [in Russian], Gidrometeoizdat, Leningrad (1987), 248 pp.
[7] A. N. Bobrovskiĭ, V. D. Bondaryuk, A. A. Kirillov *et al.*, Kvantovaya Élektron. (Moscow) **17**, 859 (1990) [Sov. J. Quantum Electron. **17**, 778 (1990)].
[8] L. N. Sinitsa, Opt. Atmos. Okeana **8**, 157 (1995).
[9] A. M. Prokhorov (ed.), *Laser Handbook* [in Russian], Sov. Radio, Moscow (1978), Vol. 1, pp. 382, 385.
[10] H. Weber, M. Bass, T. Varitumos, and D. Bua, Sov. J. Quantum Electron. **QE-9**, 1079 (1973).
[11] N. Sugimoto, N. Sims, K. Chan, and D. K. Killinger, Opt. Lett. **15**, 302 (1990).
[12] M. D. Aksenenko and M. L. Baranochnikov, in *Optical Radiation Detectors. A Handbook* [in Russian], Radio i Svyaz', Moscow (1987), pp. 68, 69.
[13] Kh. I. Zil'bershteĭn (ed.), *Spectral Analysis of Pure Substances* [in Russian], Khimiya, St. Petersburg (1994), 336 pp.
[14] H. Inaba and T. Kobayasi, Opto-Electronics **4**(2), 101 (1972).

Translated by D. H. McNeill

Theory of nonlinear directional couplers

P. I. Khadzhi and O. K. Orlov

T. G. Shevchenko Dnestr State University, 278000 Tiraspol, Moldavia
(Submitted June 30, 1998)
Zh. Tekh. Fiz. **69**, 69–71 (August 1999)

A method is proposed for obtaining exact analytical solutions for the system of nonlinear equations for the propagation of electromagnetic waves in directional couplers with an arbitrary nonlinearity in the propagation constant. The resulting solutions can be used to determine the operating characteristics of nonlinear directional couplers. © *1999 American Institute of Physics.* [S1063-7842(99)01308-2]

It has been shown[1,2] that using nonlinear directional couplers opens up broad possibilities for creating ultrafast, purely optical switching devices. In the overwhelming majority of published papers, it is assumed that the nonlinear directional couplers are based on optically nonlinear media with a Kerr nonlinearity in the refractive index or in the propagation constant. In such media, however, the threshold power for switching is rather high. By now, a satisfactory theory for the propagation of light in nonlinear directional couplers with a Kerr nonlinearity has been developed, exact analytical solutions have been obtained for the system of nonlinear equations for the intensities of the propagating waves in terms of elliptical functions, and the switching characteristics have been studied.[1] In addition, couplers based on semiconductors, with their inherent giant resonant nonlinearities, are of great interest, since they can ensure a substantial reduction in the switching power. Media which are modeled by a system of two-level atoms inherently have a saturation effect which substantially modifies the refractive index and absorption coefficient. Numerical studies have been made[3–5] of the effect of saturation in the refractive index of the medium and it has been shown that the operational characteristics of the directional couplers are qualitatively different from those of Kerr media. Evidently, media with excitonic and biexcitonic nonlinearities offer even greater promise.[6] Here the dependence of the refractive index of the medium on the intensity of a propagating wave can be rather complex. We know of no papers in which an analytical solution has been obtained for the nonlinear equations describing the propagation of light in nonlinear directional couplers with a non-Kerr nonlinearity. There is, therefore, some interest in obtaining these solutions for an arbitrary dependence of the ''propagation constant'' on the wave intensity. Here we shall show that the traditionally employed system of first order, nonlinear differential equations for coupled waves propagating in the same direction in two channels of a nonlinear directional coupler can be solved exactly in quadrature form for an arbitrary dependence of the ''propagation constant'' β on the intensity J of the propagating wave.

Let us consider a directional coupler consisting of two different optical waveguides whose propagation constants β_1 and β_2 depend on the intensities J_1 and J_2 of the light propagating in each of the waveguides as

$$\beta_1=\beta_{01}+f_1(J_1), \quad \beta_2=\beta_{02}+f_2(J_2), \tag{1}$$

where β_{01} and β_{02} are constants and $f_1(J_1)$ and $f_2(J_2)$ are arbitrary continuous functions of the intensity.

We shall assume that the coupling constant γ of the coupler is independent of the intensity, which is essentially always true.[1,2] In this case, the nonlinear differential equations for the amplitudes E_1 and E_2 of the coupled waves propagating in the x direction in each of the optical waveguides of the coupler have the form

$$\frac{dE_1}{dx}=-i\beta_{01}E_1-if_1(J_1)E_1+i\gamma E_2,$$

$$\frac{dE_2}{dx}=-i\beta_{02}E_2-if_2(J_2)E_2+i\gamma E_1. \tag{2}$$

In essentially all published papers, this system of equations is solved under the assumption that each of the fields has its own amplitude and phase and one goes from Eq. (2) to a system of nonlinear equations for the amplitudes and phase differences.[1,2] We believe the following approach, which leads to a solution for Eqs. (2) in quadrature form, is more convenient. Let us introduce the new functions

$$J_1=\frac{c}{8\pi}|E_1|^2, \quad J_2=\frac{c}{8\pi}|E_2|^2,$$

$$Q=\frac{c}{8\pi}(E_1^*E_2-E_2^*E_1), \quad R=\frac{c}{8\pi}(E_1^*E_2+E_2^*E_1). \tag{3}$$

Using Eq. (2) and the system of coupled (2) equations, we obtain the following system of nonlinear differential equations for the new functions:

$$\frac{dJ_1}{dx}=i\gamma Q, \quad \frac{dJ_2}{dx}=-i\gamma Q, \tag{4}$$

$$\frac{dQ}{dx}=i(\beta_{01}-\beta_{02})R+i[f_1(J_1)-f_2(J_2)]R + 2i\gamma(J_1-J_2), \tag{5}$$

and

$$\frac{dR}{dx}=i(\beta_{01}-\beta_{02})Q+i[f_1(J_1)-f_2(J_2)]Q. \tag{6}$$

We shall find solutions for this system with the simplest boundary conditions. Let laser light of intensity $J_1|_{x=0}=J_0$ be incident on the entrance end of one of the waveguides in the coupler (say, the first). Then $J_2|_{x=0}=Q|_{x=0}=R|_{x=0}=0$. Equation (4) easily yields the following integral of the motion:

$$J_1+J_2=J_0, \tag{7}$$

which is a consequence of the conservation of energy in the system. Using Eqs. (7), (4), and (6), we obtain

$$R=\frac{\Delta\beta_0}{\gamma}(J_1-J_0)+\frac{1}{\gamma}[F_1(J_1)+F_2(J_2)-F_1(J_0)], \tag{8}$$

where the new function $F(J)$ is introduced in the form

$$F_i(J)=\int_0^J f_1(J')dJ'; \quad i=1,2, \tag{9}$$

with $F_1(0)=F_2(0)=0$ and $\delta\beta_0=\beta_{10}-\beta_{20}$. After this, with the aid of Eqs. (8) and (4) we obtain from Eq. (5) an integral of the motion:

$$Q^2=-4J_1J_2+\frac{1}{\gamma^2}[\Delta\beta_0J_2+F_1(J_0)-F_1(J_1)-F_2(J_2)]^2. \tag{10}$$

Using Eqs. (4) and (10), one can obtain a representation of the phase trajectory of the solution in the $(J_2, dJ_2/dx)$ plane for the intensity of a wave propagating in the second waveguide of the nonlinear directional coupler,

$$\left(\frac{dJ_2}{dx}\right)^2+W(J_2)=0, \tag{11}$$

where

$$W(J_2)=[\Delta\beta_0J_2+F_1(J_0)-F_1(J_0-J_2)-F_2(J_2)]^2 -4\gamma^2J_2(J_0-J_2) \tag{12}$$

plays the role of the potential energy of a conservative nonlinear oscillator. The condition $W(J_2)=0$ can be used to determine the minimum and maximum intensities J_2 of the wave in the second waveguide of the nonlinear directional coupler. Similar expressions are easily written down for J_1, as well.

Finally, from Eq. (1) it is easy to obtain a solution for $J_2(x)$ in quadrature form,

$$\int_0^{J_2}\frac{dJ_2}{\sqrt{-W(J_2)}}=x. \tag{13}$$

The solution (13), together with Eq. (7), completely describes the spatial distribution of the wave intensities in both waveguides of the nonlinear directional coupler.

Assuming that the optical waveguides are linear, i.e, assuming that $f_1(J_1)=f_2(J_2)=0$, from Eq. (13) we obtain the standard result[7,8]

$$J_2(x)=\frac{J_0}{1+\left(\frac{\Delta\beta_0}{2\gamma}\right)^2}\sin^2\sqrt{1+\left(\frac{\Delta\beta_0}{2\gamma}\right)^2}\,\gamma x. \tag{14}$$

If the optical waveguides are nonlinear with Kerr corrections to the propagation constants, $f_1=\alpha_1J_1$ and $f_2=\alpha_2J_2$, where α_1 and α_2 are constants, then from Eq. (12) we obtain

$$W(J_2)=J_2\left[\frac{1}{4}(\alpha_1+\alpha_2)^2J_2^3-(\Delta\beta_0+\alpha_1J_0)(\alpha_1+\alpha_2)J_2^2 +[4\gamma^2+(\Delta\beta_0+\alpha_1J_0)^2]J_2-4\gamma^2J_0\right], \tag{15}$$

while the integral (13) can be expressed in terms of the Jacobi elliptic functions.[1] Rather than present all the possible solutions, we shall note just one unique simple solution. Assuming that one waveguide of the nonlinear directional coupler is self-focusing, while the other is self-defocusing (i.e., $\alpha_2=-\alpha_1$), we obtain a solution characteristic of a linear directional coupler,

$$J_2=\frac{4\gamma^2}{4\gamma^2+(\Delta\beta_0+\alpha_1J_0)^2}J_0 \times\sin^2\sqrt{\gamma^2+\frac{1}{4}(\Delta\beta_0+\alpha_1J_0)^2}\,x. \tag{16}$$

Here the coupling length and transfer efficiency depend on the intensity J_0 of the light delivered to the front end of the first waveguide.

If we assume that both waveguides of the nonlinear directional coupler are identical and their ''propagation constants'' are characterized by a saturation effect $\beta_1=\beta_2=\beta_0+\alpha(1+J/J_s)^{-1}$,where J_s is the saturation intensity, then from Eq. (12) we easily obtain

$$W(J_2)=-4\gamma^2\left\{J_2(J_0-J_2)-\frac{\alpha^2}{4\gamma^2} \times\left[\frac{\ln\left(1+\frac{J_2}{J_s}\right)\ln\left(1+\frac{J_0-J_2}{J_s}\right)}{\ln\left(1+\frac{J_0}{J_s}\right)}\right]^2\right\}. \tag{17}$$

Equation (13) can be integrated further using Eq. (17) only numerically.

Note that the assumption of nonzero transfer from the front ends of both waveguides of the nonlinear directional coupler, i.e., $J_1|_{x=0}=J_{01}$ and $J_2|_{x=0}=J_{02}$, does not, in principle, make it any more difficult to obtain the general solutions.

[1] A. Maĭer, Usp. Fiz. Nauk **165**, 1037 (1995).
[2] M. Jensen, IEEE J. Quantum Electron. **QE-18**, 1580 (1982).
[3] G. I. Stegeman, C. T. Seaton, C. N. Ironside *et al.*, Appl. Phys. Lett. **50**, 1035 (1987).
[4] F. Kh. Abdullaev and R. Gulyamov, Pis'ma Zh. Tekh. Fiz. **18**(20), 10 (1992) [Sov. Tech. Phys. Lett. **18**(10), 653 (1992)].

[5] J. D. Begin and M. Cada, IEEE J. Quantum Electron. **QE-30**, 3006 (1994).
[6] P. I. Khadzhi, *Nonlinear Optical Processes in Exciton and Biexciton Systems in Semiconductors* [in Russian], Stiintsa, Kishinev (1985).
[7] A. Yariv, IEEE J. Quantum Electron. **QE-9**, 919 (1973).
[8] A. Yariv and P. Yeh, *Optical Waves in Crystals* [Wiley, New York (1984); Mir, Moscow (1987)].

Translated by D. H. McNeill

Some features of the polarization characteristics of strongly absorbing helical periodic media

A. A. Gevorgian

Erevan State University, 375049 Erevan, Armenia
(Submitted July 8, 1997; resubmitted April 6, 1998)
Zh. Tekh. Fiz. **69**, 72–78 (August 1999)

Results are presented on the polarization characteristics (rotation of the plane of polarization and polarization ellipticity) as a function of the layer thickness and the absorption anisotropy in strongly absorbing media having a helical structure. A strong resonancelike change in the polarization ellipticity is found as a function of the anisotropy of the absorption at frequencies of diffractional interaction of the light with the medium. A change in the sign of rotation of the plane of polarization of the light is observed as the layer thickness is varied. It is established that sign of the rotation also changes as the absorption anisotropy varies. These effects are studied under conditions of interaction of light with a half space and with a layer of medium of finite thickness. Some new features are identified in the previously observed effect wherein the absorption of radiation in media having a periodic structure decreases as the layer thickness increases. © *1999 American Institute of Physics.* [S1063-7842(99)01408-7]

INTRODUCTION

Studies of the optical properties of strongly absorbing helical periodic media were reported in Refs. 1–5. These media include cholesteric liquid crystals, chiral smectics, helical magnetic media, and artificial ferromagnetic helical structures. These media come within the definition of gyrotropy given by Fedorov,[6] although in the natural state they cannot exhibit spatial dispersion and cannot be magnetoactive. Some characteristics of the Faraday effect in helical periodic media under conditions of oblique incidence were studied in Ref. 7. The optical activity spectra were investigated in Ref. 8 the propagation of light in media possessing dielectric and magnetic helicity was examined in Ref. 9. Some features of the amplitude characteristics of strongly absorbing helical media were studied in Refs. 10 and 11, and an effect wherein the absorption decreases with increasing layer thickness and increasing absorption anisotropy. There it was shown that when light is incident normally on a planar layer of helical periodic medium, three diffraction mechanisms take place: diffraction of light by the periodic helicity caused by refraction anisotropy, diffraction of light by the periodic helicity caused by absorption anisotropy, and diffraction of light in a bounded volume caused by the finite nature of the layer thickness. The present paper reports the results of a further study of the properties of strongly absorbing helical periodic media. Some unique effects are identified, i.e., a change in the sign of rotation with varying layer thickness and with varying absorption anisotropy, and also a resonancelike variation of the ellipticity as a function of the absorption anisotropy.

Whereas the wavelength dependences of the polarization characteristics of media having a helical structure and also the influence of isotropic and anisotropic absorption on this dependence have been studied in fairly great detail,[1–5] the same cannot be said of the dependence of the polarization characteristics on the layer thickness. We merely note Ref. 12, in which the rotation of the plane of polarization was studied experimentally as a function of the layer thickness. As far as we are aware, no studies have been made of the polarization characteristics of media having a helical structure as a function of the absorption anisotropy. In the present paper we attempt to fill this gap.

In order to facilitate our analysis of the mechanisms responsible for the observed behavior for a layer of finite thickness, we first deem it necessary to analyze the interaction of light with a half space.

BOUNDARY-VALUE PROBLEM FOR A HALF SPACE

We shall consider the case of light normally incident on a half space filled with a medium having a helical structure whose axis is perpendicular to the boundary surface. The field in the medium at the distance z from the boundary has the form[1,2]

$$\mathbf{E}(z,t)=\{[E_1^+ \exp(ik_1z)+E_2^+ \exp(ik_2z)] \exp(iaz)\mathbf{n}_+ + [\xi_1E_1^+ \exp(ik_1z)+\xi_2E_2^+\exp(ik_2z)] \times\exp(-iaz)\mathbf{n}_-\} \exp(-i\omega t), \quad (1)$$

where

$$\xi_{1,2}=-\delta/[1-(\chi\pm b_{1,2})^2];$$
$$k_{1,2}=2\pi b_{1,2}\sqrt{\varepsilon_m}/\lambda; \quad a=2\pi/\sigma;$$
$$b_{1,2}=\sqrt{1+\chi^2\pm\gamma}; \quad \gamma=\sqrt{4\chi^2+\delta^2};$$
$$\delta=\varepsilon_a/\varepsilon_m; \quad \chi=\lambda/\sigma\sqrt{\varepsilon_m};$$
$$\varepsilon_m=(\varepsilon_1+\varepsilon_2)/2; \quad \varepsilon_a=(\varepsilon_1-\varepsilon_2)/2; \quad (2)$$

$\mathbf{n}_\pm=(\mathbf{x}\pm i\mathbf{y})/\sqrt{2}$ are the unit vectors of the circular polarizations, ε_1 and ε_2 are the principal values of the permittivity

tensor in the plane perpendicular to the axis of the medium, λ is the wavelength in vacuum, σ is the pitch of the helix, and the amplitudes $E^+_{1,2}$ are determined from the boundary conditions.

The influence of the periodicity of the structure and the helix parameters on the change in the field and also the wavelength dependence of the fields excited in the medium have been analyzed in detail in many studies (see, in particular, Refs. 1–3). In accordance with formula (1), the rotation of the plane of polarization may be expressed in the form

$$\varphi = \varphi_0 - az, \tag{3}$$

where φ_0 is the rotation of the plane of polarization if the field in the medium is expressed in the form

$$\mathbf{E}(z,t) = \{[E_1^+ \exp(ik_1 z) + E_2^+ \exp(ik_2 z)]\mathbf{n}_+ + [\xi_1 E_1^+ \exp(ik_1 z) + \xi_2 E_2^+ \exp(ik_2 z)]\mathbf{n}_-\} \times \exp(-i\omega t). \tag{4}$$

Since the analytical formulas are cumbersome, the dependences of the polarization characteristics on the distance z from the boundary and also on the imaginary part of the dielectric anisotropy ε_a'' are best analyzed by means of numerical calculations using the formulas put forward above. A single prime will denote the real part of a particular quantity and a double prime will denote the imaginary part.

Figure 1a gives the rotation φ_0 of the plane of polarization and Fig. 1b gives the polarization ellipticity e as a function of the distance z from the boundary for anisotropic absorption at the following characteristic wavelengths of the incident light: *1* — $\lambda = \sigma\sqrt{|\varepsilon_m|(1-|\delta|)} \approx 0.615\ \mu$m and *2* — $\lambda = \sigma\sqrt{|\varepsilon_m|(1+|\delta|)} \approx 0.635\ \mu$m (near the boundary of the selective reflection region), *3* — $\lambda = 0.2\ \mu\text{m} < \sigma\sqrt{|\varepsilon_m|}$, and *4* — $\lambda = 1.5\ \mu\text{m} > \sigma\sqrt{|\varepsilon_m|}$ (far from the selective reflection region). Note that here and subsequently, to be specific, we consider the case $\varepsilon_a'' > 0$ when studying the dependences of φ_0 and e both on ε_a'' and on the distance z from the boundary (under conditions of anisotropic absorption).

It can be seen from Fig. 1 that under conditions of anisotropic absorption ''saturation'' of the rotation is observed at the frequencies of diffractional interaction of the light with the medium: after passing through a peak the rotation undergoes damped oscillations about a certain value. Saturation of the rotation has also been observed experimentally.[12]

An interesting pattern is observed near the long-wavelength boundary. Here, before reaching saturation the rotation decreases, goes to zero, changes sign, and only then reaches saturation. The calculations show that this pattern only occurs for specific values of the absorption anisotropy. For $\varepsilon_a'' < 0$ the opposite pattern is observed. In particular, the change in the sign of the rotation is observed near the short-wavelength boundary of the selective reflection region.

It can be seen from the figures that the polarization ellipticity also saturates at the frequencies of diffractional interaction of the light with the medium. In this case, whereas near the short-wavelength boundary of the selective reflection region the ellipticity saturates by decreasing, near the long-wavelength boundary it saturates by increasing.

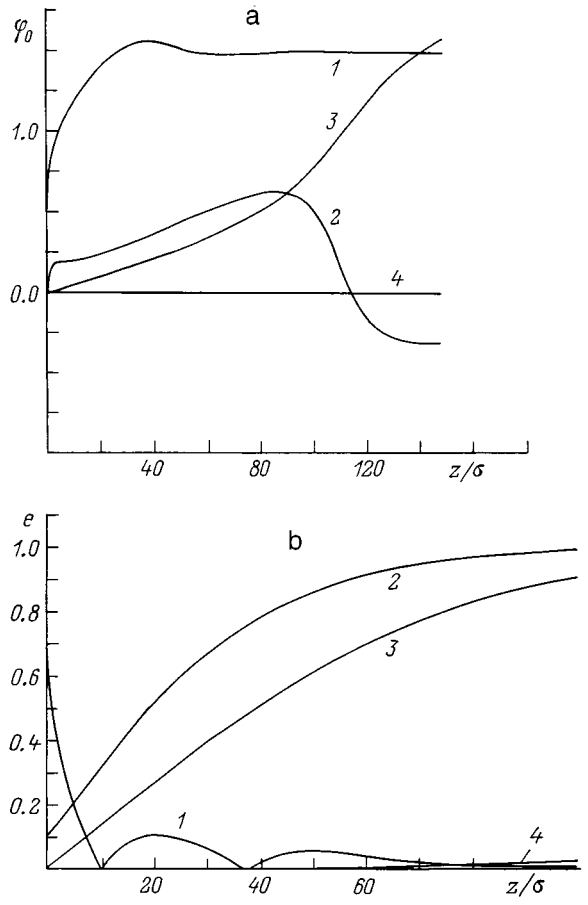

FIG. 1. Rotation of the plane of polarization (a) and the ellipticity (b) as functions of the parameter z/σ for anisotropic absorption: $\varepsilon_1 = 2.29$, $\varepsilon_2' = 2.143$, $\varepsilon_1'' = 0.1$, $\varepsilon_2'' = 0$, and $\sigma = 0.42\ \mu$m.

Figure 2a gives the rotation φ_0 and Fig. 2b gives the polarization ellipticity e as a function of the absorption anisotropy [the parameter $\ln(2\varepsilon_a'')$] for the same wavelengths of the incident light as in Fig. 1. These figures demonstrate resonancelike behavior of the ellipticity as a function of the absorption anisotropy at wavelengths near the selective reflection region. It can also be seen that at specific wavelengths the sign of rotation also changes as the absorption anisotropy varies.

DISCUSSION

In order to identify the mechanisms responsible for the observed behavior, we shall use an expression for the wave field in the medium. In cases of weak anisotropy the amplitude of one of the natural waves having nondiffracting circular polarization is much smaller than the amplitudes of the other waves, so that Eq. (4) can be expressed in the form

$$\mathbf{E}(z,t) = \{[E_1^+ \exp(ik_1 z) + E_2^+ \exp(ik_2 z)]\mathbf{n}_+ + \xi_2 E_2^+ \times \exp(ik_2 z)\mathbf{n}_-\} \exp(-i\omega t). \tag{5}$$

This is a fairly good approximation and can explain many of the characteristic features of the properties of helical periodic media. In particular, the saturation of the rotation and the ellipticity under conditions of anisotropic absorption

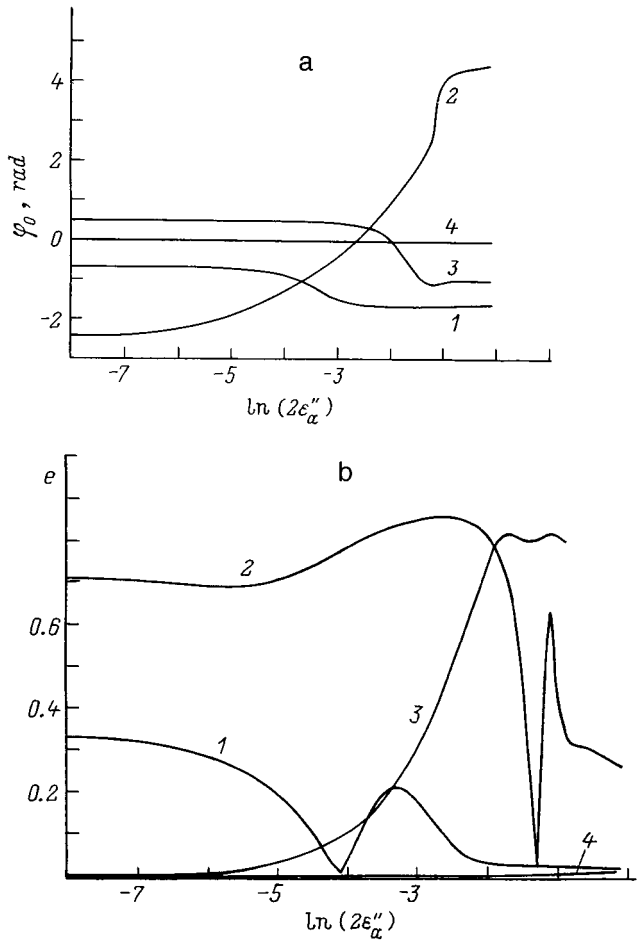

FIG. 2. Rotation of the plane of polarization (a) and the ellipticity (b) as functions of the parameter $\ln(2\varepsilon_a'')$ for anisotropic absorption: $z=50\sigma$; the other parameters are the same as in Fig. 1.

near the short-wavelength boundary can be attributed to the fact that the different natural waves undergo different absorption. This can be seen from the expressions for k_1 and k_2. It can also be confirmed analytically using for the imaginary parts of the wave numbers the following approximate expressions obtained under the condition $1-|\chi|$, $|\chi| \gg \varepsilon_a''$, ε_a', ε_m'':

$$k_1'' \approx \pi\varepsilon_m''/(\lambda\sqrt{\varepsilon_m'}),$$

$$k_2'' \approx \frac{\pi}{\lambda\sqrt{\varepsilon_m'}}\left[\varepsilon_m'' - \frac{\varepsilon_a''\varepsilon_a'}{2\varepsilon_m'|\chi|(1-|\chi|)}\right], \quad |\chi| = \lambda/(\sigma\sqrt{\varepsilon_m'}). \tag{6}$$

The results of an exact calculation of k_1'' and k_2'' as functions of the wavelength for various values of ε_a'' and ε_m'' (Fig. 3) confirm what we have said. Near the short-wavelength boundary $\lambda_1 = \sigma\sqrt{|\varepsilon_m'|(1-|\delta|)}$ we find $k_2'' \sim 0$ and thus $k_1'' \gg k_2''$ (absorption suppression effect). An increase in thickness leads to a rapid decrease in the amplitudes of the natural waves, proportional to $\exp(-k_{1,2}''z)$. Since $k_1'' \gg k_2''$, those amplitudes which are proportional to $\exp(-k_1''z)$ decrease rapidly and become negligible as z increases. Consequently, at some distance from the boundary lying near the short-wavelength boundary, the field may be expressed in the form

$$\mathbf{E}(z,t) = E_2^+ \exp(ik_2z)(\mathbf{n}_+ + \xi_2\mathbf{n}_-) \exp(-i\omega t). \tag{7}$$

It therefore follows that the ellipticity is $e=(|\xi_2|-1)/(|\xi_2|+1)$ and does not depend on z. Moreover, since $|\xi_2| \approx 1$ near the selective reflection region, we find $e \approx 0$.

The rotation of the plane of polarization at some distance from the boundary is determined only by the real and imaginary parts of ξ_2 and also does not depend on z. After comparatively rapid variations (in regions of rapidly varying

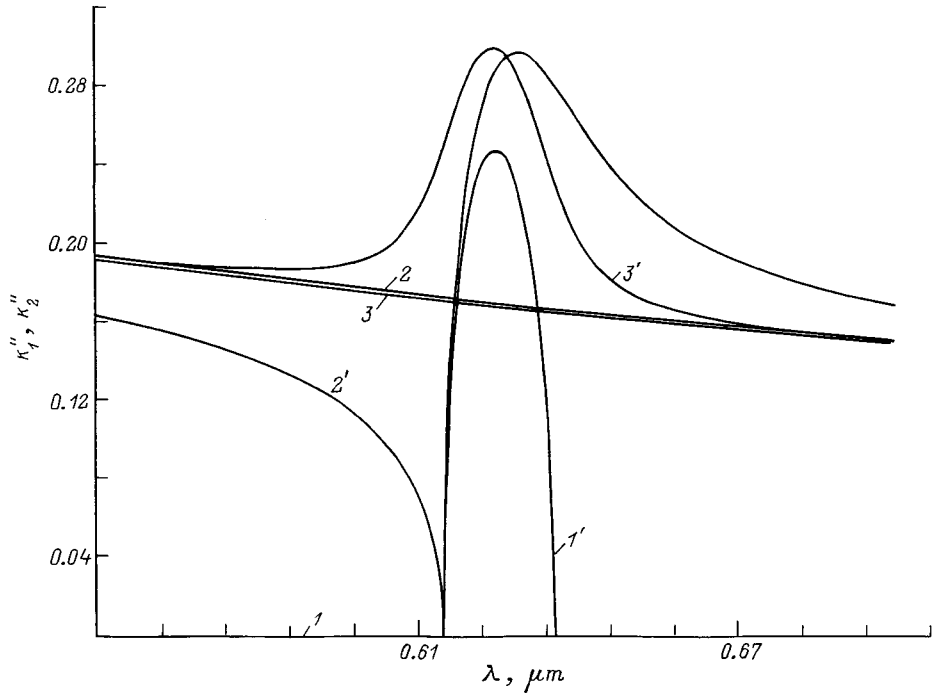

FIG. 3. Imaginary parts of the nonresonant k_1'' and resonant k_2'' wave numbers as functions of the wavelength λ for various values of the absorption anisotropy ε_a'' and the average absorption ε_m'': *1* — $\varepsilon_a''=\varepsilon_m''=0$; *2* — $\varepsilon_a''=\varepsilon_m''=0.05$; *3* — $\varepsilon_a''=0$, $\varepsilon_m''=0.05$; the other parameters are the same as in Fig. 1; *1–3* — k_1'', *1′–3′* — k_2''.

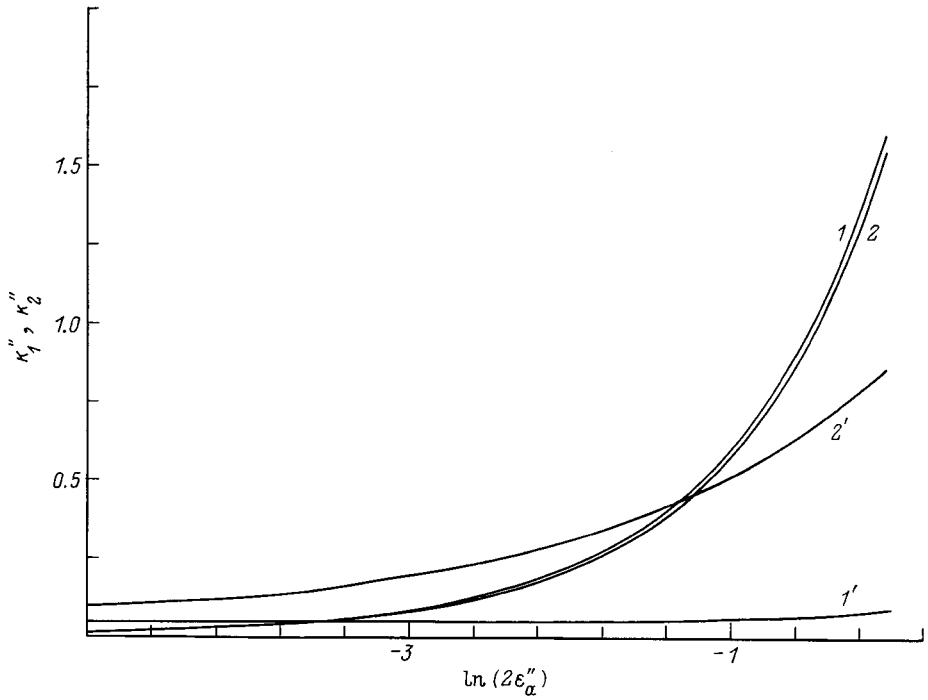

FIG. 4. Imaginary parts of the nonresonant and resonant wave numbers as functions of the absorption anisotropy: *1* — $\lambda_1=0.615\,\mu$m, *2* — $\lambda_2=0.635\,\mu$m; the other parameters are the same as in Fig. 1; *1, 2* — k_1''; *1′, 2′* — k_2''.

wave amplitudes), the rotation goes to saturation. The change in the sign of rotation observed near the long-wavelength boundary of the selective reflection region occurs because, under certain conditions of anisotropic absorption, the natural wave, whose amplitude is usually neglected, begins to play an important role. Near the long-wavelength boundary we find $k_1''<k_2''$ (anomalously strong absorption). This implies that whereas for small z the field can be represented as the sum (5), as z increases further the amplitudes of the natural waves which are proportional to $\exp(-k_2''z)$ decrease more rapidly, and beyond a certain value of z they become smaller than the amplitude of the natural wave whose field was neglected [i.e., the natural wave with amplitude proportional to $\xi_1\exp(-k_1''z)$]. The amplitude of this wave decreases far more slowly with increasing z. The direction of rotation begins to change near these values of z. As z increases further, the field can be expressed as the sum

$$\mathbf{E}(z,t)=E_1^+\ \exp(ik_1z)(\mathbf{n}_+ + \xi_1\mathbf{n}_-)\ \exp(-i\omega t). \qquad (8)$$

Thus whereas for small z the field is formed by the sum of circular waves (5) and the rotation takes place in one direction, for large z the field has the form (8) and the rotation changes direction, since the fast circular component is replaced by a slow one. Similarly, i.e., by studying the characteristics of the natural waves in the medium, we can explain other features in the dependences of the polarization characteristics on the distance z from the boundary. Characteristic features in the dependences of φ_0 and e on the parameter $\ln(2\varepsilon_a'')$ can also be explained.

Near the short-wavelength boundary, the wave amplitudes proportional to $\exp(-k_2''z)$ vary negligibly as $\ln(2\varepsilon_a'')$ increases, whereas the amplitude of the other wave, proportional to $\exp(-k_1''z)$, decreases rapidly (in the absence of absorption this amplitude is greater than the other two). Thus resonancelike behavior is observed in the dependence of e on $\ln(2\varepsilon_a'')$.

Note that the sign of rotation changes with varying absorption anisotropy at $\lambda=0.2\,\mu$m (far from the selective reflection region; short-wavelength region). This effect is also caused by the different damping of the natural waves in the medium as the absorption anisotropy varies and by the increased effect of this difference as this anisotropy varies.

Near the long-wavelength boundary, as $\ln(2\varepsilon_a'')$ increases, the wave amplitudes proportional to $\exp(-k_2''z)$ begin to decrease more rapidly than those proportional to $\exp(-k_1''z)$. Moreover, as near the short-wavelength boundary, this amplitude is greater than the other two in the absence of absorption. Here, however, a peculiarity appears. As the absorption anisotropy increases further, the wave amplitudes proportional to $\exp(-k_2''z)$ begin to decrease more slowly than the other wave amplitude, proportional to $\exp(-k_1''z)$ (the diffraction mechanism caused by the absorption anisotropy begins to have an influence, with the result that the mechanism of absorption suppression begins to come into play). Figure 4 gives k_1'' and k_2'' as functions of the absorption anisotropy at wavelengths near the short-wavelength and long-wavelength boundaries of the selective reflection region, which confirm what we have said. Thus above a certain value of $\ln(2\varepsilon_a'')$, the amplitude proportional to $\exp(-k_1''z)$ becomes smaller than the other two. As a result of this behavior of the natural wave amplitudes as functions of the parameter $\ln(2\varepsilon_a'')$, we observe a resonancelike change in the ellipticity and a change in the sign of rotation as functions of this parameter near the long-wavelength boundary of the selective reflection region.

This reasoning suggests that the reduced absorption of radiation in periodic media with increasing layer thickness,

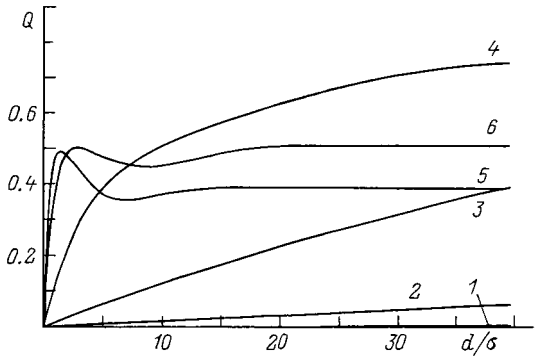

FIG. 5. Results of the absorption of radiation Q in a layer as a function of the parameter d/σ for ε_a'': *1* — 0.00005, *2* — 0.0005, *3* — 0.005, *4* — 0.05, *5* — 0.25, *6* — 0.5; $\varepsilon_m''=\varepsilon_a''$, $\varepsilon_1'=2.29$, $\varepsilon_2'=2.25$, $\sigma=0.42$ μm, and $\lambda=0.635$ μm.

discovered in Refs. 10 and 11, may also be observed near the long-wavelength boundary of the selective reflection region (when $\varepsilon_a''>0$). This reduced absorption is observed at an absorption anisotropy for which the diffraction mechanism attributed to the absorption anisotropy is already appreciable. Calculations made for a layer of finite thickness confirm this. Figure 5 gives the value of $Q=1-(R+T)$ (characterizing the optical energy absorbed in the medium) plotted as a function of the layer thickness (where R is the reflection coefficient and T is the transmission coefficient) for various values of the absorption anisotropy near the long-wavelength boundary of the selective reflection region. It can be seen that above a certain level of absorption anisotropy the value of Q decreases with increasing layer thickness. We also note that, as can be seen from Fig. 5, for the given parameters of the medium the decreasing radiation absorption with increasing layer thickness is observed at ''enormous'' levels of absorption anisotropy (near the long-wavelength boundary a similar effect is observed at a much lower level of absorption anisotropy). Naturally, the lower the refraction anisotropy (first diffraction mechanism), the sooner (i.e., at a lower level of absorption anisotropy) the decrease in absorption with increasing layer thickness set in near the long-wavelength boundary (since this effect is a manifestation of the diffraction of light on the periodic structure created by the absorption anisotropy). Numerical calculations confirm this statement. In fact, the calculations show that for the parameters $\varepsilon_1'=2.29$, $\varepsilon_2'=2.285$ ($\delta\approx0.001$), and $\sigma=0.42$ μm, this effect begins to appear for $\varepsilon_a''=0.02$.

To conclude this section, we note that as the calculations have shown, changes in the sign of rotation with varying layer thickness and varying absorption anisotropy for a cholesteric liquid crystal having the parameters $\varepsilon_1'=2.29$, $\varepsilon_2'=2.143$ ($\delta=0.033$), and $\sigma=0.42$ μm are observed for transmission coefficients T of the order of $10^{-11}-10^{-12}$, i.e., there is no transmitted wave. However, this does not imply that the identified effects are purely ''theoretical.'' The calculations show that these effects depend strongly on the refraction anisotropy ε_a'. As the refraction anisotropy decreases, the transmission coefficient increases rapidly in those regions where the sign of rotation changes. For example, for the parameters $\varepsilon_1'=2.29$, $\varepsilon_2'=2.285$ ($\delta=0.001$), $\sigma=0.42$ μm, a change in the sign of rotation is observed for values of the transmission coefficient T of the order of $10^{-1}-10^{-2}$. Thus, these effects are fully ''experimental,'' i.e., they are amenable to measurement. From this it also follows that the change in rotation identified in this study is a manifestation of the diffraction of light by the periodic helicity caused by the absorption anisotropy.

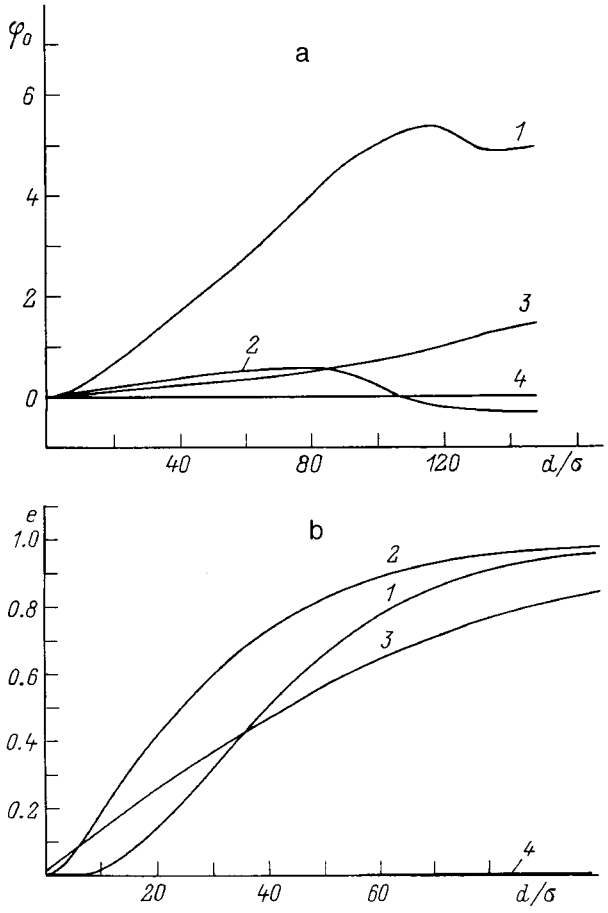

FIG. 6. Rotation of the plane of polarization (a) and the ellipticity (b) as functions of the parameter d/σ for various wavelengths of the incident light for anisotropic absorption. The other parameters are the same as in Fig. 1.

BOUNDARY-VALUE PROBLEM FOR A LAYER

Let us analyze the normal transmission of light through a layer of helical periodic medium whose axis is perpendicular to the boundary surfaces. Figure 6a gives φ_0 and Fig. 6b gives e plotted as functions of the layer thickness d for the same wavelengths of the incident light as in Fig. 1 but for different values of ε_a'' and ε_m'' in the case when the light is transmitted through the layer.

Figure 7a gives φ_0 and Fig. 7b gives e plotted as function of the parameter $\ln(2\varepsilon_a'')$ for the same wavelengths of the incident light as in Fig. 1.

The characteristics and relationships observed in this case can also be attributed to the behavior of the amplitudes

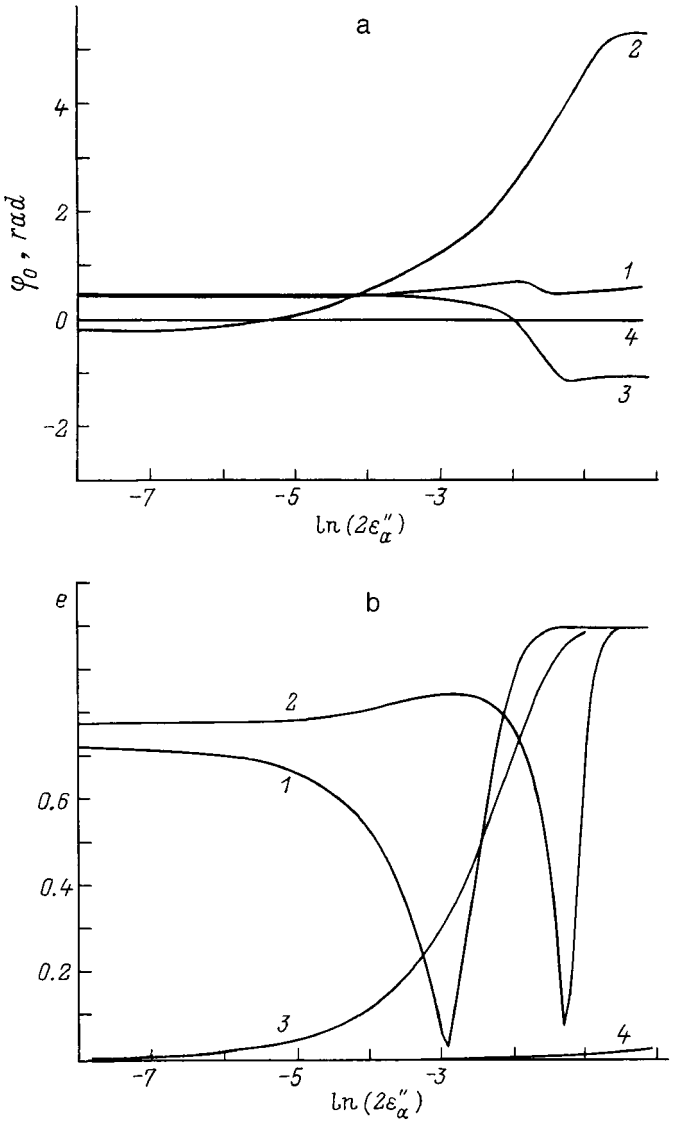

FIG. 7. Rotation of the plane of polarization (a) and the ellipticity (b) as functions of the absorption anisotropy $\ln(2\varepsilon_a'')$ for anisotropic absorption: $d=50\sigma$. The other parameters are the same as in Fig. 1.

and phases of the waves excited in the medium. In this case, however, the field in the medium has the form (at the second boundary)

$$\begin{aligned}\mathbf{E}(d,t)=\{&[(\gamma-\delta-2\chi)+[(2b_1-l_1)\exp(ik_1d)\\&+(2b_1+l_1)\exp(-ik_1d)]/(2b_1)\\&+(\gamma+\delta+2\chi)[(2b_2+l_2)\exp(ik_2d)\\&+(2b_2-l_2)\exp(-ik_2d)]/(2b_2)]\mathbf{n}_+\\&+[(\gamma-\delta+2\chi)[(2b_1-l_1)\exp(ik_1d)\\&+(2b_1+l_1)\exp(-ik_1d)]/(2b_1)]+[(\gamma+\delta-2\chi)\\&\times[(2b_2-l_2)\exp(ik_2d)+(2b_2-l_2)\\&\times\exp(-ik_2d)]/(2b_2)]\mathbf{n}_-\}\exp(-i\omega t)/(4\gamma a_1a_2),\end{aligned}$$

where $l_{1,2}=\gamma\pm2$, $a_{1,2}=\cos(k_{1,2}d)\mp iul_{1,2}\sin(k_{1,2}d)/(k_{1,2}d)$, $u=\pi d/\sqrt{\varepsilon_m}/\lambda$.

A comparison between the curves in Figs. 1a and 1b and those in Figs. 4a and 4b shows that in most cases they are similar. However, there are some important differences. In particular, the dependence of the ellipticity e on z near the short-wavelength boundary under conditions of anisotropic absorption differs substantially from the dependence of e on the layer thickness d under similar conditions. This difference demonstrates once again the difference between the half-space problem and the layer problem (although the layer is arbitrarily thick).

We reiterate that the effects identified in this study are observed in periodic helical media near the selective reflection region, and they occur because the interaction between light and these media is accompanied by the excitation of different natural waves whose amplitudes, phase velocities, and attenuation differ. Thus, depending on the variation in the parameters of the medium responsible for these wave characteristics, quite unusual patterns of wave interaction with the medium can be obtained.

In conclusion, we note that these results can be applied, in particular, to ellipsometry in the design of various ellipsometric systems using layers of media of finite thickness having a helical structure.

The authors are grateful to G. A. Vardanyan and O. S. Eritsyan for valuable discussions.

This work was carried out under Topic No. 96-895, financed by the State Centralized Sources of the Republic of Armenia.

[1] V. A. Belyakov and A. S. Sonin, *Optics of Cholesteric Liquid Crystals* [in Russian], Nauka, Moscow (1982).
[2] V. A. Belyakov, *Diffraction Optics of Complex Periodic Structures* [in Russian], Nauka, Moscow (1988).
[3] O. S. Eritsyan, *Optics of Gyrotropic Media and Cholesteric Liquid Crystals* [in Russian], Aïastan, Erevan (1988).
[4] V. S. Rachkevich, Ber. Bunsenges. Phys. Chem. **93**, 1137 (1989).
[5] V. A. Belyakov, A. A. Gevorgyan, O. S. Eritsyan, and N. V. Shipov, Kristallografiya **33**, 574 (1988) [Sov. J. Crystallogr. **33**, 337 (1988)]; Zh. Tekh. Fiz. **57**, 1418 (1987) [Sov. Phys. Tech. Phys. **32**, 843 (1987)].
[6] F. I. Feodorov, *Theory of Gyrotropy* [in Russian], Nauka i Tekhnika, Minsk (1976).
[7] V. A. Kienya and I. V. Semchenko, Kristallografiya **39**, 514 (1994) [Crystallogr. Rep. **39**, 457 (1994)].
[8] E. K. Galanov and V. G. Medvedev, Opt. Spektrosk. **76**, 79 (1994) [Opt. Spectrosc. **76**, 72 (1994)].
[9] V. N. Kapshaĭ, V. A. Kienya, and I. V. Semchenko, Kristallografiya **36**, 822 (1991) [Sov. J. Crystallogr. **36**, 459 (1991)].
[10] G. A. Vardanyan and A. A. Gevorgyan, Kristallografiya **42**, 316 (1997) [Crystallogr. Rep. **42**, 276 (1997)].
[11] G. A. Vardanyan and A. A. Gevorgyan, Kristallografiya **42**, 723 (1997) [Crystallogr. Rep. **42**, 663 (1997)].
[12] Yu. V. Denisov, V. A. Kizel', E. P. Sukhenko *et al.*, Kristallografiya **21**, 991 (1976) [Sov. J. Crystallogr. **21**, 568 (1976)].

Translated by R. M. Durham

Periodic surface photoreliefs in glassy and hyperelastic polymers

V. V. Mogil'nyĭ, Yu. V. Gritsaĭ, and S. V. Kovalev

Belarus State University, 220050 Minsk, Belarus
(Submitted July 13, 1998)
Zh. Tekh. Fiz. **69**, 79–83 (August 1999)

The formation and degradation of periodic photoreliefs on the surface of polymer layers having significantly different glass-transition temperatures are investigated for various process activation temperatures. It is established that the main factor limiting the resolution of periodic relaxation photoreliefs at the surface of glassy polymer layers containing dimerizing anthracene derivatives is the presence of shear stresses. Their action is suppressed by the thermal decomposition of dimers, a process which gives rise to inverted reliefs of higher spatial frequencies. It is shown that the resolution can be enhanced by more than an order of magnitude by using a polymer matrix in the hyperelastic state. © *1999 American Institute of Physics.* [S1063-7842(99)01508-1]

Reversible photoreliefs formed by heating exposed glassy layers of polymethyl methacrylate containing photodimerizing derivatives of anthracene are described in Ref. 1. The reliefs were formed as a result of the expansion of exposed regions caused by the thermally activated relaxation of nonequilibrium centers (photoproducts and their neighborhoods). The technological simplicity of producing these photoreliefs suggests that they may be used to recorded relief-phase holograms and to form other optical elements. Of particular importance in this context are periodic reliefs and the extent to which they can reproduce the spatial frequencies of an image. Quite clearly, the resolution of the photoreliefs is determined by the properties of the polymer matrix, which in turn depend on the relationship between the temperature at which the relief begins to form and the glass-transition temperature of the material. Here we investigate the formation and degradation of periodic photoreliefs on the surface of polymer layers with significantly different glass-transition temperatures for various process activation energies.

Polymer layers $25-27\,\mu$m thick were prepared by pouring ingredients dissolved in chloroform (polymer $+10$ mol. % 9-anthraldehyde) onto a glass substrate and then drying for 24 h at 293 and 323 K to a constant weight. Polymethyl methacrylate (PMMA) and polybutyl methacrylate (PBMA) were used as the polymer base. In order to achieve a uniform photoproduct distribution over thickness, the samples were exposed to radiation from a DKSSh-1000 xenon lamp (ZhS-16 light filter) or an argon laser ($\lambda = 488$ nm) at the long-wavelength edge of the absorption spectrum of 9-anthraldehyde. The exposure field was formed by an optical test pattern (having the period $d = 25-88\,\mu$m) or by the interference of two laser beams ($d = 5-20\,\mu$m). For the selected layer thickness, the diffraction spreading of the boundaries for exposure through the optical test pattern did not exceed 10% of the smallest period. The concentration of 9-anthraldehyde and the degree of photoconversion were determined using the electron absorption spectra. The layer thickness l_0, the relief height l, and its modulation amplitude Δl (Fig. 1) were determined using an MII-4 microinterferometer. The value of Δl was taken positive if the thickness of the exposed region exceeded that of the unexposed region. For small periods ($d < 20\,\mu$m) Δl and the degree of photoconversion were determined by a holographic method. This involved measuring the diffraction efficiencies of the transmission and reflection relief holograms and using the formula for the diffraction efficiency of two-dimensional phase gratings[2] to determine Δl and the amplitude of modulation of the refractive index Δn. Know-

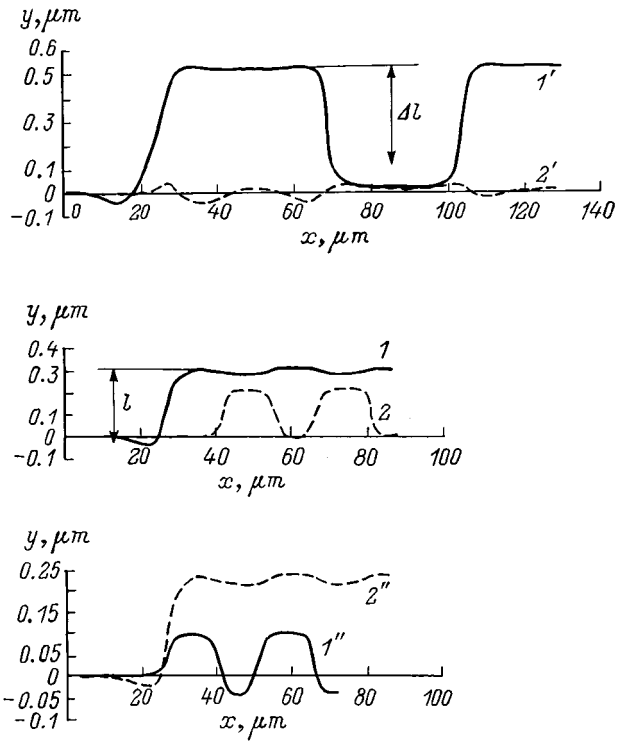

FIG. 1. Profile of periodic relief at the surface of PMMA ($1,1',2,2'$) and PBMA films ($1'',2''$) after annealing for 20 ($1,1'$) and 515 min ($2,2'$) at 364 K and annealing for 270 ($1''$) and 4800 min ($2''$) at 308 K. Spatial period 70 ($1,2$) and 25 μm ($1',1'',2',2''$).

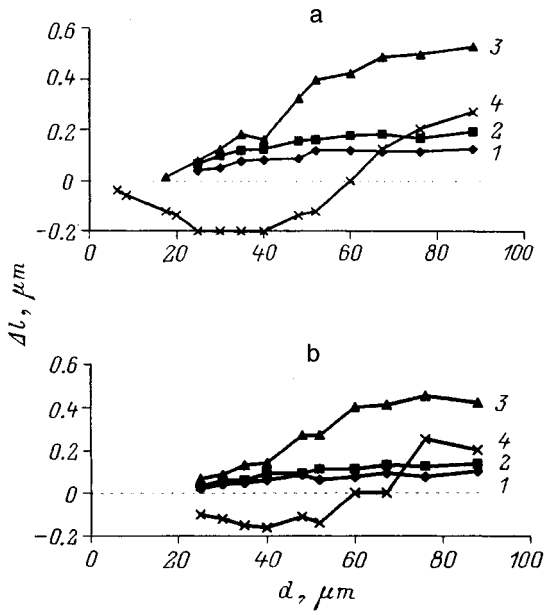

FIG. 2. Amplitude of modulation of the periodic relief at the surface of PMMA films as a function of the spatial period after annealing for 95 (*1*) and 320 min (*2*) at 326 K and for 20 (*3*) and 515 min (*4*) at 364 K. Molecular weight of PMMA 200 000 (a) and 42 000 (b).

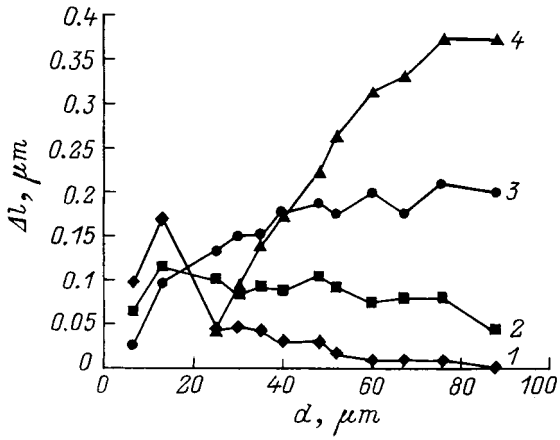

FIG. 3. Amplitude of modulation of the periodic relief at the surface of PBMA films as a function of the spatial period after annealing for 0.5 (*1*), 1 (*2*), 4.5 (*3*), and 13.5 h (*4*) at 308 K.

ing the molecular refractions of 9-anthraldehyde and its photoproduct (photodimer),[3] we determined the degree of photoconversion. The holograms were reconstructed using an He–Ne laser, whose radiation is not absorbed by 9-anthraldehyde, which first eliminated any further photoconversion and second, allowed us to consider the holograms to be pure phase ones. During the reconstruction of reflection holograms, the influence of light reflected by the rear boundary of the substrate was eliminated by its wedge shape.

Photoreliefs can be formed on the surface of PMMA layers (the glass-transition temperature of the pure polymer[4] is $T_g = 373$ K) at temperatures both close to and considerably lower than T_g. Figure 2 gives the modulation amplitude Δl for reliefs obtained using the optical test pattern ($d \geqslant 25\ \mu$m) plotted as a function of the spatial period for various annealing regimes. For $d > 70\ \mu$m the kinetics of the modulation amplitude is similar to that of the height of the nonperiodic relief described in Ref. 1. During annealing at 326 K the amplitude increases monotonically to around 1% of the initial layer thickness (with $\approx 80\%$ photoconversion). An increase in temperature to 364 K leads to a trebling of the amplitude followed by a drop caused by decomposition of the photodimers[1] (Fig. 1, curves *1* and *2* and Fig. 2, curves *3* and *4*). For smaller periods ($d \sim 25\ \mu$m) Δl increases monotonically at the low-temperature annealing stage, but does not exceed 0.2–0.3% of the layer thickness. An increase in temperature initially barely increases Δl but then causes a steeper drop and a reversal of the sign of this amplitude (inversion of the relief) (Fig. 1, curves *1′* and *2′* and Fig. 2, curves *1–4*).

At the maximum positive modulation amplitude of the reliefs we observe an appreciable (tenfold) drop in Δl when moving to smaller periods in the range 70–25 μm (Figs. 2a and b, curves *3*). This drop in Δl is the result of a reduction in the depth of the dip (Fig. 1, curve *1′*), i.e., it can be represented as the result of the unexposed material being ''entrained'' by the expanding exposed material. The overall relief height l is appreciable ($\approx 0.3\ \mu$m) even when Δl tends to 0. The drop in Δl is caused by approximately the same increase in the thickness of the unexposed regions and decrease (compared with the maximum possible) in the exposed regions.

The decrease in the amplitude Δl and the ensuing limitation of the resolution of the periodic reliefs are probably caused by shear stresses. The relatively rapid expansion induces shear deformations and consequently shear stresses at the boundary of the expanding exposed material and leads to entrainment of the unexposed polymer. Naturally, the effect is intensified as the period decreases. In this interpretation the effect is similar to induced plasticity.[5] It may then be possible to suppress the effect by reducing the mechanical moduli and the relaxation times of the stresses. This can easily be achieved by increasing the annealing temperature to values close to or exceeding T_g. However, heating the matrix above 350 K causes thermolysis of photodimers[6] and thus cannot be used directly in the PMMA layers.

Annealing at temperatures exceeding T_g without thermolyzing the photodimers was achieved in PBMA layers having a glass-transition temperature of 293 K (Ref. 4). Figure 3 gives the modulation amplitude of the relief as a function of the period after short-term annealing at 308 K. The most significant difference compared with the PMMA layers is the considerably larger range of reproducible spatial frequencies (whose maximum increases by more than an order of magnitude). A reduction in the shear stresses is evidently caused by the polymer being transferred from the glassy to the hyperelastic state. Another difference is the increase in Δl with decreasing d at the initial stage of annealing (Fig. 3, curves *1* and *2*). This behavior is evidently attributable to the increasing contribution of cold flow[5] to the process of relief formation. The exposed polymer becomes deformed, but keeps its specific volume, which causes shrinkage of the unexposed material (Fig. 1, curve *1″*). Under longer-term annealing (up to 40 h) the modulation amplitude only increases for large

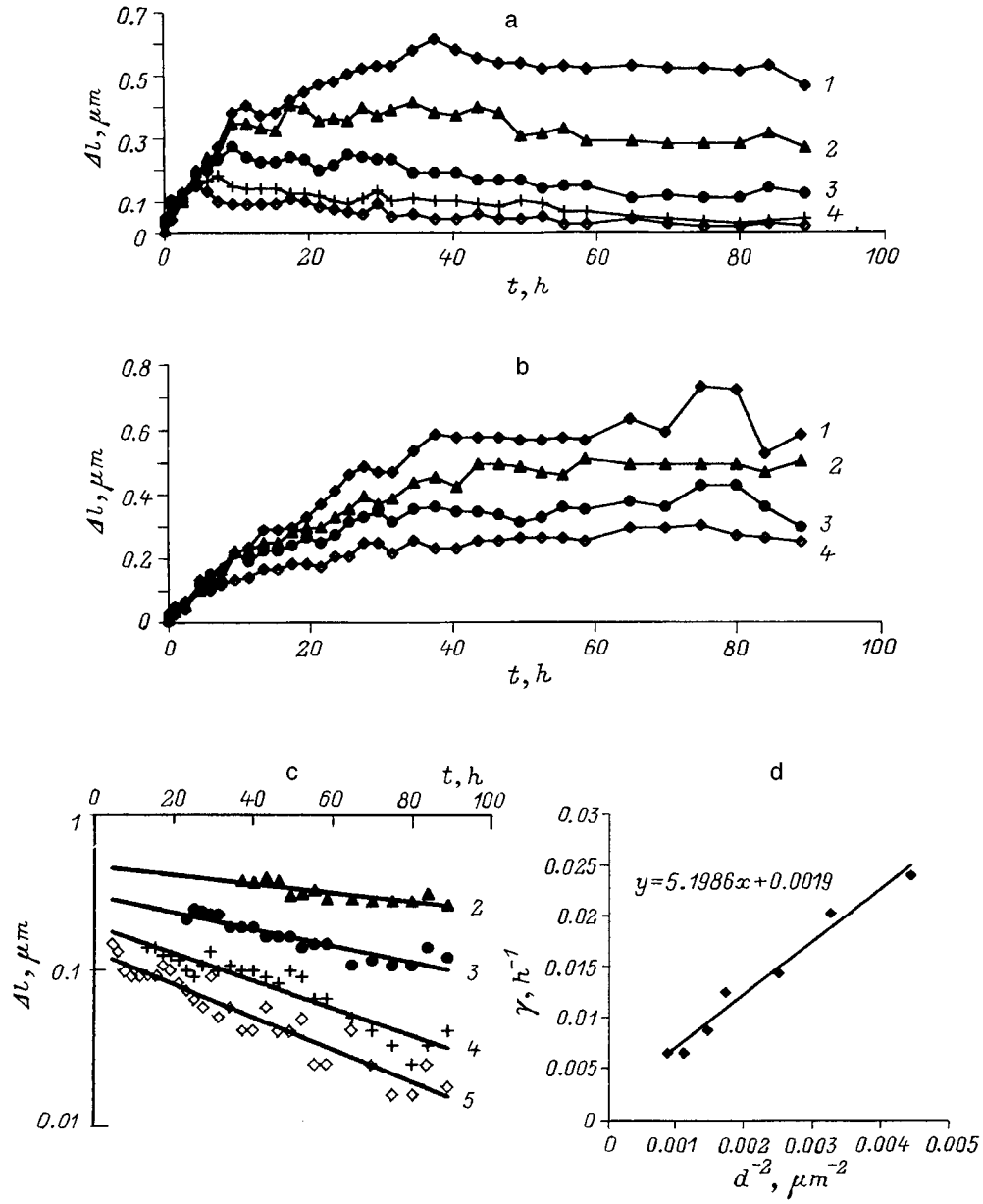

FIG. 4. Kinetics of the modulation amplitude (a) and the relief height (b) at the surface of PBMA layers at 308 K for spatial periods of 88 (*1*), 67 (*2*), 48 (*3*), 35 (*4*), 30 (*5*) μm; c — approximation of descending sections of the curves in Fig. 4, a — by exponential functions; d — dependence of γ on d^{-2} demonstrating the diffusion behavior of the relief degradation.

periods and begins to decrease for small periods, passing through a maximum. Further annealing (for more than 40 h) leads to a reduction in Δl for all spatial periods (Fig. 4a). At higher temperatures this decay takes place within shorter times.

Potential causes of this degradation of the periodic relief are residual shear stresses, surface tension forces, and also diffusion of 9-anthraldehyde and its photodimer. Diffusion of 9-anthraldehyde increases its concentration near the photorelief maxima, i.e., cannot cause a reduction in Δl. Thus, any diffusion degradation of the relief can only be attributed to migration of the photodimer.

In cases where the initial sinusoidal distribution of the photoproduct is destroyed by diffusion, the modulation amplitude of its concentration Δc will vary with time as given by[7]

$$\Delta c(t)=\Delta c_0 \exp(-\gamma t), \tag{1}$$

where Δc_0 is the initial amplitude of the modulation of the concentration, $\gamma=D(2\pi/d)^2$, and D is the diffusion coefficient.

For a rectangular initial distribution, we can transform the right-hand side of Eq. (1) into a sum of exponential functions whose leading term, as before, has the form (1).[7] Assuming that $\Delta l \sim \Delta c(t)$, we obtain the exponential kinetics of the relief modulation amplitude

$$\Delta l(g)=\Delta l_0 \exp(-\gamma t). \tag{2}$$

A similar time dependence is observed for mechanical reliefs on the surface of plastic materials which deteriorate under the action of surface tension forces and possibly residual mechanical stresses. The fundamental difference here is the dependence $\gamma \sim d^{-1}$ (Ref. 8). Approximating the descending sections of the t dependences of Δl with exponential functions (Fig. 4c), we established that γ depends linearly on d^{-2} (Fig. 4d) which argues in favor of the diffusion mechanism of relief degradation. The slope of the line in Fig. 4d gives the calculated diffusion coefficient $D \approx 3.5 \times 10^{-17}\,\mathrm{m^2/s}$ at 308 K. Monitoring the diffraction efficiency of the transmission holograms revealed that Δn decreases to zero over the degradation times of the periodic relief, which provides direct evidence in support of the diffusion mechanism. As a result of the photodimer concentration being equalized, the overall relief height was half that of the maximum.

We have therefore established that using the hyperelastic state of polymers can substantially reduce the limiting role of shear stresses, although the formation of a relief under these conditions is impeded by the evolving material diffusion of the photoproduct. It has been noted that on the surface of the PMMA layers reliefs may also form near T_g (Fig. 2, curve *3*). Recording transmission phase holograms showed that for $d \geqslant 6\ \mu$m, no appreciable diffusion effects occur under these conditions. However, the high rate of the process and the ensuing absence of any relaxation stresses cause an appreciable drop in the frequency characteristic of the reliefs at spatial frequencies exceeding $30\,\mathrm{mm^{-1}}$. The only method of varying the rate of relief formation caused by photodimerization is to vary the annealing temperature. A reduction in temperature, while slowing the relief formation, also increases the stress relaxation time, so that the resolution cannot be increased. In addition, as the temperature decreases, the induced plasticity threshold is reached, so that at $T < 315$ K no relief forms at all.

Slow deformation without any reduction in temperature was achieved during the thermal decomposition of the photodimers.[6] Figures 2a,b give Δl as a function of the period d under high-temperature (364 K) annealing (curves *3* and *4*). Decomposition of the dimers leads to erasure of the relief at large periods and the appearance of an antiphase periodic relief with $d < 35\ \mu$m (Fig. 1, curves *1*,*1′*,*2*, and *2′*). The different behavior of the relief at the formation and degradation stage is attributable to the significantly different rate of these processes. Shrinkage of the exposed regions is limited by the rate of photodimer decomposition and is considerably slower than the expansion at the same temperature. This creates conditions for the relaxation of shear stresses below the induced plasticity threshold and for the conservation of strains in the exposed zones. Slow compression can create an antiphase photorelief with its amplitude maximum shifted toward higher spatial frequencies. We established by means of a holographic method that the resolution of the periodic photorelief increased fivefold when its profile was inverted (Fig. 2a, curve *4*). The almost complete absence of any periodic photorelief at the beginning of high-temperature annealing and the appearance of an antiphase relief at the same temperature (and consequently the same surface tension) means that the surface tension cannot be considered to be a significant factor under these conditions.

The influence of the molecular weight of PMMA on the relief formation processes confirms the proposed mechanism (Fig. 2a,b). In fact, a linear polymer having a lower molecular weight has a lower glass-transition temperature,[4] and consequently the mechanical relaxation processes are faster. This then increases Δl for small periods and reduces the maximum negative amplitude for inversion of the relief (Fig. 2b).

Hence, the main factor limiting the resolution of periodic relaxation photoreliefs at the surface of glassy polymer layers containing dimerizing anthracene derivatives are the shear stresses which build up during annealing. Their action is suppressed by the thermal decomposition of photodimers, which forms inverted reliefs of higher spatial frequencies. Hyperelastic polymer matrices can increase the resolution by more than an order of magnitude. A factor limiting the stability of the photoreliefs in this case is their diffusional degradation.

[1] V. V. Mogil'nyĭ and Yu. V. Gritsaĭ, Pis'ma Zh. Tekh. Fiz. **21**(24), 66 (1995) [Tech. Phys. Lett. **21**, 1024 (1995)].
[2] K. K. Shvarts, *Physics of Optical Recording in Semiconductors and Dielectrics* [in Russian], Zinatne, Riga (1986), 230 pp.
[3] V. V. Mogil'nyĭ and Yu. V. Gritsaĭ, Opt. Spektrosk. **83**, 832 (1997) [*sic*].
[4] J. F. Rabek, *Experimental Methods in Polymer Chemistry*, Part 2 [Wiley, New York, 1980; Nauka, Moscow, 1983, 480 pp.].
[5] J. D. Ferry, *Viscoelastic Properties of Polymers*, 2nd ed. [Wiley, New York, 1970; IL, Moscow, 1963, 535 pp.].
[6] W. J. Tomlinson, E. A. Chandross, R. L. Fork *et al.*, Appl. Opt. **11**, 533 (1972).
[7] A. I. Raĭchenko, *Mathematical Theory of Diffusion and Its Applications* [in Russian], Naukova Dumka, Kiev (1981), 396 pp.
[8] Yu. A. Shevlyakov, N. G. Nakhodkin, and V. P. Nemtsev, Ukr. Fiz. Zh. **13**, 1737 (1968).

Translated by R. M. Durham

Effect of pulsed excitation of a bounded medium on a plane electromagnetic wave

A. G. Nerukh, O. N. Rybin, and I. V. Shcherbatko

Kharkov State Technical University of Radio Electronics, 310726 Kharkov, Ukraine
(Submitted December 9, 1997)
Zh. Tekh. Fiz. **69**, 84–92 (August 1999)

An analysis is made of the transformation of a plane monochromatic electromagnetic wave in response to a temporal change in the permittivity and conductivity of a semibounded medium. The change in the parameters of the medium takes the form of a rectangular pulse of arbitrary duration and amplitude. The detailed structure of the electric field and its evolutional redistribution are determined. The asymptotic formation of a backward wave is demonstrated, whose amplitude may exceed that of the primary wave for parameters typical of a semiconductor. © *1999 American Institute of Physics.* [S1063-7842(99)01608-6]

INTRODUCTION

The transformation of electromagnetic waves in media with time-varying parameters has attracted widespread interest because of its numerous applications (such as the propagation of electromagnetic waves in a nonsteady-state medium, radio communications, geophysical probing, probing of unstable objects by short electromagnetic pulses, amplification and conversion technology, and so on). The need to solve these problems also arises in studies of ultrafast electromagnetic phenomena in semiconductor and quantum electronics. For instance, the generation of optical pulses in semiconductor structures may be controlled at such a rate that the accompanying transient processes can no longer be ignored.[1]

The transformation of electromagnetic fields accompanying time variations in the parameters of a medium have been studied by many authors (for example, Refs. 2–6) with the most detailed studies usually being made for an abrupt change in a single parameter.[7,8] In the present paper we investigate how a plane monochromatic electromagnetic wave is influenced by a synchronous pulsed change in the permittivity $\varepsilon(t)$ and conductivity $\sigma(t)$ of a medium which takes place in the half space $x \geqslant 0$ and begins at zero time. Up to zero time the medium is homogeneous, nonabsorbing, and has permittivity ε_0. The pulse is rectangular and has a duration τ, so that in the half space $x \geqslant 0$ the change in the parameters of the medium is described by

$$\varepsilon(t)=\varepsilon_0[\Theta(-t)+\Theta(t-\tau)]+\varepsilon_1[\Theta(t)-\Theta(t-\tau)],$$
$$\sigma(t)=\sigma_1[\Theta(t)-\Theta(t-\tau)], \tag{1}$$

where ε_1 and σ_1 are the permittivity and conductivity of the medium in the half space $x \geqslant 0$ over the time interval $t \in [0,\tau]$, $\Theta(t)$ is the Heaviside unit step function, and the medium in the half space $x<0$ remains nonconducting throughout and has the permittivity ε_0.

If $E_0(t,x)$ is the primary field in the unperturbed medium, the electromagnetic field in a nonsteady-state medium occupying the half space $x \geqslant 0$ is described by a Volterra integral equation of the second kind, which in this case has the form[4]

$$E(t,x)=E_0(t,x)+\int_0^\infty dt' \int_0^\infty dx' K(t,t',x,x')E(t',x'). \tag{2}$$

Here $K(t,t',x,x')$ is the kernel of the integral equation,

$$K(t,t',x,x')=-(1/\overline{a^2}(t))\left\{\overline{\sigma}(t)+(1/2)(1-\overline{a^2}(t))\frac{\partial}{\partial t}\right\}\times\delta(\nu_0(t-t')-|x-x'|), \tag{3}$$

$\overline{a}(t)=\sqrt{\varepsilon_0/\varepsilon(t)}$, $\nu_0=c/\sqrt{\varepsilon_0}$, $\overline{\sigma}(t)=2\pi\sigma(t)/\varepsilon_1$; c is the speed of light in vacuum, and $\delta(t)$ is the Dirac delta function.

The solution of Eq. (2) in the region $x \geqslant 0$ is written in terms of the resolvent $R(t,t',x,x')$ by means of the integral

$$E(t,x)=E_0(t,x)+\int_0^\infty dt' \int_0^\infty dx' R(t,t',x,x')E_0(t',x'). \tag{4}$$

The resolvent can be obtained from the equation

$$R(t,t',x,x')=K(t,t',x,x')+\int_0^\infty dt'' \int_0^\infty dx'' K(t,t'',x,x'')R(t'',t',x'',x') \tag{5}$$

and for the case $\varepsilon(t)=\varepsilon_1=\text{const}$, $\sigma(t)=\sigma_1=\text{const}$ consists of two terms,

$$R(t,t',x,x')=R_1(t,t',x,x')+R_2(t,t',x,x'), \tag{6}$$

which are determined by means of an inverse Laplace transformation as follows:

$$R_n(t,t',x,x')=\exp(-\overline{\sigma}_1(t-t'))\int_{\alpha-i\infty}^{\alpha+i\infty}(dp/2\pi i)\times S_n(p,t,t'x,x')\exp(p(t-t')), \quad n=1,2.$$

Here we find

$$S_1(p)=\frac{a}{2\nu_0}\left(a^2\frac{p-\overline{\sigma_1}}{p+\overline{\sigma_1}}-1\right)\sqrt{p^2-\overline{\sigma_1^2}}\times\exp\left\{-pt'-\frac{|x-x'|}{\nu_1}\sqrt{p^2-\overline{\sigma_1^2}}\right\},$$

$$S_2(p)=\frac{a}{2\nu_0}\left(a\sqrt{\frac{p-\overline{\sigma_1}}{p+\overline{\sigma_1}}}-1\right)^2\sqrt{p^2-\overline{\sigma_1^2}}\times\exp\left\{-pt'-\frac{x+x'}{\nu_1}\sqrt{p^2-\overline{\sigma_1^2}}\right\}, \quad (7)$$

where $a=\sqrt{\varepsilon_0/\varepsilon_1}$, $\alpha>\overline{\sigma_1}$, $\overline{\sigma_1}=2\pi\sigma_1/\varepsilon_1$, $\nu_1=c/\sqrt{\varepsilon_1}$, and $\mathrm{Re}\sqrt{p^2-\overline{\sigma_1^2}}>0$.

The first term in Eq. (6) is the resolvent of the unbounded problem, and the second term allows for the influence of the boundary formed at zero time and separating the steady-state and nonsteady-state half spaces. Hence, expression (4) with the resolvent (6) determines the field inside the nonsteady-state medium after an abrupt change in its parameters to the values $\overline{\sigma_1}$ and ε_1 over the entire time interval $t\in[0,\tau]$ of the excitation of the medium.

In the region $x<0$ the field is determined using the same relation (2) which in this case in an integral formula expressing the external field in terms of the internal one.

FIELD IN A REGION OF EXCITED MEDIUM IN THE TIME INTERVAL $[0,\tau]$

Let us assume that the primary field is a plane monochromatic wave $E_0(t,x)=\exp\{i(\omega t-kx)\}$, $k=\omega/\nu_0$. After the abrupt excitation of the medium (i.e., after a jump in the permittivity and conductivity of the medium in the region $x\geqslant 0$), in the time interval $t\in[0,\tau]$ the field is described by the integral obtained from formula (4) by substituting the expressions (6) and (7)

$$E_2(t,x)=E_1(t,x)-a\int_{\alpha-i\infty}^{\alpha+i\infty}\frac{dp}{2\pi i}\frac{p-\overline{\sigma_1}+i\omega}{p-\overline{\sigma_1}-i\omega}\times\frac{a(p-\overline{\sigma_1})-\sqrt{p^2-\overline{\sigma_1^2}}}{p^2-\overline{\sigma_1^2}+k^2\nu_1^2}\times\exp((p-\overline{\sigma_1})t-(x/\nu_1)\sqrt{p^2-\overline{\sigma_1^2}}),$$

$$\mathrm{Re}(\sqrt{p^2-\overline{\sigma_1^2}})>0. \quad (8)$$

Here $E_1(t,x)$, which is the result of integrating the first part of the resolvent, R_1, is the sum of two plane waves

$$E_1(t,x)=B_1^+\exp(-\overline{\sigma_1}t)\exp\{i(\Omega t-kx)\}+B_1^-\exp(-\overline{\sigma_1}t)\exp\{-i(\Omega t+kx)\} \quad (9)$$

having the amplitudes

$$B_1^{\pm}=-a^2[(a^2-1)\omega^2\pm 2\overline{\sigma_1}(i\Omega\mp\overline{\sigma_1})]\times\{2i\Omega[i\Omega\mp(\overline{\sigma_1}+i\omega)]\}^{-1}$$

and the frequency $\Omega=\sqrt{a^2\omega^2-\overline{\sigma_1^2}}$.

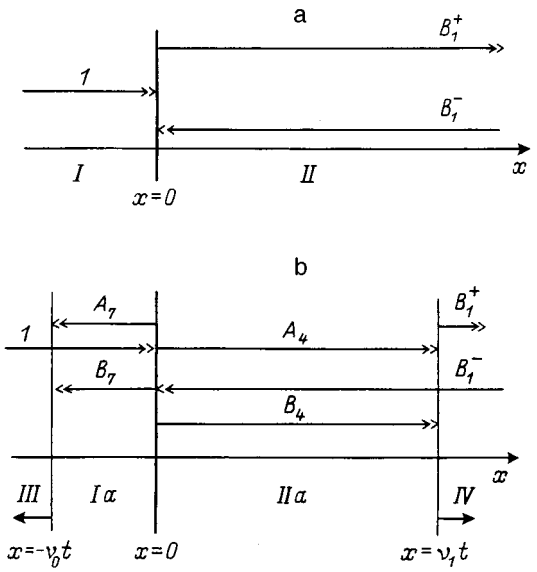

FIG. 1.

The waves (9) are also the transformed field for an unbounded medium and comprise the well-known result,[7,8] i.e., a primary plane wave, wave *1* in Fig. 1 (here and below we will use the term ''waves'' to mean their amplitude coefficients). After an abrupt change in the parameters of the medium, wave *1* initially splits into two plane waves: forward and backward. The forward wave [the first term in expression (9), B_1^+ in Fig. 1a] propagates in the same direction as the primary wave $E_0(t,x)$, while the backward wave [second term in expression (9), B_1^- in Fig. 1a] propagates toward the interface formed between the two media, $x=0$. These waves have the same wave number as the primary wave but a new frequency Ω. In cases of high conductivity $\overline{\sigma_1}>k\nu_1$, the frequency becomes imaginary and the transformed wave becomes aperiodic. Both waves decay exponentially with time as a result of the appearance of the conductivity σ_1 in the region $x\geqslant 0$. In this case of a bounded medium, expression (9) describes the entire transformed wave in the region $x\geqslant 0$ immediately after the excitation of the medium.

The influence of the boundary of the excited half space is given by the integral term in formula (8). This is only nonzero in the band $x\in[0,\nu_1 t]$, which is the region of influence of the boundary of the nonsteady-state half space on the field. For $t\to\infty$ the poles of the integrand in formula (8) immediately give steady-state waves since the integrals over the sections asymptotically go to zero. However, the entire transient process remains outside the field of view and in cases of short-lived excitation of the medium, important information on the field structure is lost. Thus, the evolution of the field at all stages of variation in the medium is of interest. For this purpose we shall analyze approximations of expression (8) in two limiting cases: $x\approx\nu_1 t$ $(x<\nu_1 t)$, i.e., the field near the moving frontier of influence of the boundary, and $x\ll\nu_1 t$, i.e., late-time approximation. We rewrite expression (8) in the form

$$E_2(t,x)=E_1(t,x)+A_2e(t,x,-\overline{\sigma_1}-i\omega)+B_2^+e(t,x,-i\Omega)+B_2^-e(t,x,i\Omega)-\Theta(t-x/\nu_1)aI_0\times(\overline{\sigma_1}\sqrt{t^2-(x/\nu_1)^2})+A_3f(t,x,-\overline{\sigma_1}-i\omega)+B_3^+f(t,x,-i\Omega)+B_3^-f(t,x,i\Omega),\quad(10)$$

where $I_0(t)$ is a modified Bessel function of the first kind, and the constants are given by

$$A_2=-a^2\frac{2\omega}{(1-a^2)\omega-2i\overline{\sigma_1}},$$

$$B_2^{\pm}=\pm a^2\frac{\Omega\mp\overline{\sigma_1}}{2i\Omega}\frac{i\Omega\mp(\overline{\sigma_1}-i\omega)}{i\Omega\mp(\overline{\sigma_1}+i\omega)},$$

$$A_3=-a^3\frac{2i\omega^2}{(1-a^2)\omega-2i\overline{\sigma_1}},$$

$$B_3^{\pm}=\pm a^3\frac{\omega^2}{2i\Omega}\frac{i\Omega\mp(\overline{\sigma_1}-i\omega)}{i\Omega\mp(\overline{\sigma_1}+i\omega)}.$$

The functions e and f in formula (10) are nonzero in the region $x<\nu_1t$ and are determined by the following expressions, in accordance with Ref. 9

$$e(t,x,\beta)=\int_{\alpha-i\infty}^{\alpha+i\infty}(dp/2\pi i)\times\exp\{pt-(x/\nu_1)\sqrt{p^2-\overline{\sigma_1^2}}\}(p+\beta)^{-1},$$

$$f(t,x,\beta)=-\int_{\alpha-i\infty}^{\alpha+i\infty}(dp/2\pi i)\times\exp\{pt-(x/\nu_1)\sqrt{p^2-\overline{\sigma_1^2}}\}\times(p+\beta)^{-1}(p^2-\overline{\sigma_1^2})^{-1/2}.\quad(11)$$

Applying the Efros[10] theorem to these functions, we rewrite expression (10) in the form

$$E_2(t,x)=E_1(t,x)+\Theta(t-x/\nu_1)\exp(-\overline{\sigma_1}t)\times\Big\{A_2\exp\{(\overline{\sigma_1}+i\omega)(t-x/\nu_1)\}+B_2^+\exp\{i\Omega(t-x/\nu_1)\}+B_2^-\exp\{-i\Omega(t-x/\nu_1)\}-aI_0(\overline{\sigma_1}\sqrt{t^2-(x/\nu_1)^2})+\frac{\overline{\sigma_1}x}{\nu_1}\times\int_{x/\nu_1}^{t}d\xi(A_2\exp\{(\overline{\sigma_1}+i\omega)(t-\xi)\}+B_2^+\exp\{i\Omega(t-\xi)\}+B_2^-\exp\{-i\Omega(t-\xi)\})\times I_1(\overline{\sigma_1}\sqrt{t^2-(x/\nu_1)^2})(t^2-(x/\nu_1)^2)^{-1/2}-\int_{x/\nu_1}^{t}d\xi(A_3\exp\{(\overline{\sigma_1}+i\omega)(t-\xi)\}+B_3^+\exp\{i\Omega(t-\xi)\}+B_3^-\exp\{-i\Omega(t-\xi)\})I_0\times(\overline{\sigma_1}\sqrt{t^2-(x/\nu_1)^2})\Big\}.\quad(12)$$

Formula (12) completely describes the electromagnetic field distribution in the half space after excitation of the medium in the time interval $t\in[0,\tau]$.

Near the frontier of influence of the boundary, the approximate expression for the field to within the first order of smallness with respect to $\overline{\sigma_1}(t-x/\nu_1)$ has the form

$$E_2(t,x)\approx E_1(t,x)+\{\exp(-\overline{\sigma_1}x/\nu_1)F_1(x)\exp\{i(\omega t-kx)\}+\exp(-\overline{\sigma_1}t)F_2(x)\exp\{i\Omega(t-x/\nu_1)\}+\exp(-\overline{\sigma_1}t)F_3(x)\exp\{-i\Omega(t-x/\nu_1)\}+\exp(-\overline{\sigma_1}t)F_4(x)\}.\quad(13)$$

Here we have

$$F_1(x)=2a\omega((1-a^2)\omega-2i\overline{\sigma_1})^{-1}(\overline{\sigma_1}+i\omega)^{-1}\times(\overline{\sigma_1}+i(1-a)\omega+\overline{\sigma_1^2}x/\nu_1),$$

$$F_2(x)=-\frac{a^2}{2i\Omega}\frac{\overline{\sigma_1}-i(\omega+\Omega)}{\overline{\sigma_1}+i(\omega-\Omega)}\times(\overline{\sigma_1}-i\Omega+i\Omega^{-1}a\omega^2+i\Omega^{-1}\times(\overline{\sigma_1}-i\Omega)\overline{\sigma_1^2}x/\nu_1),$$

$$F_3(x)=-B_4(\overline{\sigma_1}+i\Omega+i\Omega^{-1}a\omega^2+i\Omega^{-1}(\overline{\sigma_1}-i\Omega)\overline{\sigma_1^2}x/\nu_1)(\overline{\sigma_1}-i(\omega-\Omega))^{-1},$$

$$B_4=a^2i(2\Omega)^{-1}[\overline{\sigma_1}-i(\omega-\Omega)]^2[\overline{\sigma_1}+i(\omega+\Omega)]^{-1},$$

$$F_4(x)=-a+a((1-a^2)\omega-2i\overline{\sigma_1})^{-1}(a^2(1+a^2)\times\omega^3\Omega^{-2}-2ia^2\omega^2(\overline{\sigma_1}+i\omega)^{-1}-(\overline{\sigma_1^2}/2)\times[2\omega(\overline{\sigma_1}+i\omega)^{-1}+i\Omega^{-1}((1+a^2)\times\omega+2i\overline{\sigma_1})]x/\nu_1).$$

It can be seen from expression (13) that a jump in the parameters of the medium in a bounded region leads to the appearance of a wave spectrum having nonuniform amplitudes near the frontier of influence of the boundary. The first wave $\sim F_1(x)$ is an analog of an ordinary transmitted wave. The wave $\sim F_2(x)$ is a ''quasi-antipode'' of the wave B_1^+ from the field E_1 (9), since its phase has the form $i\Omega\times(t-x/\nu_1)=i(\Omega t-\sqrt{1-a^2(\overline{\sigma_1}/\omega)^2}kx)$. It will be shown below that this wave recombines asymptotically with the wave B_1^+. It should be noted that in the absence of a conductivity jump, no continuous spectrum forms, and these waves recombine almost immediately. The third wave $\sim F_3(x)$ is the result of the wave B_1^- being reflected from the interface formed between the two media. Hence, the pres-

ence of a conductivity jump in a bounded medium gives rise to a transient process which brings the field to a steady state only after a certain time.

It also follows from expressions (9) and (13) that a jump in the permittivity of the medium in the half space gives a field discontinuity at the moving plane of influence of the boundary

$$x=\nu_1 t: \quad E_2(x/\nu_1+0,x)-E_2(x/\nu_1-0,x)$$
$$=a(a-1)\exp(-\bar{\sigma}_1 t).$$

The evolution of the electromagnetic field toward the steady-state regime will be described by the late-time approximation of expression (10). In order to obtain this approximation we shall use a representation[9] of the integrals $e(t,x,\beta)$ and $f(t,x,\beta)$, in terms of Lifshitz–Hankel functions [formulas (A1) and (A2) in the Appendix]. In this representation the zeroth-order Lifshitz–Hankel function is given by a Neumann series[11] [formula (A3) in the Appendix]. Retaining only the first term in this series and substituting the late-time ($t \gg x/\nu_1$) approximations for the functions $e(t,x,\beta)$ and $f(t,x,\beta)$ into formula (10), we find that the following approximate expression for the field $E_2(t,x)$ holds for long times:

$$E_2(t,x)\approx B_1^- \exp(-\bar{\sigma}_1 t)\exp\{-i(\Omega t+kx)\}$$
$$+A_4\exp(-\bar{\sigma}_1 x/\nu_1)\exp\{i\omega(t-x/\nu_1)\}$$
$$+B_4\exp(-\bar{\sigma}_1 t)\exp\{-i(\Omega t-kx)\}$$
$$+\exp(-2\bar{\sigma}_1 t)Q_1(t,x), \qquad (14)$$

where

$$A_4A_2\frac{\bar{\sigma}_1+i(1-a)\omega}{\bar{\sigma}_1+i\omega}.$$

Thus, in the region $0 \ll x \ll \nu_1 t$ the following asymptotically remain (Fig. 1b): the forward transmitted wave A_4, initiated by wave *1*, the backward wave B_1^-, transferred from the region $x > \nu_1 t$, and the forward wave B_4 generated by its being reflected from the interface with the media. The wave A_4 decays with increasing distance from the boundary ($\sim\exp(-\bar{\sigma}_1 x/\nu_1)$), unlike the waves $B_1^{\pm}$ and B_4, which decay with time ($\sim\exp(-\bar{\sigma}_1 t)$). The absence of the wave B_1^+ in expression (14) and its presence in expression (13) indicates that this wave gradually ''disappears'' with increasing distance from the plane $x=\nu_1 t$. The term Q_1 in formula (14) is the continuous wave spectrum caused by the jump in the conductivity of the medium and we find $Q_1(t,x)\to 0$ for $t\to\infty$. An expression for Q_1 is given in Sec. 4 of the Appendix.

RESPONSE OF AN EXTERNAL FIELD TO THE EXCITATION OF A BOUNDED REGION OF MEDIUM

We obtain an expression for the electric field in the region of unperturbed medium $x \leqslant 0$ in the time interval $t \in [0,\tau]$ by substituting expression (8) into formula (2)

$$E_3(t,x)=E_0(t,x)+\Theta\left(t+\frac{x}{\nu_0}\right)\left\{\bar{\sigma}_1+\frac{1}{2}(1-a^2)\frac{\partial}{\partial t}\right\}$$
$$\times\left\{A_5\exp\left\{i\omega\left(t+\frac{x}{\nu_0}\right)\right\}\right.$$
$$+(a^2/2i\Omega+B_5^+)\exp\left\{i\Omega\left(t+\frac{x}{\nu_0}\right)\right\}$$
$$+(a^2/2i\Omega+B_5^-)\exp\left\{i\Omega\left(t+\frac{x}{\nu_0}\right)\right\}$$
$$+C_5\exp\left\{-(1+a^2)(1-a^2)^{-1}\bar{\sigma}_1\left(t+\frac{x}{\nu_0}\right)\right\}$$
$$+e\left(-\bar{\sigma}_1\left(t+\frac{x}{\nu_0}\right)\right)A_6 Z\left(t+\frac{x}{\nu_0},-\bar{\sigma}_1-i\omega\right)$$
$$+B_6^+ Z\left(t+\frac{x}{\nu_0},i\Omega\right)+B_6^- Z\left(t+\frac{x}{\nu_0},-i\Omega\right)$$
$$\left.+C_6 Z\left(t+\frac{x}{\nu_0},(1+a^2)(1-a^2)^{-1}\bar{\sigma}_1\right)\right\}. \qquad (15)$$

Here we have

$$A_5=-8a^2\frac{2\bar{\sigma}_1+i(1+a^2)\omega}{(2\bar{\sigma}_1+i(1-a^2)\omega)^2},$$

$$A_6=4a^3\frac{(\omega-2i\bar{\sigma}_1)\omega}{(2\bar{\sigma}_1+i(1-a^2)\omega)^2},$$

$$B_5^{\pm}=\pm\frac{a^2(1+a^2)i(\bar{\sigma}_1\pm i\Omega)}{2\Omega((1+a^2)\bar{\sigma}_1\pm(1-a^2)i\Omega)},$$

$$B_6^{\pm}=\pm\frac{a^5 i\omega}{(1+a^2)\bar{\sigma}_1\pm(1-a^2)i\Omega},$$

$$C_5=\frac{16a^2\bar{\sigma}_1}{(2\bar{\sigma}_1+i(1-a^2)\omega)^2}, \quad C_6=-\bar{\sigma}_1 C_5,$$

$$Z\left(t+\frac{x}{\nu_0},\beta\right)=-\int_{\alpha-i\infty}^{\alpha+i\infty}\left(\frac{dp}{2\pi i}\right)\exp\left(p\left(t+\frac{x}{\nu_0}\right)\right)$$
$$\times(p+\beta)^{-1}(p^2-\bar{\sigma}_1^2)^{-1/2}. \qquad (16)$$

An asymptotic approximation ($t \gg -x/\nu_0$) for the function $Z(t+x/\nu_0,\beta)$ can be obtained from the corresponding approximation for the function $f(t,x,\beta)$, by making the substitution $t\to t+x/\nu_0$, $x/\nu_0\to 0$ [formula (A4) in the Appendix]. Substituting the expression thus obtained for the function $Z(t+x/\nu_0,\beta)$ into formula (15), we obtain the late-time

approximation for the electric field outside the nonsteady-state half space

$$E_3(t,x)\approx E_0(t,x)+\Theta(t+x/\nu_0)\Big\{A_7\exp\{i(\omega t+kx)\}+B_7\exp\{-i\Omega(t+x/\nu_0)\}+\Big[\overline{\sigma_1}+(1/2)(1-a^2)\frac{\partial}{\partial t}\Big]Q_2(t,x)\Big\},\qquad(17)$$

where

$$A_7=-4a^2(2\overline{\sigma_1}+i(1-a^2)\omega)^{-1}(\overline{\sigma_1}+i(1+a^2)\omega-a\omega(\omega-2i\overline{\sigma_1})(2\overline{\sigma_1}+2i\omega)^{-1}),$$

$$B_7=-a^2(2i\Omega)^{-1}(2\overline{\sigma_1}-i(1-a^2)\Omega).$$

It follows from expression (17) that at zero time the plane $x=-\nu_0 t$ separates from the interface formed between the two media $x=0$ and propagates toward negative x at the velocity ν_0. This plane defines the left-hand boundary of the region of external field formed by reflection of the primary wave *1*, which results in the formation of the wave A_7, and also by the propagation of the wave B_1^- across the interface between the media, resulting in the formation of the wave B_7 (Fig. 1b). The phase velocities of these waves are the same as the phase velocity of the primary wave. The frequency and wave number of the wave A_7 are the same as the corresponding characteristics of the primary wave, while the frequency and wave number of the wave B_7 are Ω and Ω/ν_0, respectively. The function $Q_2(t,x)$ (Appendix Sec. 4) in formula (17) describes the continuous wave spectrum in the region $x<0$, which vanishes when $t\to\infty$.

We find the electric field in the right-hand neighborhood of the plane $x=-\nu_0 t$ by rewriting the integral (16) using the Efros theorem, in the form

$$Z(t+x/\nu_0,\beta)=-\int_0^{t+x/\nu_0}d\xi\times\exp\{-\beta(t+x/\nu_0-\xi)\}I_0(\overline{\sigma_1}\xi).\qquad(18)$$

Then, to within the first order of smallness with respect to $\omega(t+x/\nu_0)$, we obtain an approximate expression for the field $E_3(t,x)$ near the frontier of influence of the boundary $x=-\nu_0 t$

$$E_3(t,x)\approx E_0(t,x)+\Theta(t+x/\nu_0)\{A_7\exp\{i(\omega t+kx)\}+B_7\exp\{-i\Omega(t+x/\nu_0)\}-2a((2a^3-2a-1)\omega^2-(\overline{\sigma_1}+i\omega)\overline{\sigma_1})\times(2\overline{\sigma_1}+i(1-a^2)\omega)^{-2}\}.\qquad(19)$$

The last term in formula (19) is a constant component originating from the continuous wave spectrum, and it differs radically from the corresponding term in formula (17), where the continuous spectrum of the waves forms a complex function of space–time coordinates.

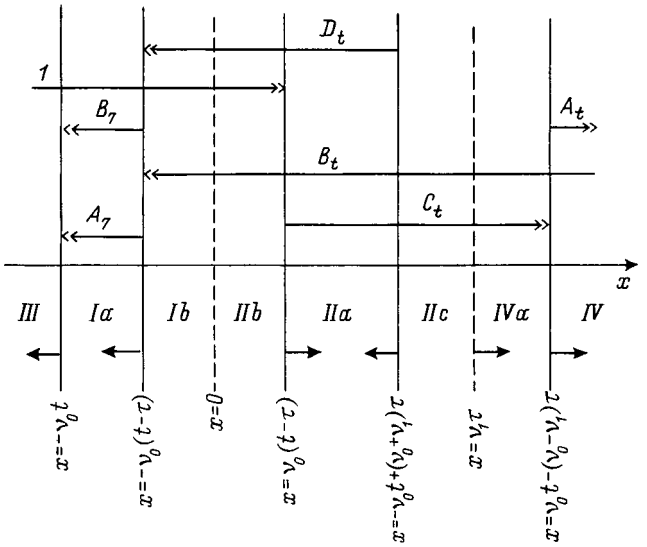

FIG. 2.

RESIDUAL EFFECTS AFTER REMOVAL OF THE EXCITATION OF THE MEDIUM

The electric field in the entire space after an abrupt return of the parameters in the region $x\geqslant 0$ to the initial state is expressed by

$$E(t,x)=E_0(t,x)+(1/a^2)\Big\{\overline{\sigma_1}+(1/2)(1-a^2)\frac{\partial}{\partial t}\Big\}\times\int_0^\tau dt'\int_0^\infty dx'\,\delta(\nu_0(t-t')-|x-x'|)E_2(t',x').\qquad(20)$$

After substituting expression (8), we find that the field consists of three groups of components occupying different regions in space (Fig. 2),

$$E(t,x)=E_0(t,x)+E_4(t,x)+E_5(t,x)+E_6(t,x).\qquad(21)$$

The first group are the waves which exist in regions *IIa*, *IIc*, and *IV* and propagate in the direction of the primary wave

$$E_4(t,x)=\Theta(x+\nu_0(\tau-t))\Big\{-\Theta(-t+x/\nu_0)E_0(t,x)+\Big[\overline{\sigma_1}+(1/2)(1-a^2)\frac{\partial}{\partial t}\Big][(\Theta(t-x/\nu_0)A_8-\Theta(-t+x/\nu_0)k^{-2}\nu_1^{-2}(B_3^+\exp(i\Omega\tau)+B_3^-\exp(i\Omega\tau)))E_0(t,x)-a^2\Omega^{-1}\times\exp(-(\overline{\sigma_1}+i\omega)\tau)\sin(\omega\tau)\exp\{-(\omega t+kx)\}+\exp\{i\Omega(t-x/\nu_0)\}(\Theta(-t+x/\nu_0)k^{-2}\nu_1^{-2}B_3^+ +\Theta(t-x/\nu_0)B_8^+)+\exp\{-i\Omega(t-x/\nu_0)\}\times(\Theta(-t+x/\nu_0)k^{-2}\nu_1^{-2}B_3^-+\Theta(t-x/\nu_0)B_8^-)+A_8e(\tau,-(1/a)(t-x/\nu_0-\tau),-\overline{\sigma_1}-i\omega)$$

$$+2a(2a\overline{\sigma_1}+(1-a^2)i\omega)(B_8^+ e(\tau,-(1/a)\times(t-x/\nu_0-\tau),i\Omega)+B_8^- e(\tau,-(1/a)\times(t-x/\nu_0-\tau),-i\Omega))]\Big\}\Theta(x). \quad (22)$$

The second group are waves propagating in the opposite direction, which exist in region *Ia*:

$$E_5(t,x)=\left[\overline{\sigma_1}+(1/2)(1-a^2)\frac{\partial}{\partial t}\right]\{A_5E_0(t,x)+B_5^+\exp\{i\Omega(t+x/\nu_0)\}+B_5^-\exp\{-i\Omega(t+x/\nu_0)\}+C_5\exp\{2|1-a^2|^{-1}\overline{\sigma_1}(t+x/\nu_0)\}-\Omega^{-1}\exp(-\overline{\sigma_1}(t+x/\nu_0))-\sin(\Omega(t+x/\nu_0))+\exp(-\overline{\sigma_1}(t+x/\nu_0))A_6Z(t+x/\nu_0,-\overline{\sigma_1}-i\omega)+B_6^+Z(t+x/\nu_0,i\Omega)+B_6^-Z(t+x/\nu_0,-i\Omega)+C_6Z(t+x/\nu_0,(1+a^2)|1-a^2|^{-1}\overline{\sigma_1})\}\Theta(-x) \quad (23)$$

and in regions *Ib, IIa, IIb*:

$$E_6(t,x)=\left[\overline{\sigma_1}+(1/2)(1-a^2)\frac{\partial}{\partial t}\right]\times\{\Theta(-x)\exp(-(\overline{\sigma_1}-i\omega)\tau)\sin(\Omega\tau)\times\exp\{-i(\omega t+kx)\}+A_5e(\tau,(1/a)(t+x/\nu_0-\tau),-\overline{\sigma_1}-i\omega)+B_5^+e(\tau,(1/a)(t+x/\nu_0-\tau),i\Omega)+B_5^-e(\tau,(1/a)(t+x/\nu_0-\tau),-i\Omega)+C_5e(\tau,(1/a)(t+x/\nu_0-\tau),\times(1+a^2)|1-a^2|^{-1}\overline{\sigma_1})+A_6f(\tau,(1/a)(t+x/\nu_0-\tau),-\overline{\sigma_1}-i\omega)+B_6^+f(\tau,(1/a)(t+x/\nu_0-\tau),i\Omega)+B_6^-f(\tau,(1/a)(t+x/\nu_0-\tau),-i\Omega)+C_6f(\tau,(1/a)(t+x/\nu_0-\tau),(1+a^2)|1-a^2|^{-1}\overline{\sigma_1})\} \quad (24)$$

Here we have

$$A_8=-\frac{1}{a}\frac{(1-a^2)\omega-2ai\overline{\sigma_1}}{(1-a^2)\omega-2i\overline{\sigma_1}},$$

$$B_8^{\pm}=\frac{i}{2\Omega}\frac{(1+a^2)\omega\pm 2i\Omega}{(1-a^2)\omega-2i\overline{\sigma_1}}.$$

Thus at time $t=\tau$, as a result of the parameters of the medium in the half space $x\geqslant 0$ returning abruptly to the initial state, two planes $x=\pm\nu_0(t-\tau)$ become detached from the boundary $x=0$ and propagate in opposite directions at the velocity ν_0. For this reason the plane $x=\nu_1 t$ propagating in the region $x\geqslant 0$ also splits into two diverging planes: $x=\pm\nu_0 t+(\nu_1\mp\nu_0)\tau$. These planes delimit regions of different field structure.

Using the approximate expressions given above for the integrals (11) and (16), we obtain the late-time approximation ($t\gg x/\nu_1$) for the field (21)

$$E(t,x)\approx A_t(t,x)E_0(t,x)+B_t(t,x)\exp\{-i(\omega t+kx)\}+C_t(t,x)\exp\{(i/a)(\omega t-kx)\}+D_t(t,x)\exp\{-(i/a)(\omega t+kx)\}+\Theta(\nu_0(\tau-t)-x)E_3(t,x)+\sum_{i=3}^{4}\Theta((\nu_1+(-1)^i\nu_0)\tau+(-1)^{i-1}\nu_0 t-x)\times\Theta(x+(-1)^i\nu_0(\tau-t))\exp(-\overline{\sigma_1}\tau)Q_i(t,x,\tau). \quad (25)$$

The field components in this expression have the following origin. Each of the waves B_1^-, B_4, and B_1^+ splits into forward and backward waves. The resulting waves, having the same phase velocities, interfere to form a forward wave

$$A_t(t,x)=\exp(-(\overline{\sigma_1}+i\omega)\tau)(\overline{\sigma_1}-(1-a^2)i\omega/2)\times(B_8^+\exp(i\Omega t)+B_8^-\exp(-i\Omega t)) \quad (26)$$

and a backward wave B_t

$$B_t(t,x)=-a^2(2\Omega)^{-1}(2\overline{\sigma_1}-i(1-a^2)\omega)\Theta(x+\nu_0(t-\tau))\times\exp(-(\overline{\sigma_1}-i\omega)\tau)\{\cos(\Omega\tau)+i\Theta((\nu_1+\nu_0)\tau-\nu_0 t-x)\sin(\Omega\tau)\}. \quad (27)$$

The phase characteristics of these waves are the same as those of the primary wave. The forward wave A_t only exists in region *IV*. The backward wave B_t, in regions *IIa, Ib*, and *IIb* has the amplitude

$$-(a^2/2\Omega)(2\overline{\sigma_1}-i(1-a^2)\omega)\exp(-(\overline{\sigma_1}-i\omega)\tau)\sin(\Omega\tau), \quad (28)$$

and in regions *IV, IIc*, and *IVa*

$$-(a^2/2\Omega)(2\overline{\sigma_1}-i(1-a^2)\omega)\exp(-(\overline{\sigma_1}-i(\omega+\Omega))\tau). \quad (29)$$

Splitting of the wave A_4 gives the forward wave C_t which exists in regions *IIa, IIc*, and *IVa*

$$C_t(t,x)=\Theta(x+\nu_0(\tau-t))\Theta((\nu_1-\nu_0)\tau+\nu_0 t-x)\times\exp(-(\overline{\sigma_1}+i(1+1/a)\omega)\tau)A_8 \quad (30)$$

and the backward wave D_t, which exists in regions *Ia, IIa*, and *IIb*,

$$D_t(t,x)=\Theta(x+\nu_0(t-\tau))\Theta((\nu_1+\nu_0)\tau-\nu_0 t-x)\times\exp(-(\overline{\sigma_1}-i(1+1/a)\omega)\tau)A_7. \quad (31)$$

These waves have frequencies and wave numbers equal to ω/a and k/a, respectively.

The terms Q_i in expression (25) describe the spatial distribution of the continuous wave spectrum, which appears at time $t=\tau$ and disappears for $t\to\infty$. The expression for these terms is given in Sec. 4 of the Appendix.

The term $E_3(t,x)$ in formula (25) describes the field in region *Ia* (Fig. 2) and is the same as the corresponding expression for the field in this region (Fig. 1b) before the medium returns abruptly to the initial state. This implies that the field structure is conserved in regions *Ia* and *III*.

Thus, after the medium has returned abruptly to the initial state in space two pairs of forward and backward waves form: A_t, B_t and C_t, D_t (Fig. 2) and also a continuous wave spectrum. The frequencies and wave numbers of A_t and B_t are the same as those of the primary wave, while the frequencies and wave numbers of C_t and D_t differ from the corresponding characteristics of the primary wave by the coefficient $\sqrt{\varepsilon_1/\varepsilon_0}$. In the steady-state regime when all the interface planes go to infinity, only the primary wave and the wave B_t remain in the space. The asymptotic expression for the electric field has the following form:

$$E(t,x)=E_0(t,x)-(a^2/2\Omega)(2\overline{\sigma_1}-i(1-a^2)\omega)\sin(\Omega\tau)\times\exp(-(\overline{\sigma_1}-i\omega)\tau)\exp\{-i(\omega t+kx)\}. \quad (32)$$

The modulus and phase of the backward wave amplitude $B_t=|B_t|\exp(i\Phi)$ will be given by

$$|B_t|=\frac{a}{2}\sqrt{\frac{4b^2a^2+(1-a^2)^2}{1-b^2a^2}}\times\sin(aT\sqrt{1-b^2a^2})\exp\{-ba^2T\}, \quad (33)$$

$$\Phi=T+\arctan((a^2-1)/2ba^2), \quad (34)$$

where the normalized quantities $b=2\pi\sigma_1/(\varepsilon_0\omega)$ and $T=\omega\tau$ have been introduced.

In this model we can introduce the attenuation of an electromagnetic wave as well as its amplification. Formally, this can be described by negative conductivity. For example, when free carriers are injected into a semiconductor, the permittivity of the semiconductor changes and the sign of its conductivity is reversed.[12,13] If the refractive index of the medium is a complex quantity $n=n'+in''$, the parameters a and b are expressed in terms of the real and imaginary parts of the refractive index as follows:

$$a^2=\frac{\varepsilon_0}{n'^2-n''^2}, \quad b=\frac{T}{\varepsilon}n'n''. \quad (35)$$

For typical values of the refractive index of a semiconductor such as InGaAsP ($n'\approx3.6$, $n''\approx0.01$) (Ref. 13), the parameters values of the order of $a\approx1$, $b\approx0.05$.

The behavior of the modulus of the backward wave amplitude as a function of the permittivity jump a is plotted in Fig. 3 for various values of the dissipation coefficient b. The negative values of b model an active medium having gain. It can be seen that for any sign of the dissipation coefficient the backward wave may undergo amplification. This effect is evidently caused by the transfer of energy from the source, which varies the parameters of the medium and the electromagnetic field.

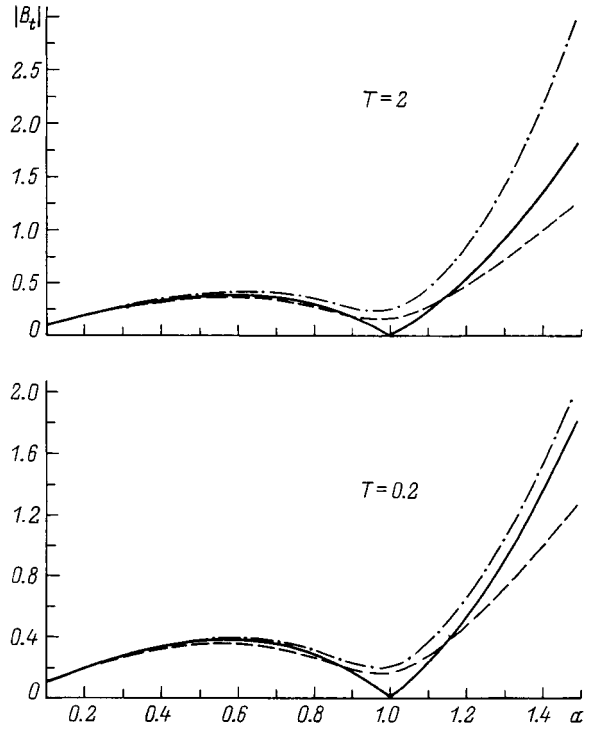

FIG. 3. Modulus of the backward wave amplitude B_t as a function of the relative change in the permittivity a. Solid curve — $b=0$, dashed curve — $b=0.1$, dot-and-dash curve — $b=-0.1$.

CONCLUSION

We have shown that the pulsed excitation of the parameters of a bounded medium over a time interval of finite duration forms a complex field structure from a primary plane monochromatic wave. This structure is a system of spatial regions separated by parallel planes propagating at velocities equal to the corresponding phase velocities of the waves. Each region is characterized by a discrete set of waves and by a continuous wave spectrum. Asymptotically, in a finite region of space, out of these sets of waves there remains a wave which is backward relative to the primary wave, having the phase characteristics of the primary wave but a different amplitude. For typical semiconductors this amplitude may exceed that of the primary wave.

APPENDIX

1. It was shown in Ref. 9 that the functions $e(t,x,\beta)$ and $f(t,x,\beta)$ can be expressed in the form

$$e(t,x,\beta)=\Theta(t-x/\nu_1)\{e^{-\beta t}\cosh((x/\nu_1)\sqrt{\beta^2-\overline{\sigma_1^2}})+(2i\overline{\sigma_1}\sqrt{t^2-(x/\nu_1)^2})^{-1}[(\beta x/\nu_1-t\sqrt{\beta^2-\overline{\sigma_1^2}})\times\exp(a_+\xi)\mathfrak{J}_{e0}(a_+,\xi)+(\beta x/\nu_1+t\sqrt{\beta^2-\overline{\sigma_1^2}})\times\exp(a_-\xi)\mathfrak{J}_{e0}(a_-,\xi)]\}; \quad (A1)$$

$$f(t,x,\beta)=\Theta(t-x/\nu_1)(\beta^2-\overline{\sigma_1^2})^{-1/2}$$
$$\times\{\exp(-\beta t)\sinh((x/\nu_1)\sqrt{\beta^2-\overline{\sigma_1^2}})$$
$$+(2i\overline{\sigma_1}\sqrt{t^2-(x/\nu_1)^2})^{-1}[(\beta x/\nu_1-t\sqrt{\beta^2-\overline{\sigma_1^2}})$$
$$\times\exp(a_+\xi)\Im_{e0}(a_+,\xi)-(\beta x/\nu_1+t\sqrt{\beta^2-\overline{\sigma_1^2}})$$
$$\times\exp(a_-\xi)\Im_{e0}(a_-,\xi)]\}; \quad \text{(A2)}$$

where $a_\pm=(-\beta t\pm(x/\nu_1)\sqrt{\beta^2-\overline{\sigma_1^2}})(i\overline{\sigma_1}\sqrt{t^2-(x/\nu_1)^2})^{-1}$, $\xi=i\overline{\sigma_1}\sqrt{t^2-(x/\nu_1)^2}$,

$$\Im_{e0}(a_\pm,z)=\int_0^z d\zeta e^{-a_\pm\zeta}\zeta^n J_n(\zeta)$$

is an nth-order Lifshitz–Hankel function and $J_n(z)$ is an nth-order Bessel function. Here we assume that $\mathrm{Re}(\sqrt{a_\pm^2+1})>0$ if $\mathrm{Re}(a_\pm)\geqslant 0$ and $\mathrm{Re}(\sqrt{a_\pm^2+1})<0$ if $\mathrm{Re}(a_\pm)<0$.

2. The zeroth-order Lifshitz–Hankel function $\Im_{e0}(a_\pm,z)$ is given by the Neumann series[11]

$$\Im_{e0}(a_\pm,z)=(a_\pm^2+1)^{-1/2}+\frac{\exp(-a_\pm z)}{\Gamma(1/2)(a_\pm^2+1)}$$
$$\times\sum_{k=0}^{\infty}(2(a_\pm^2+1)^{-1}z^{-1})^k$$
$$\times\Gamma(k+1/2)[J_{k+1}(z)-a_\pm J_k(z)]. \quad \text{(A3)}$$

3. The late-time approximation for the function $Z(t+x/\nu_0,\beta)$ has the following form

$$Z(t+x/\nu_0,\beta)\approx-\Theta(t+x/\nu_0)(\beta^2-\overline{\sigma_1^2})^{-1/2}$$
$$\times\left\{\exp(-\overline{\sigma_1}(t+x/\nu_0))\right.$$
$$-\frac{i\overline{\sigma_1}}{\beta^2-\overline{\sigma_1^2}}\frac{t\sqrt{\beta^2-\overline{\sigma_1^2}}}{t+x/\nu_0}[I_1(\overline{\sigma_1}(t+x/\nu_0))$$
$$\left.+(\beta/i\overline{\sigma_1})I_0(\overline{\sigma_1}(t+x/\nu_0))]\right\}. \quad \text{(A4)}$$

4. Functions describing the continuous wave spectra in different spatial regions

$$Q_1(t,x)=-\frac{a^2}{(1-a^2)\omega-2i\overline{\sigma_1}}\left\{\left(\frac{2i\omega\overline{\sigma_1}}{(\overline{\sigma_1}+i\omega)^2}\frac{a\omega t+(\overline{\sigma_1}+i\omega)x/\nu_1}{\sqrt{t^2-(x/\nu_1)^2}}\right.\right.$$
$$\left.-\frac{(1+a^2)a\omega^3 t-\Omega(x/\nu_1)((1+a^2)\omega\overline{\sigma_1}-2i\Omega^2)}{i\Omega\{a^2\omega^2t^2-\Omega^2(x/\nu_1)^2\}}\overline{\sigma_1}\sqrt{t^2-(x/\nu_1)^2}\right)I_1(\overline{\sigma_1}\sqrt{t^2-(x/\nu_1)^2})$$
$$\left.+\left(2\omega\frac{2(1+a^2)a\omega^3\Omega(t^2-(x/\nu_1)^2)+i\overline{\sigma_1^2}t(x/\nu_1)((1+a^2)\omega\overline{\sigma_1}+2i\Omega^2)}{\Omega(a^2\omega^2t^2-\Omega^2(x/\nu_1)^2)}\right)I_0(\overline{\sigma_1}\sqrt{t^2-(x/\nu_1)^2})\right\}. \quad \text{(A5)}$$

$$Q_2(t,x)=-i\frac{a}{2\overline{\sigma_1}+i(1-a^2)\omega}\frac{t}{t+x/\nu_0}$$
$$\times\left\{\overline{\sigma_1}\left[\frac{4a^2}{2\overline{\sigma_1}+i(1-a^2)\omega}-\frac{2a^2(1-a^2)\Omega}{\omega(2\overline{\sigma_1}-i(1-a^2)\omega)}\right]\right.$$
$$\times I_1(\overline{\sigma_1}(t+x/\nu_0))+\left[\frac{2ia^2(1+a^2)\Omega\overline{\sigma_1}}{\omega(2\overline{\sigma_1}-i(1-a^2)\omega)}\right.$$
$$-\frac{4a^2(\omega-i\overline{\sigma_1})}{2\overline{\sigma_1}+i(1-a^2)\omega}-\frac{4|1-a^4|\overline{\sigma_1}}{a(2\overline{\sigma_1}+i(1-a^2)\omega)}$$
$$\left.\left.+2\overline{\sigma_1}+i(1-a^2)\omega\right]I_0(\overline{\sigma_1}(t+x/\nu_0))\right\}; \quad \text{(A6)}$$

$$Q_3(t,x,\tau)=-\frac{ia^2(t-x/\nu_0-\tau)}{(a^2-1)\omega-2i\overline{\sigma_1}}$$
$$\times\left(\left(\frac{2i\overline{\sigma_1}}{\overline{\sigma_1}+i\omega}\frac{t-x/\nu_0-\tau}{a\sqrt{a^2\tau^2-(t-x/\nu_0-\tau)^2}}\right.\right.$$
$$-(\overline{\sigma_1}/2a)\sqrt{\tau^2-(1/a^2)(t-x/\nu_0-\tau)^2}$$
$$\left.\times\frac{((a^2+1)\omega+2i\overline{\sigma_1})}{a^4\omega^2\tau^2-\Omega^2(t-x/\nu_0-\tau)^2}\right)$$
$$\times I_1(\overline{\sigma_1}\sqrt{\tau^2-(1/a^2)(t-x/\nu_0-\tau)^2})$$
$$-I_0(\overline{\sigma_1}\sqrt{\tau^2-(1/a^2)(t-x/\nu_0-\tau)^2})(1/a)$$
$$\left.\times\left(\frac{2}{t-x/\nu_0-\tau}-\frac{\overline{\sigma_1^2}\Omega\tau((a^2+1)\omega+2ia\overline{\sigma_1})}{a^4\omega^2\tau^2-\Omega^2(t-x/\nu_0-\tau)^2}\right)\right). \quad \text{(A7)}$$

$$Q_4(t,x,\tau)=\frac{a^3}{2\bar{\sigma}_1+i(1-a^2)\omega}\Bigg(\Bigg(\frac{4i\bar{\sigma}_1}{(\bar{\sigma}_1+i\omega)^2}\,\frac{\{a\omega(\omega-2i\bar{\sigma}_1)\tau-2(\bar{\sigma}_1+i\omega)(2\bar{\sigma}_1+i(1+a^2)\omega)(t+x/\nu_0-\tau)}{(2\bar{\sigma}_1+i(1-a^2)\omega)\sqrt{a^2\tau^2-(t+x/\nu_0-\tau)^2}}$$

$$+ia(1+a^2)\frac{2a^4\bar{\sigma}_1\omega\tau+(2\bar{\sigma}_1^2+(1-a^2)\omega^2)(t+x/\nu_0-\tau)}{(2\bar{\sigma}_1-i(1-a^2)\omega)(a^4\omega^2\tau^2-\Omega^2(t+x/\nu_0-\tau)^2)}\,\bar{\sigma}_1\sqrt{a^2\tau^2-(t+x/\nu_0-\tau)^2}$$

$$-16a|1-a^2|\sqrt{a^2\tau^2-(t+x/\nu_0-\tau)^2}\bar{\sigma}_1\,\frac{a\tau-i(1+a^2)(t+x/\nu_0-\tau)}{(2\bar{\sigma}_1+i(1-a^2)\omega)((1+a^2)^2(t+x/\nu_0-\tau)^2-4a^4\tau^2)}\Bigg)$$

$$\times I_1(\bar{\sigma}_1\sqrt{\tau^2-(1/a^2)(t+x/\nu_0-\tau)^2})-\Bigg(8\frac{2\bar{\sigma}_1+i(1+a^2)\omega}{2\bar{\sigma}_1-i(1+a^2)\omega}-a(1+a^2)\bar{\sigma}_1$$

$$-\frac{2\omega\Omega(a^2\tau^2-(t+x/\nu_0-\tau)^2)-i\bar{\sigma}_1\tau(t+x/\nu_0-\tau)\Omega^{-1}(2\bar{\sigma}^2+(1+a^2)\omega^2)}{(2\bar{\sigma}_1-i(1+a^2)\omega)(a^4\omega^2\tau^2-\Omega^2(t+x/\nu_0-\tau)^2)}$$

$$+16\bar{\sigma}_1\frac{|1-a^4|(a^2\tau^2-(t+x/\nu_0-\tau)^2)-ia(1-a^2)\tau(t+x/\nu_0-\tau)}{(2\bar{\sigma}_1+i(1-a^2)\omega)((1+a^2)^2(t+x/\nu_0-\tau)^2-4a^4\tau^2)}\Bigg)I_0(\bar{\sigma}_1\sqrt{\tau^2-(1/a^2)(t+x/\nu_0-\tau)^2})\Bigg). \tag{A8}$$

[1] M. C. Tatham, G. Sherlock, and L. D. Westbrook, IEEE Photon. Technol. Lett. **PTL-5**, 1303 (1993).
[2] S. I. Averkov and V. P. Boldin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **23**, 1060 (1980).
[3] V. V. Borisov, *Transient Electromagnetic Waves* [in Russian], Leningrad State University Press, Leningrad (1987), 240 pp.
[4] A. G. Nerukh and N. A. Khizhnyak, *Contemporary Problems in Nonsteady-State Macroscopic Electrodynamics* [in Russian], NPO Test-Radio, Kharkov (1991), 280 pp.
[5] F. A. Harfoust and A. Taflove, IEEE Trans. Antennas Propag. **AP-39**, 898 (1991).
[6] K. A. Barsukov and N. Yu. Grigor'eva, Zh. Tekh. Fiz. **66**(7), 134 (1996) [Tech. Phys. **41**, 706 (1996)].
[7] V. V. Borisov, Geomagn. Aéron. **29**, 730 (1989).
[8] A. G. Nerukh and I. Yu. Shavorykina, Zh. Tekh. Fiz. **62**(5), 108 (1992) [Sov. Phys. Tech. Phys. **37**, 543 (1992)].
[9] S. L. Dvorak and D. G. Dudley, IEEE Trans. Electromagn. Compat. **EMC-37**(2), 192 (1995).
[10] M. A. Lavrent'ev and B. V. Shabat, *Methods in the Theory of the Functions of a Complex Variable* [in Russian], Nauka, Moscow (1987), 688 pp.
[11] S. L. Dvorak and E. F. Kuester, J. Comp. Physiol. **87**, 301 (1990).
[12] H.-Y. Pao, S. L. Dvorak, and D. G. Dudley, IEEE Trans. Antennas Propag. **AP-44**, 925 (1996).
[13] T. D. Visser, H. Block, and D. Lenstra, IEEE J. Quantum Electron. **QE-31**, 1803 (1995).

Translated by R. M. Durham

Electrostrictive mechanism of microwave losses in a planar strontium titanate film capacitor

O. G. Vendik and L. T. Ter-Martirosyan

St. Petersburg State Electrical Engineering University, 197376 St. Petersburg, Russia
(Submitted July 9, 1998)
Zh. Tekh. Fiz. **69**, 93–99 (August 1999)

An analysis is made of the dielectric losses in the microwave range in a planar ferroelectric capacitor caused by electrostrictively excited sound in the ferroelectric in the presence of a bias field (induced piezoeffect). An approximate expression is obtained for the dielectric loss tangent as a function of the bias field and the frequency. Numerical estimates are made for single-crystal strontium titanate at 78 K. This loss mechanism may be a determining factor for the fabrication of planar capacitors using high-quality ferroelectric films having properties close to those of the single crystal. © *1999 American Institute of Physics.* [S1063-7842(99)01708-0]

INTRODUCTION

Field-controlled microwave devices using ferroelectrics in the paraphase are opening up new possibilities for the development of devices exhibiting various advantages over existing analogs.[1–3] A decisive factor determining the possibility of using ferroelectrics in microwave technology is the acceptable level of dielectric losses in these materials in the microwave range. An important mechanism responsible for the microwave losses in ferroelectrics is the electrostrictive conversion of electromagnetic field energy into hypersonic vibrational energy in a sample whose dimensions are comparable with the hypersound wavelength. The contribution of electrostriction has been estimated qualitatively for bulk $(Ba,Sr)TiO_3$ samples.[4] The influence of acoustic vibrations excited by electrostriction in a thin $SrTiO_3$ film on the noise of a ferroelectric parametric amplifier was investigated in Refs. 1 and 5. The use of planar capacitors or coplanar lines[6,7] in microwave technology is currently attracting interest, and experimental[8] and theoretical[9] studies have been made of various mechanisms for microwave losses in materials such as $SrTiO_3$. In this context, it is interesting to study the microwave losses induced by electrostriction in a fairly thin dielectric film, when the film thickness is comparable with the wavelength of sound in the material at microwave frequencies. Numerical estimates will be made for $SrTiO_3$ (STO) at $T=78$ K. At this temperature strontium titanate in combination with a high-temperature superconductor is a promising material for microwave applications.[10]

1. FORMULATION OF THE PROBLEM

A typical design of a planar capacitor is shown in Fig. 1a, where h is the thickness of the STO film (usually $h \cong 0.5–1\ \mu$m), g is the gap width (usually $g \cong 2–10\ \mu$m), H is the substrate thickness (usually $H \cong 0.5–1.0$ mm), w is the width of the active zone of the capacitor (usually $w \cong 0.5–1.0$ mm), and L is the length of the capacitor electrodes (usually $L \cong 0.5–1.0$ mm). Figure 1a also shows the distribution of the electric displacement D. For $g \gg h$ we can assume approximately that the electric displacement in the capacitor active zone does not depend on the coordinates.

The electric displacement **D** consists of ac and dc components

$$\mathbf{D}=\mathbf{D}^{ac}+\mathbf{D}^{dc}. \tag{1}$$

The presence of a dc component $\mathbf{D}^{dc}$ leads to the formation of an induced piezoeffect and to the excitation of hypersound which is linearly related to the ac component $\mathbf{D}^{ac}$. The boundaries of the active region determine the emitter size which, being comparable with the sound wavelength, is equivalent to satisfying the spatial phase-matching conditions. In the active region of the planar capacitor nearly cylindrical hypersonic waves are excited, propagating from the STO film into the substrate. In order to simplify the calculations this wave excitation mechanism may be represented as two wave processes: a) plane waves excited in the active region which propagate along the STO film below the electrodes; b) plane waves excited in the active region which propagate along the normal to the film into the substrate. Having made this approximation, we reduce the problem to two independent one-dimensional plane problems.

1. A narrow, infinitely long, free ferroelectric rod (film) of thickness h positioned along the z axis (Fig. 1b). Between the cross sections $z=-l$ and $z=l$ $(2l=g)$ there is an active region in which the electric displacement components $\mathbf{D}^{dc}$ and $\mathbf{D}^{ac}$ do not depend on the coordinates. The strictive excitation of hypersound is attributed to the diagonal component Q of the electrostriction tensor. The presence of thin conducting electrodes in the regions $z=-l$ and $z=l$ does not influence the conditions for the excitation of hypersound. We postulate that the STO film is mechanically free both on the ''air'' side and on the substrate side. This simplification should not significantly influence the condition for the excitation of longitudinal modes propagating along the interfaces specified above.

2. A semi-infinite rod of width g positioned along the z axis (Fig. 1c). At the end of the rod there is an active region $0 \leqslant z \geqslant 2l$ $(2l=h)$, in which the electric displacement com-

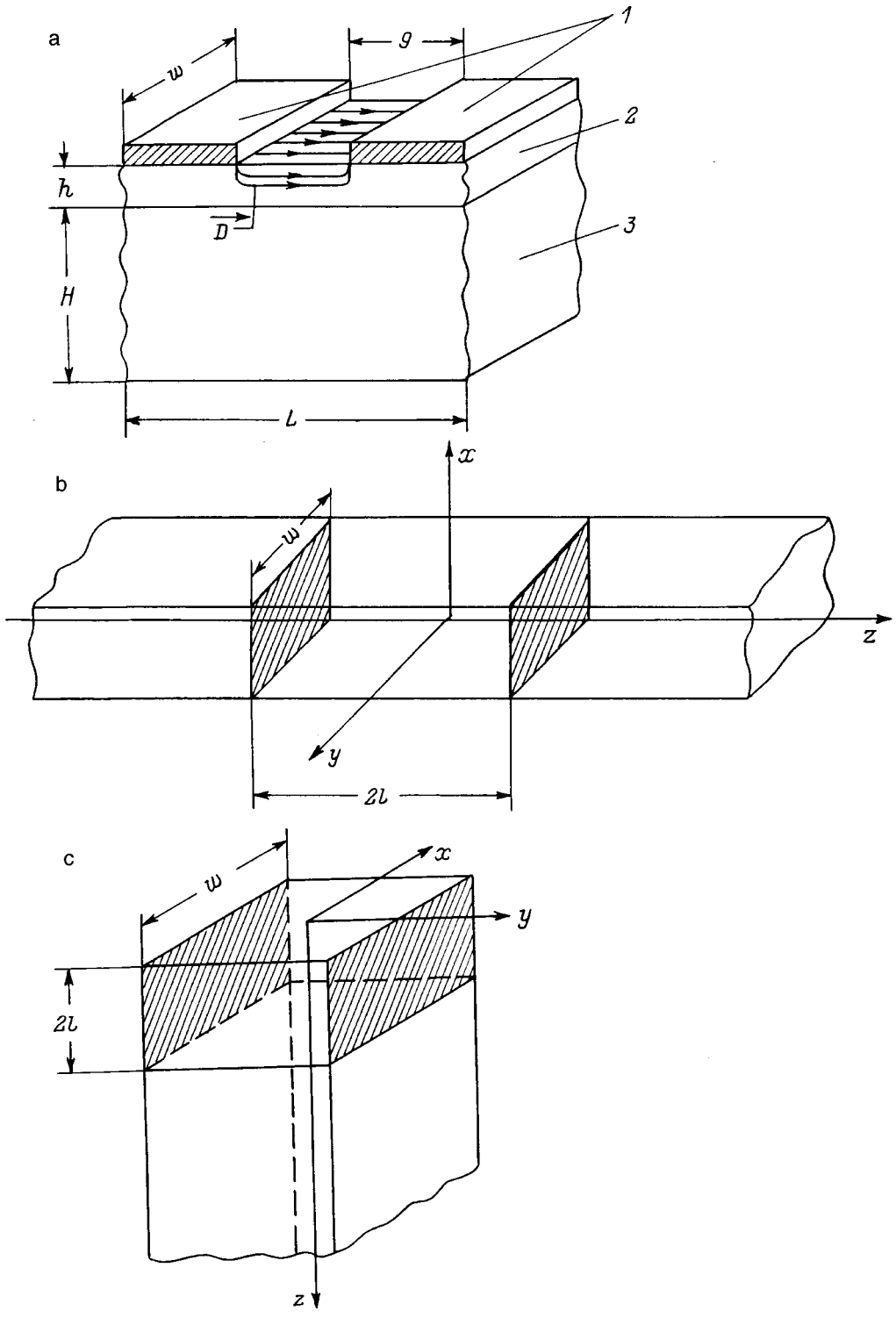

FIG. 1. Planar capacitor: a — schematic, using the following notation: *1* — electrodes, *2* — ferroelectric film, *3* — substrate; b,c — models used to calculate the excitation of longitudinal acoustic waves propagating parallel and perpendicular to the surface of the film, respectively.

ponents $\mathbf{D}^{dc}$ and $\mathbf{D}^{ac}$ are perpendicular to the z axis. The strictive excitation of hypersound is attributed to the off-diagonal component R_k of the electrostriction tensor. Assuming that the acoustic characteristics of the STO film and the substrate are similar, we eliminate reflections at the interface from our analysis; thus we are assuming that the semi-infinite rod is acoustically homogeneous.

2. FUNDAMENTAL RELATIONS BETWEEN THE DIELECTRIC AND MECHANICAL CHARACTERISTICS OF A STRONTIUM TITANATE FILM

The electric displacement vector D_m and the elastic stress tensor σ_{kl} may be used as independent variables. The usual relations between the dielectric and mechanical quantities, taking into account the electrostriction effect[11–13] give

$$u_{ij}=s_{ijkl}\sigma_{kl}+Q_{ijmn}D_mD_n,$$
$$E_m=\varepsilon_0^{-1}\varepsilon_{mn}^{-1}D_n-2Q_{ijmn}\sigma_{ij}D_n, \quad (2)$$

where u_{ij} are the components of the strain tensor, s_{ijkl} and Q_{ijmn} are the components of the elastic compliance and electrostriction tensors, respectively, E_m are the components of the electric field strength tensor, ε_{mn} are the components of the permittivity tensor, and ε_0 is the permittivity of free space.

If the electric field strength vector E_m and the elastic stress tensor σ_{ml} are taken as independent variables, we can write

$$u_{ij}=s_{ijkl}\sigma_{kl}+R_{ijmn}E_mE_n,$$
$$D_m=\varepsilon_0\varepsilon_{mn}E_n-2R_{ijmn}\sigma_{ij}E_n, \quad (3)$$

where R_{ijmn} are the components of a modification of the electrostriction tensor different to (2).

We know that strontium titanate exhibits appreciable anisotropy at low temperatures in a crystal converted to the single-domain state in terms of a structural phase transition.[14] Anisotropy in the film may also be caused by nonuniform mechanical stresses.[15] In order to simplify the problem we shall assume that the STO film is isotropic. In an isotropic medium the permittivity tensor is converted into a scalar, $\varepsilon_{mn}=\varepsilon_r$, where ε_r is the relative permittivity of the medium, and the electrostriction tensor has two independent components: Q_{ijmn} has the diagonal component Q and the off-diagonal component Q_k and the tensor R_{ijmn} has the components R and R_k, respectively. For an isotropic medium we have[12,13]

$$R=Q\cdot\varepsilon_0^2\varepsilon^2, \quad R_k=Q_k\cdot\varepsilon_0^2\varepsilon^2. \quad (4)$$

The elastic compliance tensor of an isotropic medium also has two independent components: the diagonal component s and the off-diagonal component s_k. We usually find $|s_k|<|s|$; for example, in strontium titanate at 78 K we have $s_k\cong-1.4\times10^{-12}\,\mathrm{m^2/N}$ and $s\cong4.6\cong10^{-12}\,\mathrm{m^2/N}$ (Ref. 16). For simplicity we shall neglect the contribution of the off-diagonal component of the elastic compliance tensor; in this approximation the systems of equations (2) and (3) only describe the strictive excitation of longitudinal waves.

Assuming that these simplifications have been made, we substitute Eq. (1) into Eqs. (2) and (3). Then assuming that the ac components of the electric field strength vectors and the electric displacement are much smaller than the corresponding static component, we neglect the squares of these quantities. We then obtain

$$u_z^{ac}=s\sigma_z^{ac}+2QD_z^{dc}D_z^{ac},$$
$$E_z^{ac}=\varepsilon_0^{-1}\varepsilon^{-1}D_z^{ac}-2Q\sigma_z^{ac}D_z^{dc}, \quad (5)$$
$$u_z^{ac}=s\sigma_z^{ac}+2R_kE_z^{dc}E_z^{ac},$$
$$D_z^{ac}=\varepsilon_0\varepsilon E_z^{ac}+2R_k\sigma_z^{ac}E_z^{dc}. \quad (6)$$

In both systems of equations, the time-varying z component of the strain tensor is related to the r_z component of the particle displacement vector of the medium during deformation or wave propagation

$$u_z=\frac{\partial r_z}{\partial z}. \quad (7)$$

In addition, in order to incorporate the acoustic waves in the calculations, we need to use the wave equation[17]

$$\frac{d^2r_z}{dz^2}+k^2r_z=0, \quad (8)$$

where k is the wave number of the longitudinal hypersonic wave.

3. INFINITE ROD WITH LONGITUDINAL ELECTRIC DISPLACEMENT IN THE ACTIVE ZONE

We shall consider the first problem (Fig. 1b). In this case, the regions of compression and rarefaction in the longitudinal hypersonic wave will be parallel to the electrodes (Fig. 1b) so that at any point between the electrodes, the electric displacement D_z^{ac} will be the same and the electric field strength E_z^{ac} will be a function of the coordinate z. In view of this, it is convenient to use the electric displacement vector D_m and the elastic stress tensor σ_{il} as independent variables for the calculations. Assuming that the above approximations have been made, we shall seek joint solutions of the system (5) and Eq. (8) allowing for Eq. (7). We shall use the following boundary conditions: at the interface between the active zone and the free infinite rod $z=l$ and $z=-l$ the particle displacements r_z and the mechanical stresses σ_{zz} are continuous, and only outgoing waves exist in all external regions relative to the active region.

By solving the system (5), (7), and (8) under the boundary conditions specified above (see Appendix A), we can determine the electric field strength in the active region

$$E^{ac}(z)=(1/\varepsilon_0\varepsilon_r)[1+iW\sin kl\cos kz]D^{ac}, \quad (9)$$

where we assume that $W\ll1$ and introduce the following notation:

$$W=V\cdot\Phi, \quad V=(4/s)Q^2\varepsilon_0^3, \quad \Phi=\varepsilon_r^3(E^{dc})(E^{dc})^2. \quad (10)$$

Integrating $E^{ac}(z)$ over the length of the active zone $(-l,l)$, we obtain the potential difference between the electrodes $I=i\omega D^{ac}S$, where S is the electrode area (the cross-sectional area of the rod) and we find the impedance between the electrodes

$$Z=-i[1/(\omega C)](1+iW(\sin^2 kl)/kl), \quad (11)$$

where C is the capacitance between the electrodes.

We shall assume that $\varepsilon_r=\varepsilon'-i\varepsilon''$, and $\tan\delta=\varepsilon''/\varepsilon'$. Then we have

$$\tan\delta=W(\sin^2 kl)/kl. \quad (12)$$

The factor $(\sin^2 kl)/kl$ is a rapidly oscillating function. In real samples, the spread of acoustic properties of the material within the geometric dimensions of the active region leads to averaging of these oscillations. We express the phase velocity of the longitudinal hypersonic waves in the form $v_l(x)=v_l(1+x)$, where x is a random quantity which obeys a Gaussian distribution with a zero average and an rms deviation x_0. We then have

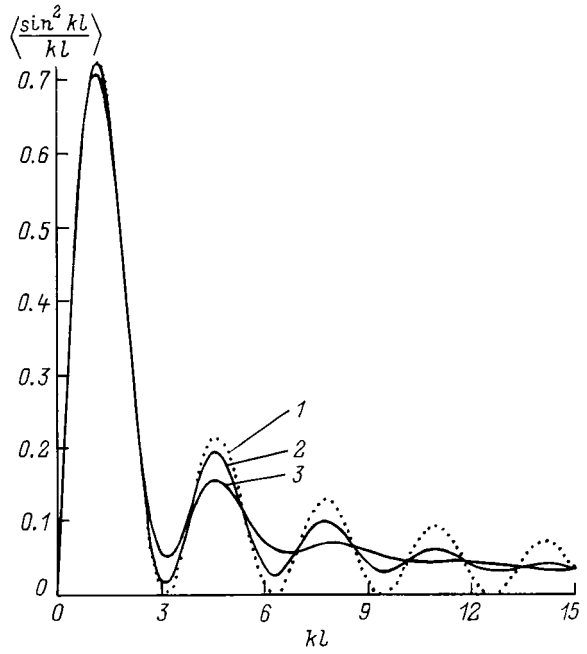

FIG. 2. Factor from Eq. (12) determining the dependence of tan δ on the microwave field frequency for various values of the rms deviation of the velocity of sound in the material: $x_0=0$ (*1*), 0.1 (*2*), 0.2 (*3*).

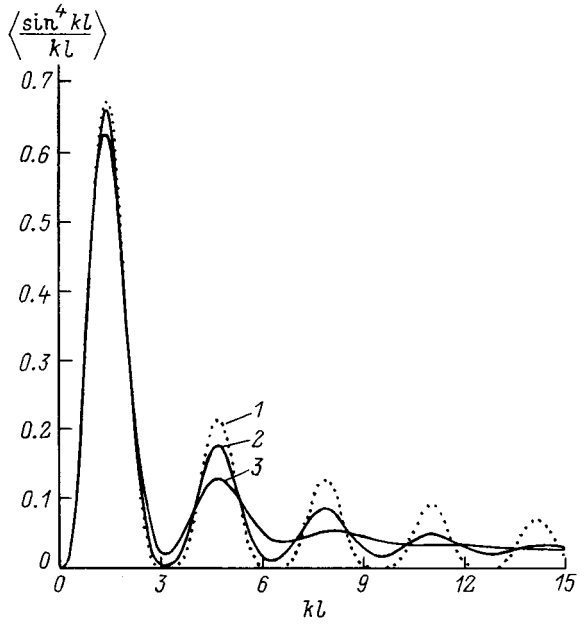

FIG. 3. Factor from Eq. (19) determining the dependence of tan δ on the microwave field frequency for various values of the rms deviation of the velocity of sound in the material: $x_0=0$ (*1*), 0.1 (*2*), 0.2 (*3*).

$$\langle(\sin^2 kl)/kl\rangle=[1/(\sqrt{\pi}x_0)]\int_{-\infty}^{\infty}\{\sin^2[kl/(1+x)]/[kl/(1+x)]\}\exp[-(x/x_0)^2]dx. \quad (13)$$

For $x_0 \ll 1$ the selected limits of integration barely affect the physical constraints on the permissible values of the hypersound velocity, but significantly simplify the integration procedure. Figure 2 gives the function $(\sin^2 kl)/kl$ and the results of calculations using formula (13) for $x_0=0.1$ and 0.2. For $kl \geqslant 5$ averaging yields the simple result

$$\langle(\sin^2 kl)/kl\rangle \cong 1/2kl. \quad (14)$$

4. SEMI-INFINITE ROD WITH A TRANSVERSE ELECTRIC FIELD IN THE ACTIVE ZONE

We shall now consider the second problem (Fig. 1c). In this case, the regions of compression and rarefaction in the longitudinal hypersonic wave will be perpendicular to the electrodes (Fig. 1c) so that at all points between the electrodes the electric field strength E_z^{ac} will be the same, and the electric displacement D_z^{ac} will be a function of the coordinate z. It is therefore convenient to use the electric field strength vector E_m and the elastic stress tensor σ_{kl} as independent variables for the calculations. Assuming that the above approximations have been made, we shall seek a joint solution of the system (6) and Eq. (8), with Eq. (7) taken into account. To do this we shall use the following boundary conditions: at $z=0$, $\sigma_z=0$ (free end); at $z=2l$ the particle displacements r_z and the mechanical stresses σ_z are continuous at the interface. Moreover, only outgoing waves exist in the external region relative to the active zone.

By solving the system (6)–(8) under the boundary conditions specified above (see Appendix B) we can determine the electric displacement in the active region

$$D^{ac}=\varepsilon_r\varepsilon_0[1+iW_k(\cos 2kl-1)\sin kz]E^{ac}, \quad (15)$$

where we assume that $W_k \ll 1$, and taking Eq. (4) into account, we introduce the following notation:

$$W_k=V_k\cdot\Phi, \quad V_k=(4/s)Q_k^2\varepsilon_0^3, \quad \Phi=\varepsilon_r^3(E^{dc})(E^{dc})^2. \quad (16)$$

The potential difference between the electrodes is $U=E^{ac}g$; the displacement current is

$$I=i\omega(S/2l)\int_0^{2l}D^{ac}(z)dz, \quad (17)$$

where $S=w\cdot 2l$ is the electrode area.

The impedance between the electrodes is then

$$Z=-i(1/\omega C)(1+i2W_k(\sin^4 kl)/kl). \quad (18)$$

Consequently, we have

$$\tan\delta=2W_k(\sin^4 kl)/kl. \quad (19)$$

The average of the oscillating function $(\sin^4 kl)/kl$, obtained by analogy with Eq. (13) is plotted in Fig. 3. For $kl \geqslant 5$, averaging yields the simple result

$$\langle(\sin^4 kl)/kl\rangle \cong (3/8)\cdot(1/kl). \quad (20)$$

5. DISCUSSION

It was noted in Sec. 1 that the dielectric losses in a planar capacitor caused by electrostriction in the presence of a dc bias field are the total losses obtained using the two calculation models, where the size of the active region in the first model is equal to the gap width $(2l=g)$ and that in the

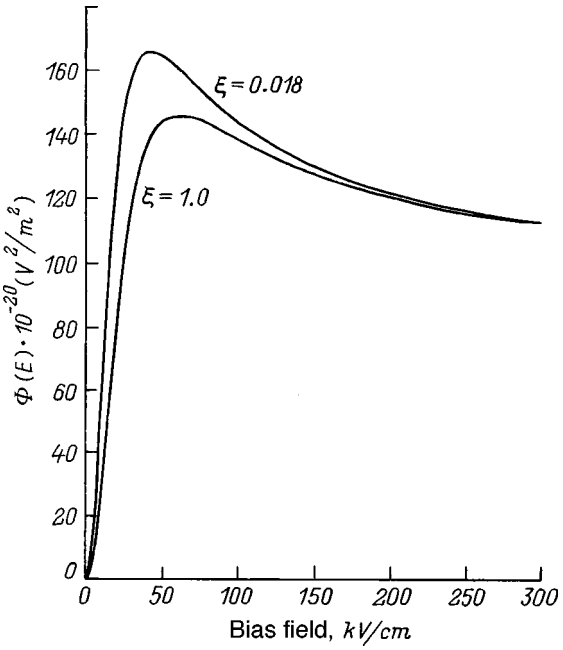

FIG. 4. The parameter Φ as a function of the bias field strength for strontium titanate at $T=78$ K.

second is equal to the thickness of the ferroelectric film ($2l=h$). Combining formulas (12) and (19), allowing for Eqs. (10) and (16), we obtain

$$\tan\delta_{\Sigma}=[V\langle[\sin^2(kg/2)]/(kg/2)\rangle + 2V_k\langle[\sin^4(kh/2)]/(kh/2)\rangle]\Phi. \quad (21)$$

Let us give quantitative estimates of the quantities appearing in this formula,

$$k=\omega/v_l, \quad (22)$$

where ω is the frequency of the microwave field in the planar capacitor and v_l is the longitudinal velocity of sound in the ferroelectric. For STO we have[16] $v_l\cong 7500$ m/s.

In order to estimate the value of the parameter $\Phi(E^{dc})$ defined in Eqs. (10) and (16), it is convenient to use a model description of the dielectric properties of the ferroelectric.[1,18] Figure 4 gives the results of calculating the parameter $\Phi(E^{dc})$ for strontium titanate at $T=78$ K. In accordance with the model used,[1,18] the parameter ξ_s is a measure of the degree of structural imperfection of the material: for a high-quality single crystal we have $\xi_s=0.018$, and for a deposited film $\xi_s\cong 1$. Note that the parameter $\Phi(E^{dc})$ of interest depends weakly on the density of material defects. Usually, a bias field of at least $E^{dc}=3$ MW/m must be applied to have a controlling effect on an STO film. From Eqs. (10) or (16) we find that $E^{dc}\geqslant 3$ MV/m and $\Phi\cong 1.4\times 10^{22}$ V^2/m^2.

Bearing in mind that in STO we have $s=s_{11}\cong 4.6\times 10^{-12}$ m^2/N, $Q\cong 0.066$ m^4/C^2, and $|Q_k|=0.01$ m^4/C^2 (Refs. 16 and 19), we find in accordance with Eqs. (10) and (16) that $V=2.62\times 10^{-24}$ m^2/V^2 and $V_k=6.01\times 10^{-26}$ m^2/V^2. Using formula (21) we can then take these quantitative estimates to calculate the dielectric loss tangent in a ferroelectric layer of a planar capacitor caused by the electrostrictive conversion of microwave electric field energy into hypersonic energy, which is ultimately dissipated in heating the material. Figure 5 gives the calculated values of $\tan\delta$ as a function of frequency for various sizes of planar capacitor.

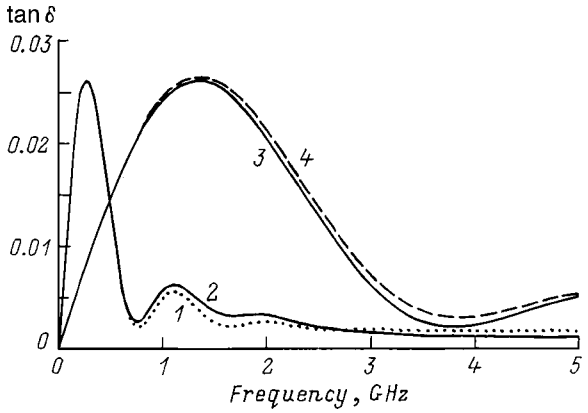

FIG. 5. Calculated values of tan δ for a single-crystal strontium titanate film as a function of the microwave field frequency at $T=78$ K and the following dimensions of the active zone of the planar capacitor (in micron): $g=10$ (*1*, *2*) and 2 (*3*, *4*); $h=2$ (*1*), 0.5 (*2*), 0.2 (*3*), and 1 (*4*).

In planar capacitors, at frequencies of 1–10 GHz the experimental values of tan δ_{Σ} lie in the range 0.02–0.05 (Refs. 1, 7, and 8) for $E^{dc}=0$ and generally decrease with increasing bias field. For a suitable choice of capacitor dimensions and microwave field frequency the calculated values of tan δ plotted in Fig. 5 may be comparable with the experimental values. The contribution of this loss mechanism may be a decisive factor if, as a result of refinements in the film technology, all sources of losses are eliminated apart from the fundamentally unavoidable scattering of the soft mode on thermal vibrations of the crystal lattice.[1,9]

CONCLUSIONS

An approximate solution has been obtained for the problem of the excitation of longitudinal hypersonic waves by a microwave electric field in a planar ferroelectric sample as a result of the piezoeffect induced by the dc electric field. Moreover the boundaries of the active region ensure that spatial phase-matching conditions are satisfied.

In the planar capacitor structures using strontium titanate films studied so far in the microwave range at $T=78$ K, the dielectric losses as a result of this effect are not a dominant factor although, when the film technology is refined and the overall microwave losses are therefore reduced, this loss mechanism may prove a decisive factor and will require further study.

The authors are grateful to S. P. Zubko for assistance with the calculations.

This work was carried out as part of the Government Program ''Physics of Condensed Media'' (Project No. 98055).

APPENDIX A

We shall take the subscript I to denote the active region and the subscripts II and III to denote the left and right parts of the rod, respectively. The complex amplitudes of the displacements r_z can then be given as:

$$r_{\mathrm{I}}=A_{\mathrm{I}}e^{-ikz}+B_{\mathrm{I}}e^{ikz},\quad r_{\mathrm{II}}=B_{\mathrm{II}}e^{ikz},$$
$$r_{\mathrm{III}}=A_{\mathrm{III}}e^{-ikz}, \tag{A1}$$

and for the components u_{zz} of the strain tensor we then obtain using Eq. (7)

$$u_{\mathrm{I}}=-ikA_{\mathrm{I}}e^{-ikz}+ikB_{\mathrm{I}}e^{ikz},\quad u_{\mathrm{II}}=ikB_{\mathrm{II}}e^{ikz},$$
$$u_{\mathrm{III}}=-ikA_{\mathrm{III}}e^{-ikz}. \tag{A2}$$

We then use formula (5) to obtain the components σ_{zz} of the mechanical stress tensor in the appropriate regions

$$\sigma_{\mathrm{I}}=-i(k/s)A_{\mathrm{I}}e^{-ikz}+i(k/s)B_{\mathrm{I}}e^{ikz}-2QD_z^{\mathrm{dc}}D_z^{\mathrm{ac}}/s,$$
$$\sigma_{\mathrm{II}}=i(k/s)B_{\mathrm{II}}e^{ikz},\quad \sigma_{\mathrm{III}}=-i(k/s)A_{\mathrm{III}}e^{-ikz}. \tag{A3}$$

The boundary conditions in this problem have the form: for $z=-l$

$$\sigma_{\mathrm{I}}(-l)=\sigma_{\mathrm{II}}(-l)\qquad\text{and}\qquad r_{\mathrm{I}}(-l)=r_{\mathrm{II}}(-l),$$

and for $z=l$

$$\sigma_{\mathrm{I}}(l)=\sigma_{\mathrm{II}}(l)\qquad\text{and}\qquad r_{\mathrm{I}}(l)=r_{\mathrm{II}}(l). \tag{A4}$$

Substituting Eqs. (A1) and (A3) into Eq. (A4), we obtain a system of linear equations for the complex amplitudes A_{I}, B_{I}, B_{II}, and A_{III}. Solving this system gives

$$A_{\mathrm{I}}=i(Q/k)D_z^{\mathrm{dc}}D_z^{\mathrm{ac}}e^{-ikl},$$
$$B_{\mathrm{I}}=-i(Q/k)D_z^{\mathrm{dc}}D_z^{\mathrm{ac}}e^{-ikl}. \tag{A5}$$

Substituting Eq. (A5) into Eq. (A3) gives

$$\sigma_{\mathrm{I}}=2(Q/s)D_z^{\mathrm{dc}}D_z^{\mathrm{ac}}(e^{-ikl}\cos kz-1). \tag{A6}$$

Substituting Eq. (A6) into Eq. (5) with allowance for Eq. (10) gives an expression for $E_z^{\mathrm{ac}}(z)$, which for $W\ll 1$ is easily converted into Eq. (9).

APPENDIX B

We shall denote the active region by the subscript I and the rest of the rod by the subscript II. The complex amplitudes of the displacements r_z can then be written as:

$$r_{\mathrm{I}}=A_{\mathrm{I}}e^{-ikz}+B_{\mathrm{I}}e^{ikz},\quad r_{\mathrm{II}}=A_{\mathrm{II}}e^{ikz}, \tag{B1}$$

and for the components u_{zz} of the strain tensor we obtain using Eq. (7)

$$u_{\mathrm{I}}=-ikA_{\mathrm{I}}e^{-ikz}+ikB_{\mathrm{I}}e^{ikz},\quad u_{\mathrm{II}}=-ikA_{\mathrm{II}}e^{-ikz}. \tag{B2}$$

We then use Eq. (6) to find the components σ_{zz} of the mechanical stress tensor in the selected regions

$$\sigma_{\mathrm{I}}=-i(k/s)A_{\mathrm{I}}e^{-ikz}+i(k/s)B_{\mathrm{I}}e^{ikz}-(2/s)R_kE_z^{\mathrm{dc}}E_z^{\mathrm{ac}},$$
$$\sigma_{\mathrm{II}}=-i(k/s)A_{\mathrm{II}}e^{-ikz}. \tag{B3}$$

The boundary conditions in this problem have the form: for $z=0$

$$\sigma_{\mathrm{I}}(0)=0,$$

and for $z=2l$

$$\sigma_{\mathrm{I}}(2l)=\sigma_{\mathrm{II}}(2l)\qquad\text{and}\qquad r_{\mathrm{I}}(2l)=r_{\mathrm{II}}(2l). \tag{B4}$$

Substituting Eqs. (B1) and (B3) into Eq. (B4), we then obtain a system of linear equations for the complex amplitudes A_{I}, B_{I}, and A_{II}. Solving the system gives

$$A_{\mathrm{I}}=i(1/k)R_kE_z^{\mathrm{dc}}E_z^{\mathrm{ac}}(2-e^{-i2kl}),$$
$$B_{\mathrm{I}}=-i(1/k)R_kE_z^{\mathrm{dc}}E_z^{\mathrm{ac}}e^{-i2kl}. \tag{B5}$$

Substituting Eq. (B5) into Eq. (B3) gives

$$\sigma_{\mathrm{I}}=(2/s)R_kE_z^{\mathrm{dc}}E_z^{\mathrm{ac}}(e^{-ikz}+ie^{-i2kl}\sin kz-1). \tag{B6}$$

Substituting Eq. (B6) into Eq. (6) gives an expression for $D_z^{\mathrm{ac}}(z)$, which for $W_k\ll 1$ is easily transformed to give Eq. (15).

[1] *Ferroelectrics in Microwave Technology*, edited by O. G. Vendik [in Russian], Sovet-skoe Radio, Moscow (1979), 272 pp.
[2] O. G. Vendik and L. T. Ter-Martirosyan, Zh. Tekh. Fiz. **63**(12), 108 (1993) [Tech. Phys. **38**, 1090 (1993)].
[3] O. G. Vendik, I. G. Mironenko, and L. T. Ter-Martirosyan, Microwaves RF **33**(7), 67 (1994).
[4] O. G. Vendik, V. V. Nezhentsev, and L. M. Platonova, Izv. Leningrad Élektrotekh. Inst. No. 64, 92 (1968).
[5] L. T. Ter-Martirosyan, Radiotekh. Elektron. **20**, 2592 (1975).
[6] S. S. Gevorgian, D. I. Kaparkov, and O. G. Vendik, IEEE Proc. Microwaves Antennas Propagat. **141**, 501 (1994).
[7] O. G. Vendik, P. K. Petrov, R. A. Chakalov *et al.*, in *Proceedings of the 27th European Microwave Conference*, Jerusalem, 1997, Vol. 1, pp. 196–202.
[8] J. Krupka, R. G. Geyer, M. Kuhn *et al.*, IEEE Trans. Microwave Theory Tech. **MTT-42**, 1886 (1994).
[9] O. G. Vendik, L. T. Ter-Martirosyan, and S. P. Zubko, J. Appl. Phys. **84**, 993 (1998).
[10] O. G. Vendik, L. T. Ter-Martirosyan, A. I. Dedyk *et al.*, Ferroelectrics **144**, 33 (1993).
[11] D. A. Berlincourt, D. R. Curran, and H. Jaffe, in *Physical Acoustics*, edited by W. Mason, Vol. 1, Part A *Methods and Devices* [Academic Press, New York, 1964; Mir, Moscow, 1966, pp. 204–326].
[12] J. F. Nye, *Physical Properties of Crystals: Their Representation by Tensors and Matrices* [Clarendon Press, Oxford, 1957; Mir, Moscow, 1967, 386 pp.].
[13] I. S. Zheludev, *Physics of Crystalline Dielectrics*, Vols. 1 and 2 [Plenum Press, New York, 1971; Nauka, Moscow, 1968, 464 pp.].
[14] T. Sakudo and H. Unoki, Phys. Rev. Lett. **26**, 851 (1971).
[15] T. Schimizu, Solid State Commun. **102**, 523 (1997).
[16] W. Rewald, Solid State Commun. **8**, 1483 (1970).
[17] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity*, 2nd. ed. [Pergamon Press, Oxford, 1970; Nauka, Moscow, 1965, 204 pp.].
[18] O. G. Vendik and S. P. Zubko, J. Appl. Phys. **82**, 4475 (1997).
[19] G. Schmidt and E. Hegenbarth, Phys. Status Solidi **3**, 329 (1963).

Translated by R. M. Durham

Mechanism for the appearance and randomization of the self-modulation of high-intensity spin waves

V. E. Demidov and N. G. Kovshikov

St. Petersburg State Electrical Engineering University, 197376 St. Petersburg, Russia
(Submitted October 19, 1998)
Zh. Tekh. Fiz. **69**, 100–103 (August 1999)

Modulational instability of high-intensity spin waves exhibiting amplification is studied in magnetic films under conditions of three-magnon decay. The mechanisms for this effect and for randomization of the envelope under self-modulation are identified. © *1999 American Institute of Physics.* [S1063-7842(99)01808-5]

Stochastic self-modulation of microwave spin oscillations and waves under conditions of three-magnon decay has been observed experimentally in various studies in bulk ferromagnetic samples (see, for example, Refs. 1 and 2) and also in thin ferromagnetic films (see Refs. 3 and 4). It has been established that a ferromagnetic spin system typically exhibits a wide variety of scenarios for the transition to chaos of the spin wave envelope and a wide range of stochastic oscillation regimes. In films the number of parametrically interacting spin waves is limited because of the discrete nature of the spectrum. Consequently, oscillatory motion characterized by a few degrees of freedom is established in the spin system of films. This constraint allows us to make a more effective study of the stochastic dynamics of spin waves.

Despite the major interest in this problem demonstrated by researchers, the mechanism giving rise to modulation of high-intensity spin waves in ferromagnetic films under conditions of three-magnon decay and its subsequent randomization has not yet been definitively established. The task for the present study was therefore to make an experimental and theoretical investigation of the modulational instability of high-intensity surface spin waves in ferromagnetic films and to explain the behavior observed.

Various experiments were carried out to obtain information on the self-modulation properties of high-intensity surface spin waves. Unlike known studies on the self-modulation of spin waves, the experimental apparatus comprised a spin wave delay line similar to that used in Ref. 5 with a microwave amplifier in the feedback circuit. This type of system can produce an exponential increase in the amplitude (gain) of the surface spin wave with time, limited by nonlinear effects in the ferromagnetic film. The experimental conditions corresponded to the decay instability of first-order surface spin waves. The active element was a VNA25 Mini-Circuits Integrated amplifier and the samples were 17 μm thick epitaxial films of yttrium iron garnet, having a saturation magnetization of 1750 G and a linear dissipation parameter of 0.5 Oe with free surface spins. The spin waves were excited and detected using microstrip transducers 30 μm wide and 5 mm apart. A prototype of the delay line was placed in a magnetic field tangential to the surface of the film, whose strength varied in the range $H_0 = 400$–600 Oe. The growth rate γ of the spin wave amplitude was varied by means of a variable attenuator connected in series with the amplifier. The microwave signal generated by the system was detected in order to isolate the amplitude modulation signal and the result was fed to the analog-to-digital converter of a computer for further processing.

The results of the investigations showed[4] that low-dimensional strange attractors form in the system, whose embedding dimensions increase with increasing γ. The transition from regular to chaotic self-modulation as the growth rate of the spin wave amplitude increases is usually accompanied by one or two period-doubling bifurcations, after which the envelope oscillations become stochastic and the modulation frequency increases continuously with increasing γ. Figures 1a–1d (left-hand side) show oscilloscope traces of the envelope as γ increases systematically from 0.1 (a) to 0.5 μs^{-1} (d); the values of γ are given for the linear regime and were recorded for a magnetizing field $H_0 = 447$ Oe. The spin-wave carrier frequency was 3040 MHz and its wave number 87 cm^{-1}. The right-hand diagrams show one-dimensional point mappings of the preceding maximum on the trace to the next one, corresponding to the oscilloscope traces. It can be seen that as γ increases, the envelope oscillations change from regular to stochastic. The transition to chaos shown in the figure is the most characteristic of a spin wave envelope in ferromagnetic films under three-magnon interaction conditions. One characteristic feature of the oscillations of the spin wave power is clearly visible, i.e., that these are of a relaxation nature. Sections of slow and fast motion can be clearly identified on the traces.

In order to explain these results we made some theoretical calculations. By jointly integrating the nonlinear equation of motion for the magnetization and the magnetostatic equations using an expansion of the variable magnetization in terms of spin wave modes and the tensor Green's functions of the magnetostatic equations,[6–8] we obtained a system of nonlinear differential equations for the slowly time-varying amplitudes of three parametrically coupled spin waves:

$$\dot{a}_1(t) = c_1 a_2^*(t) a_0(t) e^{i\delta t} - \beta a_1,$$

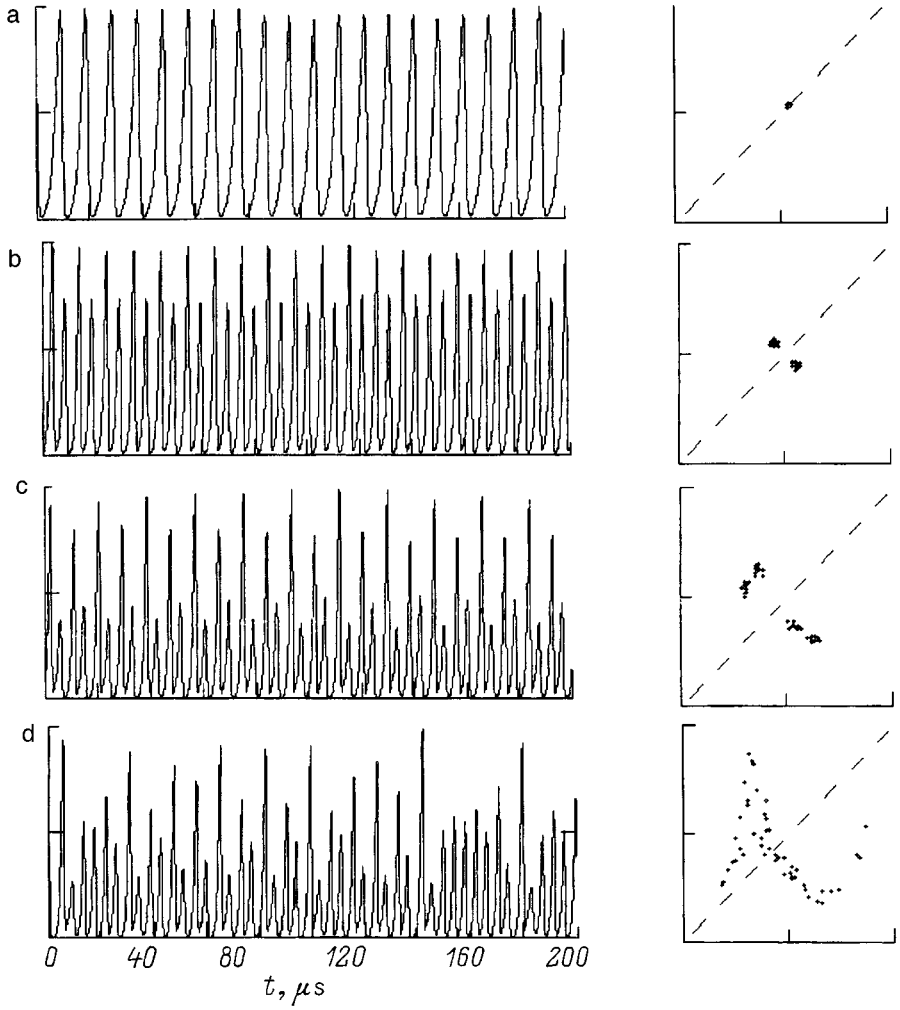

FIG. 1. Oscilloscope traces of spin wave power and point maps for $\gamma=0.1-0.5\ \mu s^{-1}$.

$$\dot{a}_2(t)=c_2 a_1^*(t)a_0(t)e^{i\delta t}-\beta a_2,$$

$$\dot{a}_0(t)=-c_3 a_1(t)a_2(t)e^{-i\delta t}+\gamma a_0, \qquad (1)$$

where a_0, a_1, and a_2 are the complex amplitudes of the initial and parametrically excited spin waves, β is the relaxation parameter of the parametric spin waves, and δ is the detuning from resonance.

The coefficients of the system c_1, c_2, and c_3 are given by

$$c_1=\omega_M(\lambda_{n_1k}\lambda_{00}+\mu_{n_1k}\mu_{00})(\lambda_{n_2k}+\mu_{n_2k})I^0_{n_10n_2}(|k|),$$

$$c_2=\omega_M(\lambda_{n_2k}\lambda_{00}+\mu_{n_2k}\mu_{00})(\lambda_{n_1k}+\mu_{n_1k})I^0_{n_20n_1}(-|k|),$$

$$c_3=-\omega_M(\lambda_{n_2k}\lambda_{00}+\mu_{n_2k}\mu_{00})(\lambda_{n_1k}+\mu_{n_1k})+(\lambda_{n_1k}\lambda_{00}+\mu_{n_1k}\mu_{00})(\lambda_{n_2k}+\mu_{n_2k})I^0_{n_20n_1}(|k|). \qquad (2)$$

Here we have $\lambda_{nk}=\sqrt{(A_{nk}+\omega_n)/2\omega_n}$ and $\mu_{nk}=\sqrt{(A_{nk}-\omega_n)/2\omega_n}$. The matrix element I^0 describes the nonlinear parametric interaction of the spin waves and is given by

$$I^0_{nn'n''}(k_\xi)=\frac{2}{L}\int_{-L/2}^{L/2}\int_{-L/2}^{L/2}G_Q(\xi,\xi',k_\xi)\times\Phi_n(\xi')\Phi_{n'}(\xi)\Phi_n(\xi)d\xi\,d\xi'.$$

All the other notation is as in Ref. 6.

When deriving the system (1), we assumed that the spin wave amplitude of the pump does not vary in space but only in time. This assumption was justified by the fact that the characteristic time of variation of the amplitude is an order of magnitude greater than the wave round-trip time in the delay line–amplifier ring system. The second assumption was that the pump wave has the wave number $k=0$ and parametrically excites spin waves whose wave vectors have the same modulus but opposite directions.

Approximate calculations using the formulas (2) show that for the experimental conditions described above, the coefficients of the system (1) have the values $c_1=745\ \mu s^{-1}$, $c_2=752\ \mu s^{-1}$, and $c_3=1497\ \mu s^{-1}$.

A system of the type (1) was first investigated in Ref. 9, where it was shown that this system may have stochastic solutions. Following Ref. 9, we write in terms of real variables

$$X=A_0\cos(\varphi+\delta t),\quad Y=A_0\sin(\varphi+\delta t),\quad Z=A_1^2=A_2^2.$$

Here A_i and φ_i are the real amplitude and phase of the ith wave, and $\varphi=\varphi_0-\varphi_1+\varphi_2$. The system (1) then has the form

$$\dot{X}=-\sigma Z+\gamma X+2\sigma Y^2-\delta Y,$$

$$\dot{Y}=\gamma Y-2\sigma XY+\delta X, \quad \dot{Z}=2Z(\sigma X-\beta), \tag{3}$$

where $\sigma=\sqrt{c_1c_2c_3}$.

The system (3) was integrated numerically using a fourth-order Runge–Kutta method. The resulting time dependences of the initial P_0 and parametric $P_{1,2}$ wave intensities for $\beta=9\ \mu s^{-1}$, $\delta=8\ \mu s^{-1}$, and $\gamma=0.2\ \mu s^{-1}$ are plotted in Fig. 2. For these values of the coefficients the envelope oscillations are regular. It can be seen that energy exchange takes place between the initial spin wave and the parametric waves. In the sections of slow motion, the amplitude of the initial wave increases exponentially with time as a result of this amplification. The amplitudes of the parametric waves remain small. In accordance with Eq. (1), in this situation the nonlinear wave interaction is very weak. Thus, removal of the fundamental wave energy by the parametric waves has little influence on the growth process. When the amplitude of the initial wave reaches a certain level, parametric amplification begins to play a significant role. Energy begins to transfer rapidly from the fundamental wave to the parametric ones, i.e., decay takes place. In this case, nonlinear attenuation of the initial wave predominates over external amplification and, as a result, the amplitude decreases rapidly. When the amplitude of the initial wave falls below a threshold level, the decay is replaced by a reverse, coalescence process, but because of dissipation and detuning from resonance, only a small fraction of the energy is returned to the fundamental wave. This process is repeated cyclically, resulting in self-modulation of the fundamental wave.

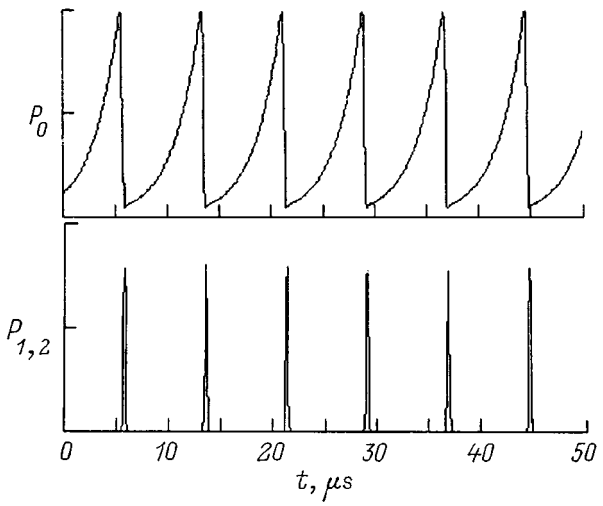

FIG. 2. Theoretical time dependence of the initial and parametric wave power for $\beta=9\ \mu s^{-1}$, $\delta=8\ \mu s^{-1}$, and $\gamma=0.2\ \mu s^{-1}$.

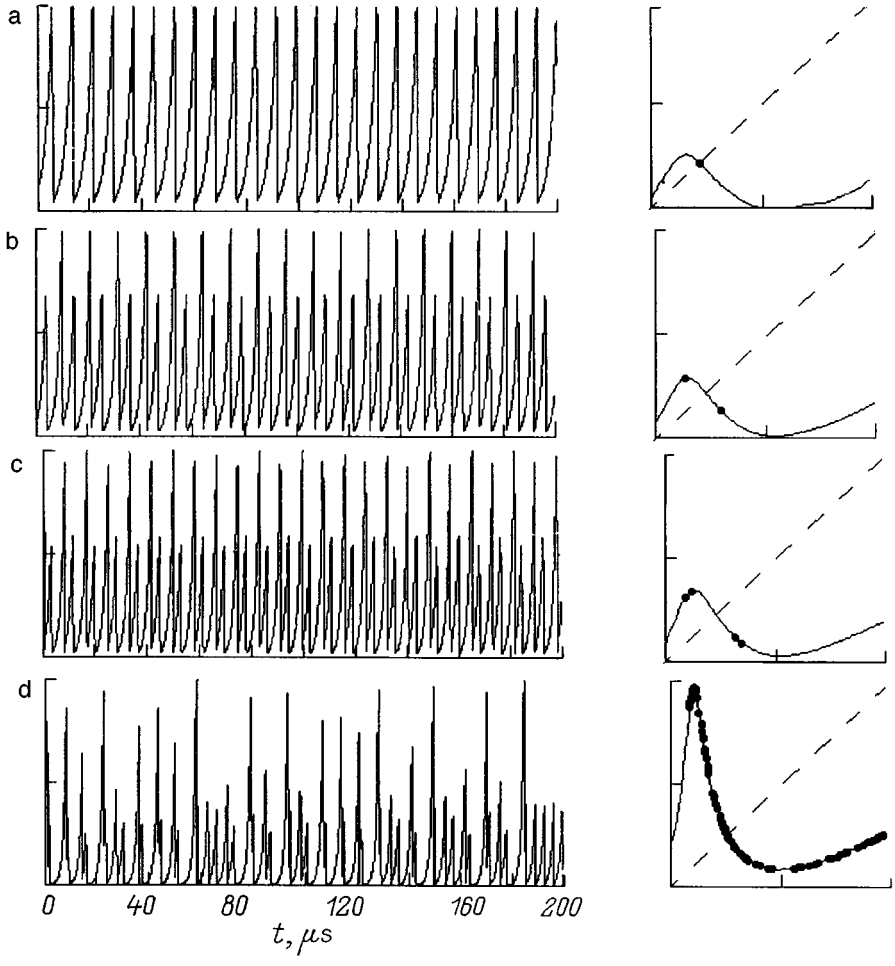

FIG. 3. Theoretical time dependences of the initial spin wave power and point maps for $\beta=9\ \mu s^{-1}$, $\delta=8$ (a), 7 (b), 6 (c), and 2 $\mu s-1$ (d); $\gamma=0.2$ (a), 0.3 (b), 0.4 (c), and 0.5 μs^{-1} (d).

Depending on the values of the parameters β and δ, the solutions of the system (3) may be singly periodic, multiply periodic, or stochastic. As the parameter γ varies, the self-modulation frequency varies but the oscillation regime remains almost unchanged. Figure 3 shows that for specific values of the parameters, the system of equations obtained has solutions qualitatively the same as the envelope oscillations observed experimentally. The point maps constructed using the calculated time series are shown on the right-hand side of Fig. 3. The circles show the fixed mappings corresponding to steady-state oscillation regimes. It can be seen that these calculated maps are similar to those observed experimentally. These observations indicate that the constructed theoretical description of stochastic self-modulation is quite adequate.

As a result of studying the system of equations (3), we established that a transition from regular to chaotic envelope oscillations similar to that observed experimentally can only take place if the parameters γ and β vary simultaneously. In this case, an increase in the gain γ increases the self-modulation frequency while a simultaneous decrease in the detuning from resonance δ makes the oscillation regime more complex. We can attempt to explain this relationship between the parameters γ and δ by the presence of a nonlinear shift of the natural frequencies of the spin waves as a result of an increase in the oscillation amplitude with increasing gain.

Thus, we can draw various conclusions. Self-modulation of nonlinear spin waves exhibiting amplification is a relaxation process. As the gain increases, the envelope oscillations become more complex and may become stochastic as a result of a simultaneous decrease in the detuning from resonance. In this case, the bifurcation values of γ, the types of regimes established, and the scenarios for the transition to chaos are determined by the relaxation frequency of the parametrically excited spin waves. This phenomenon can be explained in terms of the model of three interacting waves. In Ref. 4 the present authors showed that strongly stochastic envelope oscillations are characterized by around six degrees of freedom. By making additional allowance for the multiwave nature of the decay process, it would be possible to describe the stochastic dynamics of an envelope for large γ.

A mechanism similar to that described can also be proposed to explain stochastic self-modulation under transverse pumping in yttrium iron garnet spheres[1,2] under conditions of parametric instability of first-order spin waves. In this case, in Ref. 1 the term responsible for the gain must be replaced by a term describing the transfer of energy from an external source. The sections of increasing amplitude of the initial high-intensity wave can then be nonexponential but most of the other established dependences are broadly conserved.

These results give some idea of the reason for the appearance and randomization of the self-modulation of spin waves and may be used as a basis for further research.

This work was financed by the Russian Fund for Fundamental Research (Grant No. 96-02-19515) and by the Ministry for General and Professional Education of the Russian Federation (Grant No. 97-8.3-13).

[1] H. Benner, R. Henn, F. Roedelsperger *et al.*, Izv. Vyssh. Uchebn. Zaved. PND **3**, No. 1, pp. 32–50 (1995).
[2] A. I. Smirnov, Zh. Éksp. Teor. Fiz. **90**, 385 (1986) [Sov. Phys. JETP **63**, 222 (1986)].
[3] A. M. Mednikov, Fiz. Tverd. Tela (Leningrad) **23**(1), 242 (1981) [Sov. Phys. Solid State **23**, 136 (1981)].
[4] V. E. Demidov and N. G. Kovshikov, JETP Lett. **96**, 261 (1997).
[5] B. A. Kalinikos, N. G. Kovshikov, and A. N. Slavin, Zh. Éksp. Teor. Fiz. **94**(2), 159 (1988) [Sov. Phys. JETP **67**, 303 (1988)].
[6] B. A. Kalinikos, Izv. Vyssh. Uchebn. Zaved. Fiz. **24**(8), 42 (1981).
[7] O. G. Vendik, B. A. Kalinikos, and D. N. Chartorizhskiĭ, Fiz. Tverd. Tela (Leningrad) **19**, 387 (1977) [Sov. Phys. Solid State **19**, 222 (1977)].
[8] B. A. Kalinikos, Zh. Tekh. Fiz. **54**, 1846 (1984) [Sov. Phys. Tech. Phys. **29**, 1083 (1984)].
[9] S. Ya. Vyshkind and M. I. Rabinovich, Zh. Éksp. Teor. Fiz. **71**, 557 (1976) [Sov. Phys. JETP **44**, 292 (1976)].

Translated by R. M. Durham

Using an atomic force microscope in the surface modification regime to determine the migration energy of surface defects

E. V. Blagov and V. M. Mostepanchenko

Modus Research and Innovation Enterprise, 125047 Moscow, Russia

G. L. Klimchitskaya

Northwest Correspondence Polytechnical Institute, 191186 St. Petersburg, Russia

(Submitted August 12, 1997; resubmitted October 20, 1998)
Zh. Tekh. Fiz. **69**, 104–110 (August 1999)

The lines of constant force and the profiles of the horizontal force component are calculated for the scanning of the tip of an atomic force microscope over a surface vacancy in a close-packed lattice with allowance for atomic displacements. The character of the lines of force is studied in all three scanning regimes that arise for different values of the force: without modification of the surface by the tip, migration of a single vacancy by a single interatomic distance in the direction opposite to the motion of the tip, and ''dragging'' of a vacancy by the tip. It is shown that the profiles of the horizontal force component can be used to calculate the activation energy for surface migration of a vacancy. An estimate is made of the scanning force for which these effects may be observed experimentally. [S1063-7842(99)01908-X]

INTRODUCTION

Recently, atomic force microscopy has found numerous applications, not only in studies of surface microstructure (see, for example, Refs. 1 and 2) but also for studying atomic friction.[3–5] We know that in the contact mode, when the dominant forces between the tip of the atomic force microscope (AFM) and the surface of the sample are repulsive forces of an exchange nature, the highest horizontal resolution, a few angstroms, can be achieved by using a sharp tip.[6] Experimental and theoretical investigations (see, for example, Refs. 7–13) have shown that the nature of the scanning and the resulting constant-force surfaces depend on the atomic structure of the apex of the tip. In particular, in the simplest case of a ''single-atom'' tip (when only one atom at the very apex interacts with the surface atoms) at fairly high scanning forces in the contact mode the constant-force surfaces have breaks[10–13] into which the tip inevitably sinks, becoming set into the surface.[4,14] In order to achieve continuous scanning, a fairly large atomic cluster should be located at the apex of the tip[9–11] (''cluster'' tip).

Results reported in Refs. 11, 13, 15, and 16 indicate that an AFM can be used for the diagnostics of point defects localized at the surface of a solid. This conclusion was recently confirmed experimentally[17] in a comparative study of vacancies on a pre-irradiated surface, using scanning tunneling and atomic force microscopy. The next step could be an attempt to use AFM to determine the energy characteristics of the defect migration process, particularly vacancies.[18] Information currently available on these characteristics is quite inadequate. Although some results were obtained by mathematical modeling (see Refs. 19–21 and the literature cited therein), these were extremely sensitive to the choice of potential (see the discussion of this topic in Ref. 20). In addition, the migration process is strongly influenced by thermal fluctuations,[21] which also complicates the calculations. Thus, the possibility of directly determining the activation energy for defect migration by AFM is of some considerable interest.

In order to study the energy characteristics of a defect, we need to use an AFM in the surface modification regime, where the AFM tip can change the positions of the atoms, for example, by shifting an atom into an adjacent vacancy (the possibility of these processes for an interstitial atom at the surface of silicon was discussed in Ref. 4). In order to interpret these experimental results, we need to make calculations which take into account the motion of the atoms under the influence of the tip. For this purpose it is convenient to use a simple model developed in Ref. 7 which, however, can fairly accurately describe scanning in the contact mode when the forces of attraction between the tip and the surface can be neglected (this model was also used in Ref. 22 to determine the influence of atomic mobility on the breaks in the constant-force surface). In this model, which is developed here to describe scanning processes in the surface modification regime, the potential energy U of the sample plus AFM tip system is expressed as the sum of the interaction energy U_{ts} between the tip atoms and the sample atoms and the interaction energies U_s and U_t of the sample and tip atoms among themselves. The first of these terms describing the exchange repulsion of the tip and the surface can be represented approximately as the sum of the potentials α/r^{12}, where α is the phenomenological constant of the exchange interaction and r is the distance between the interacting atoms. The terms U_s and U_t are considered to be the lattice energies in the harmonic approximation (apart from the contribution of atoms adjacent to a vacancy, Sec. I). As a result,

the constant-force surfaces $z=z(x,y,F_0)$ and also the positions of the various sample and tip atoms $\mathbf{r}_i$ may be determined as solutions of the system

$$-\frac{\partial U}{\partial z}=F_0, \quad \mathrm{grad}_{\mathbf{r}_i}U=0. \tag{1}$$

In the present study we calculate the constant-force lines and also the horizontal force profiles when a cluster tip scans in the contact mode above the surface of a close-packed lattice containing a vacancy. We show how the force lines look in various scanning regimes which depend on the force: without any modification of the surface, with a change in the position of the defect by a single lattice parameter in the direction opposite to the direction of scanning, and with complex modification of the surface when the AFM tip ''drags'' the vacancy during the scanning process.

1. CONSTANT-FORCE LINES DURING SCANNING ABOVE A VACANCY IN A CLOSE-PACKED LATTICE

In order to eliminate the dependence of the results on the phenomenological parameter of the exchange interaction α, whose value is not known accurately, we shall use dimensionless values of the coordinates, forces, and energies:

$$\tilde{\mathbf{r}}=\frac{\mathbf{r}}{a}, \quad \tilde{F}=\frac{F}{F_0}, \quad \tilde{U}=\frac{U}{F_0 a}, \tag{2}$$

where a is the equilibrium distance between the atoms in the unperturbed close-packed lattice and $F_0=12\alpha/d^{13}$ is the force acting on a single-atom tip situated at the initial height d above the surface atom.

Taking into account the reasoning put forward in the Introduction, the dimensionless interaction energy between the tip and the sample is written as

$$\tilde{U}_{ts}=\frac{1}{12}\left(\frac{d}{a}\right)^{13}\sum_{i,j}\frac{1}{\tilde{r}_{ij}^{12}}, \tag{3}$$

where $\tilde{r}_{ij}$ is the dimensionless distance between the tip and sample atoms.

Following Refs. 10, 11, and 13, we can use a tip model which can describe the continuous scanning above a vacancy for any initial heights d (the tip is simulated by a paraboloid[23] with radius of curvature $R=5a/3$ and having a cluster of seven atoms at its apex, with one atom positioned at the distance $h\approx 0.22a$ beneath the center of a regular hexagon with sides of length $b=0.85a$ formed by the other six atoms; for further details see Refs. 10, 11, and 13). The subscript j in Eq. (3) then has values between 1 and 7. The results obtained in Ref. 22 indicate that allowing for the mobility of the cluster atoms at the apex of the tip has only a negligible influence on the profile of the constant-force surfaces and the profiles of the horizontal force component. For this reason we shall subsequently only take into account the mobility of the sample atoms. Hence, the tip energy U_t is assumed to be constant during scanning and may be neglected.

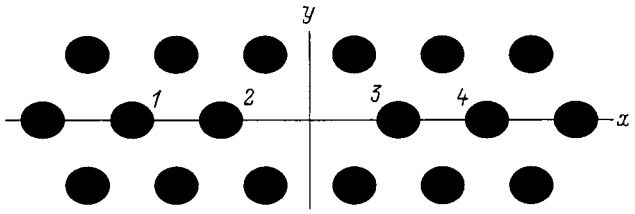

FIG. 1. Configuration of atoms near a vacancy.

If the lattice energy U_s is considered in the harmonic approximation, the general expression for the elastic displacement energy[24] allowing for the symmetry of a close-packed lattice may be expressed in the form

$$\tilde{U}_s=\frac{1}{2}\sum_{i,k}\tilde{D}^{(i,k)}(\Delta\tilde{\mathbf{r}}_i,\Delta\tilde{\mathbf{r}}_k), \tag{4}$$

where $\Delta\tilde{\mathbf{r}}_i=\tilde{\mathbf{r}}-\tilde{\mathbf{r}}_i^{(0)}$ describe the displacements of the equilibrium positions $\tilde{\mathbf{r}}_i$ of the lattice atoms during scanning from their equilibrium positions $\tilde{\mathbf{r}}_i^{(0)}$ in a lattice which does not interact with the tip (as we know, the scanning process can be considered to be quasistatic).

The dimensionless constants $\tilde{D}^{(i,k)}$ are related to the matrix of the force constants of a particular solid by

$$\tilde{D}^{(i,k)}=\frac{ad^{13}}{12\alpha}D^{(i,k)}. \tag{5}$$

It was shown in Ref. 22 that calculations of the continuous force surfaces can be confined to the simplified energy representation (4), assuming that during scanning only those atoms in the surface layer of the sample exposed to the direct action of the tip undergo displacements and neglecting the ensuing small displacements of neighboring atoms. This assumption implies that the force constant matrix $\tilde{D}^{(i,k)}=\tilde{D}\delta_{ik}$ is diagonal.

All the above reasoning has referred to the elastic displacements of atoms in a perfect lattice. We shall now consider the scanning of an AFM tip above a surface vacancy. In this case, the force acting from the tip may be so great that it causes elastic displacements of atoms adjacent to the vacancy, directed toward the vacancy. Let us assume that a vacancy is positioned at the point $\tilde{x}_v^{(0)}=0$ and four lattice atoms having coordinates $\tilde{x}_i^{(0)}=-2,-1,1,2$ for $i=1,2,3,4$, respectively, lie on a single straight line (Fig. 1). Here we neglect atomic relaxation in the nearest neighborhood of a vacancy for a sample which does not interact with the tip. This is justified by the fact that the calculations performed below are only a modeling of those data which should be obtained experimentally using an AFM. These data automatically allow for the influence of relaxation, fixing the actual position of all the atoms.

When describing the inelastic displacements of atoms adjacent to a vacancy, we need to bear in mind that displacements along the x axis play a dominant role for the vacancy migration process (Fig. 1). Thus to a first approximation we need only change the dependence on Δx in the energy of these atoms and replace the corresponding terms (4) by[25]

$$\tilde{U}_n(\Delta\tilde{x}_i)=\frac{\tilde{D}_n}{2\pi^2}\sin^2(\pi\Delta\tilde{x}_i), \tag{6}$$

where the dimensionless constant $\tilde{D}_n$ is related to the dimensional constant D_n by (5).

In fact, for displacements of an atom away from an adjacent vacancy (according to Fig. 1 this occurs when $\Delta\tilde{x}_2<0$ and $\Delta\tilde{x}_3>0$) we can assume $\tilde{D}_n\approx\tilde{D}$, neglecting any small change in this quantity near the vacancy. When the displacement is toward the vacancy, we generally find $\tilde{D}_n<\tilde{D}$. For small displacements the energy (6) has the same form as the corresponding contribution in Eq. (4). At the points $\Delta=\pm0.5a$, i.e., midway between the vacancy and the neighboring atoms, this energy has a maximum corresponding to a potential barrier of height (in conventional units)

$$\Delta E=\frac{D_n a^2}{2\pi^2}. \tag{7}$$

If, under the action of the tip, atom *2* in Fig. 1 is displaced to the right by $\Delta x_2=0.5$, it takes the place of the vacancy, while the vacancy is displaced to the left. Equation (6) will then describe the potential energy of atom *2* in the new position $\tilde{x}_2^{(0)}=0$.

We shall now calculate the constant-force lines when the AFM tip scans from the initial position above atom *1* with the coordinate $\tilde{x}_1^{(0)}=-2$ to the final position above atom *4* with the coordinate $\tilde{x}_4^{(0)}=2$. In all cases the initial tip height is $d=0.5a$. The parameters $\tilde{D}$ and $\tilde{D}_n$ can be conveniently expressed in the form

$$\tilde{D}=\frac{1}{\Delta\tilde{l}},\quad \tilde{D}_n=\frac{1}{\Delta\tilde{l}_n}, \tag{8}$$

where $\Delta\tilde{l}$ has the meaning of the elastic displacement of the lattice atom under the action of a force equal to the scanning force and $\Delta\tilde{l}_n$ is introduced by analogy.

The solution of system (1) can be used to describe three scanning regimes for various values of $\Delta\tilde{l}_n$ (note that the result is almost independent of the value of $\Delta\tilde{l}$, so that we shall only use a single value here, $\Delta\tilde{l}=0.05$). The first regime, corresponding to scanning without any modification of the surface, is established for $\Delta\tilde{l}_n<0.6$. In this case, even if atom *2* undergoes inelastic displacements, they remain fairly small. When the tip approaches the point $x=-1$, the atom describes half a closed trajectory with the maximum displacement $\Delta\tilde{x}_2<0.5$. When the tip is situated above the point $\tilde{x}=-1$, the atom is again located at $\tilde{x}=-1$, but with $\Delta\tilde{z}_2=-0.05$, i.e., below its initial position. For illustration Fig. 2a gives the constant-force curve obtained by solving the system (1) with the potential energy equal to the sum (3), (4), and (6) which is almost independent of the specific value of $\Delta\tilde{l}_n<0.6$. Thus, if the atom does not pass through the potential barrier and no modification of the surface occurs, the atomic displacements have no significant influence on the constant-force lines even if they are so large that they cannot be considered to be elastic.

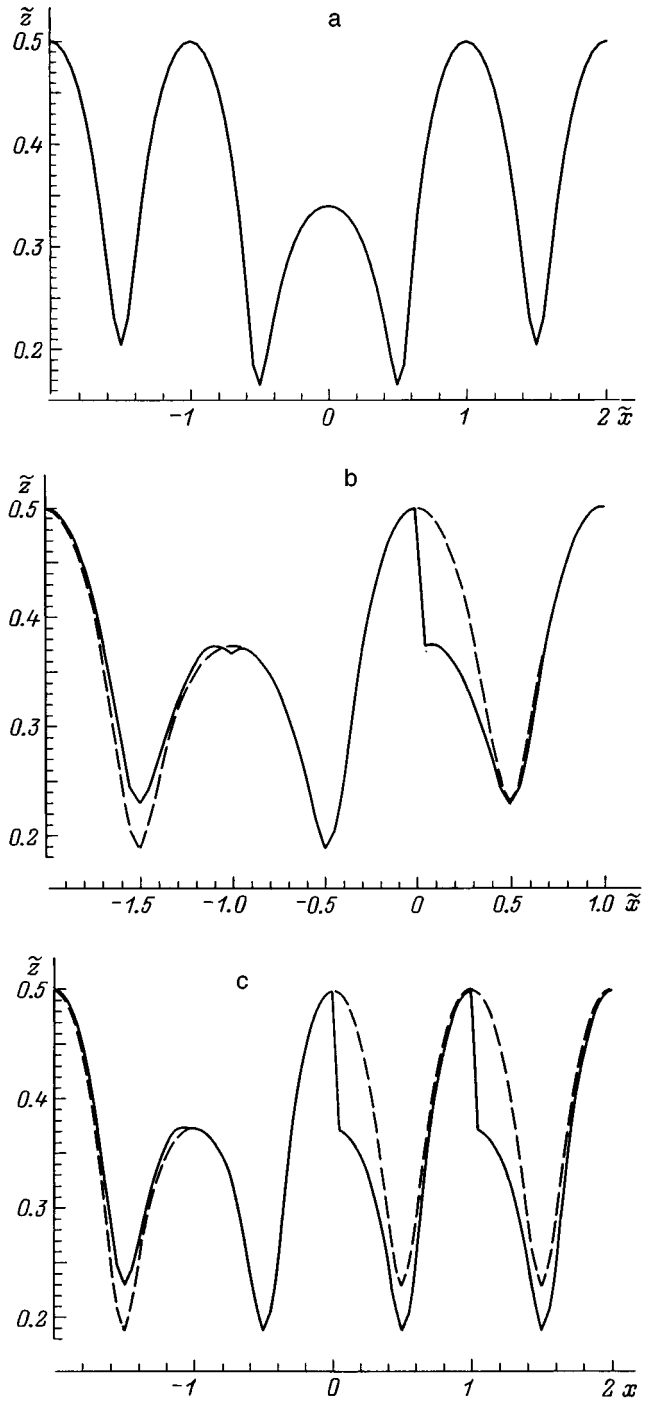

FIG. 2. Constant-force lines when an AFM tip scans above a vacancy.

The second possible scanning regime, with modification of the surface, occurs when $\Delta\tilde{l}_n\approx0.6$. In this case, atom *2* is displaced by the tip over the distance $\Delta\tilde{x}_2=0.5$ and passes through the maximum of the potential barrier when the tip reaches the point $\tilde{x}=-1$ (i.e., the position above the equilibrium position of atom *2* in the unperturbed lattice). In Fig. 2b the solid curve gives the constant-force line obtained by solving the system (1) and the dashed curve gives that which would be obtained by scanning in the first regime (i.e., without any modification of the surface) above a vacancy initially

positioned at the point $\tilde{x}_v^{(0)}=-1$. After transition through the potential barrier, atom *2* is found at the point $\tilde{x}_2^{(0)}=0$, and the scanning continues as if the vacancy had been initially located at the point $\tilde{x}_v^{(0)}=-1$. It can be seen from Fig. 2b that it is difficult to determine from the behavior of the constant-force line in the interval $(-2,0)$ whether the vacancy was initially positioned at the point -1 or had migrated from the point 0. The only difference between these two cases is the smaller depth of the minimum (by an order of magnitude of $0.05a$) at the point -1.5 compared with point 0.5 if migration has taken place.

However, scanning in the region $\tilde{x}>0$ leads to more appreciable changes on the constant-force line. After the AFM tip has reached the region slightly to the right of the new position of atom *2* (note that the vacancy is positioned to the left), this atom becomes appreciably displaced to the left, although it does not pass through the potential barrier (the possibility of such a transition is discussed below). This displacement leads to substantial changes in the constant-force line: the tip falls by $\Delta z\approx 0.15$ and the maximum of the constant-force line has an unusual tapered profile (Fig. 2b). Consequently, the constant-force lines can be used to distinguish between the first and second scanning regimes.

We shall now analyze the third possible regime for scanning of the AFM tip above a vacancy, which is established at higher forces (in our calculations for $\Delta\tilde{l}_n>0.6$) and leads to deeper modification of the surface compared with the second regime. A characteristic feature of the third regime is that atom *2* passes through the potential barrier (and the vacancy migrates to its initial position) when the tip is to the left of the point $\tilde{x}=-1$. This leads to major changes in the entire subsequent scanning process.

When the tip is in the region to the right of atom *2* in its new equilibrium position at the point $\tilde{x}_2^{(0)}=0$, it transfers this atom through the potential barrier to the left to its initial position. This conclusion is an obvious consequence of the complete symmetry between the initial situation and that encountered after the first migration event. As a result, atom *3* is again closest to the vacancy on the right. After the AFM tip has moved to the right of this atom, it transfers it to the vacancy site. In this way, the vacancy is displaced to the right and occupies a position to the left of atom *4*, and so on.

Hence, at the beginning of scanning the tip displaces the vacancy to the left by a single interatomic distance a and during the subsequent scanning process drags it to the right. The tip thus transfers each successive atom past which it scans to the left through the potential barrier. The corresponding constant-force line (solid curve) calculated by solving the system (1) is plotted in Fig. 2c. Also plotted is the dashed constant-force line obtained by scanning the tip without any surface modification above a vacancy situated initially at the point $\tilde{x}_v^{(0)}=-1$. It can be seen from a comparison of Figs. 2b and 2c that it is impossible to determine the nature of the surface modification by the AFM tip from an analysis of scanning over the interval $(-2,1)$. However, the nature of the modification can easily be determined by analyzing the interval $x>1$. If the vacancy only undergoes a single migration to the left, the subsequent behavior of the constant-force line will be the same as that for a perfect lattice. If the tip drags the vacancy, these lines have clearly defined characteristic features.

2. HORIZONTAL FORCE COMPONENT AND ESTIMATE OF THE ACTIVATION ENERGY OF THE VACANCY MIGRATION PROCESS

As we have shown in the previous Section, when an AFM tip scans above a vacancy in the contact mode, various scenarios may take place, without and with different types of surface modification. The type of interaction between the tip and the vacancy is reflected not only on the constant-force lines but also on the profiles of the horizontal force component. In addition, as we shall show subsequently, these profiles contain important additional information which can be used to give a quantitative estimate of the activation energy of the vacancy migration process.

During scanning the horizontal forces acting on the tip can be measured independently (see, for example, Refs. 2–5) and they can be calculated to interpret the experimental results. Having determined the point $\tilde{z}$ on the constant-force surface and also the position vectors $\mathbf{r}_i$ of the sample atoms for each tip position $(\tilde{x},\tilde{y})$ as a result of solving the system (1), it is easy to determine the horizontal forces F_x and F_y acting on the tip during adiabatic scanning.[26] In dimensionless units, we have

$$\tilde{F}_x=\left(\frac{d}{a}\right)^{13}\sum_{i,j}\frac{\tilde{x}+\tilde{x}_{(t)j}-\tilde{x}_i}{\tilde{r}_{ij}^{14}}, \tag{9}$$

where $\tilde{x}_{(t)j}$ is the component of the position vector of the cluster atoms at the apex of the tip (a similar expression is obtained for the y component).

We shall now give the results of calculating the horizontal force using formula (9) for all three scanning regimes of the AFM tip above a vacancy calculated in the previous Section. For the first regime (no surface modification, $\Delta\tilde{l}_n<0.6$), the horizontal force profiles are plotted in Fig. 3a. As for the constant-force lines, the results here are also almost independent of Δl_n.

For the second regime, corresponding to the simplest modification of the surface ($\Delta\tilde{l}_n\approx 0.6$), the horizontal force profiles are shown by the solid curve in Fig. 3b (the dashed curve gives the horizontal force profile when no modification occurs and the vacancy occupies the position $\tilde{x}_v^{(0)}=-1$). It can be seen from Fig. 3b that when migration occurs, the force profile becomes asymmetric. This asymmetry is most appreciable when the tip scans directly above the vacancy, i.e., over the interval $(-1.5,0.5)$, and provides the fundamental possibility of determining the height of the potential barrier overcome by atom *2* (see below).

The profile of the horizontal force component for the third scanning regime is shown by the solid curve in Fig. 3c (the dashed curve, as in Fig. 3b corresponds to a vacancy at the point $\tilde{x}_v^{(0)}=-1$). As for the constant-force lines, a difference between Figs. 3c and 3b is only observed when the scanning is analyzed over the interval $\tilde{x}>1$.

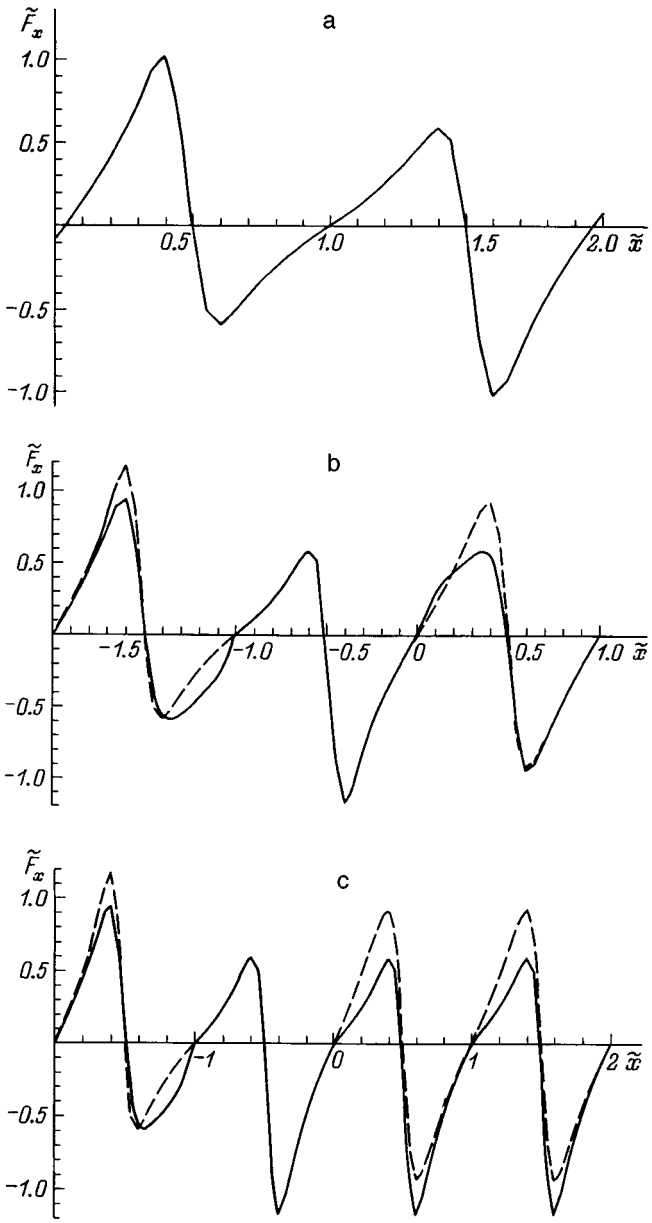

FIG. 3. Profiles of the horizontal force component when an AFM tip scans above a vacancy.

We shall now estimate the activation energy of the vacancy migration process using measured profiles of the horizontal force component similar to those shown in Figs. 3b and 3c. For this purpose we shall consider scanning of the tip near atom *2*, which is displaced to the right by the tip. In this region the scanning is fairly smooth, without any breaks in the constant-force line. We also note that the horizontal force component is determined not only by interaction of the tip with atom *2* but also with other atoms in the cell.

We shall postulate that an AFM was used to determine horizontal force profiles for two different scenarios: when the tip shifts atom *2* to the site of a vacancy ($\tilde{F}_x$) during scanning and when scanning takes place without any modification of the surface ($\tilde{F}_x^{(0)}$). These data were simulated above and are given by the solid and dashed curves in Fig. 3b, respectively. The work of the tip in displacing atom *2* to the vacancy site can then be expressed as (in conventional units)

$$A=-F_0 a\int_{\tilde{x}_1^{(0)}}^{\tilde{x}_2^{(0)}}(\tilde{F}_x-\tilde{F}_x^{(0)})d\tilde{x}, \tag{10}$$

where $\tilde{x}_1^{(0)}=-2$ and $\tilde{x}_2^{(0)}=-1$ are the x coordinates of atoms *1* and *2* in the unperturbed lattice.

This work is obviously proportional to the area between the solid and dashed curves in Fig. 3b.

It is expected that only one profile, $\tilde{F}_x$, will be obtained in an experimental study in the surface modification regime. However, it should be noted that when the tip scans in the range between $\tilde{x}_2^{(0)}=-1$ and $\tilde{x}_v^{(0)}=0$, we obtain $\tilde{F}_x=\tilde{F}_x^{(0)}$, as can be seen from Fig. 3b, i.e., the solid and dashed curves are the same. In addition, the profile $\tilde{F}_x^{(0)}$ is symmetric relative to the point $\tilde{x}_2^{(0)}=-1$. In consequence, expression (10) for the work may be rewritten as

$$A=-F_0 a\int_{\tilde{x}_1^{(0)}}^{\tilde{x}_v^{(0)}}\tilde{F}_x d\tilde{x}. \tag{11}$$

It is interesting to examine how the work (10) and the height of the potential barrier (7) agree in our model. Taking into account the relation $\Delta\tilde{l}_n=F_0/(aD_n)$, derived from expressions (5) and (8), we can express the latter as a function of $\Delta\tilde{l}_n$

$$\Delta E=\frac{F_0 a}{2\pi^2\Delta\tilde{l}_n}. \tag{12}$$

For $\Delta\tilde{l}_n\approx 0.6$, for which vacancy migration occurs, we have from expression (12) $\Delta E=0.084F_0 a$. On the other hand, calculations using formulas (10) or (11) and the profiles in Fig. 3b give $A=0.083F_0\,a$ (in order to obtain this result we performed numerical integration with the step $0.05a$, i.e., the same step as that used to calculate the curves in Fig. 3b). The quantitative agreement between the values obtained for A and ΔE confirms the consistency of the proposed method and the possibility of using formula (11) to determine the activation energy of the vacancy migration process using the horizontal force profiles obtained by scanning.

We shall now use our model to obtain approximate values of the scanning force for which a vacancy can be displaced by the AFM tip to a new position. We shall take the data for aluminum ($a=2.87$ Å) as an example. According to different estimates, the migration energy $E_m=\Delta E=A$ is $E_m^{(1)}\approx 0.9$ eV or $E_m^{(2)}\approx 0.5$ eV (Refs. 25 and 27). Thus, the scanning force $F_0=A/(0.083a)$ at which migration takes place is $F_0^{(1)}\approx 6.0$ nN or $F_0^{(2)}\approx 3.4$ nN. Such forces are quite normal for AFMs operating in the contact mode. Thus we expect that vacancy migration under the influence of an AFM tip will be recorded experimentally in the very near future.

3. DISCUSSION OF RESULTS

The foregoing simple model of harmonic displacements of surface atoms under the influence of an AFM tip (with the

exception of the displacements of atoms adjacent to a vacancy, which are considered as anharmonic) was used to calculate constant-force lines and profiles of the horizontal force component for scanning above a vacancy in a close-packed lattice. It was found that depending on the parameter $\Delta\tilde{l}_n$, which characterizes the mobility of the atoms adjacent to the vacancy, three fundamentally different scanning regimes are possible: without surface modification, migration of a single vacancy over a single interatomic distance in the direction opposite to the direction of scanning, and dragging of the vacancy by the AFM tip. These regimes were calculated assuming that the tip scans along the line of the surface atoms. However, it is easy to establish that vacancy migration under the influence of the tip takes place similarly when the tip scans at a small angle to this axis.

The physically different values of the dimensionless normalized parameter $\Delta\tilde{l}_n$ correspond to different values of the scanning force above the same sample and not to different elastic properties of the sample material. Thus, by systematically scanning above a vacancy with increasing forces and determining the profiles of the horizontal force component at each time, we can determine the force at which the first scanning regime is replaced by the second.

The most important conclusion reached in this study is that the activation energy for the migration of a surface vacancy can definitely be determined using measured profiles of the horizontal force component in scanning regimes where the surface is modified by the AFM tip. Estimates made using published data show that these regimes can be achieved for forces typical of an AFM in the contact mode. This suggests that the results obtained above may be confirmed experimentally.

The authors are grateful to V. I. Panov for interest in this work and for fruitful discussions.

[1] C. F. Quate, Surf. Sci. **299**, 980 (1994).
[2] S. N. Magonov and Myung-Hwan Whangbo, *Surface Analysis with STM and AFM* (VCH Verlagsgesellschaft mbH, Weinheim, 1996).
[3] C. A. J. Putman, M. Igarashi, and R. Kaneko, Appl. Phys. Lett. **66**, 3221 (1995).
[4] L. Howald, R. Lüthi, E. Meyer, and H. -J. Güntherodt, Phys. Rev. B **51**, 5484 (1995).
[5] M. R. Sørensen, K. W. Jacobsen, and P. Stoltze, Phys. Rev. B **53**, 2101 (1996).
[6] T. R. Albrecht and C. F. Quate, J. Vac. Sci. Technol. A **6**, 271 (1988).
[7] S. A. C. Gould, K. Burke, and P. K. Hansma, Phys. Rev. B **40**, 5363 (1989).
[8] S. Ciraci, A. Baratoff, and I. P. Batra, Phys. Rev. B **41**, 2763 (1990).
[9] H. Tang, C. Joachim, and J. Devillers, Surf. Sci. **291**, 439 (1993).
[10] E. V. Blagov, G. L. Klimchitskaya, A. A. Lobashev, and V. M. Mostepanenko, Pis'ma Zh. Tekh. Fiz. **21**(3), 73 (1995) [Tech. Phys. Lett. **21**, 121 (1995)].
[11] E. V. Blagov, G. L. Klimchitskaya, A. A. Lobashov, and V. M. Mostepanenko, Surf. Sci. **349**, 196 (1996).
[12] B. S. Good and A. Banerjea, J. Phys. **8**, 1325 (1996).
[13] E. V. Blagov, G. L. Klimchitskaya, A. A. Lobashev, and V. M. Mostepanenko, Zh. Tekh. Fiz. **67**(6), 77 (1997) [Tech. Phys. **42**, 655 (1997)].
[14] F. F. Abraham and I. P. Batra, Surf. Sci. **209**, L125 (1989).
[15] E. V. Blagov, G. L. Klimchitskaya, V. M. Mostepanenko *et al.*, Pis'ma Zh. Tekh. Fiz. **19**(8), 73 (1993) [Tech. Phys. Lett. **19**, 254 (1993)].
[16] E. V. Blagov, G. L. Klimchitskaya, and V. M. Mostepanenko, Surf. Rev. Lett. **4**(2), 2 (1997).
[17] J. R. Hahn, H. Kang, S. Song, and I. C. Jeon, Phys. Rev. B **53**(4), R1725 (1996).
[18] E. V. Blagov, G. L. Klimchitskaya, V. M. Mostepanenko *et al.*, Pis'ma Zh. Tekh. Fiz. **20**(1), 71 (1994) [Tech. Phys. Lett. **20**, 34 (1994)].
[19] R. Stumpf and M. Scheffler, Phys. Rev. Lett. **72**(2), 254 (1994).
[20] G. Boisvert, L. J. Lewis, M. J. Puska, and R. M. Nieminen, Phys. Rev. B **52**, 9078 (1995).
[21] U. Kürpich, A. Kara, and T. S. Rahman, Phys. Rev. Lett. **78**, 1086 (1997).
[22] E. V. Blagov, G. L. Klimchitskaya, and V. M. Mostepanenko, Preprint UFPB-DF-002 (1997).
[23] M. Bordag, G. L. Klimchitskaya, and V. M. Mostepanenko, Surf. Sci. **328**, 129 (1995).
[24] M. Born and K. Huang, *Dynamic Theory of Crystal Lattices* (Clarendon Press, Oxford, 1968).
[25] C. P. Flynn, *Point Defects and Diffusion* (Clarendon Press, Oxford, 1972).
[26] E. V. Blagov, G. L. Klimchitskaya, V. M. Mostepanenko, and M. Z. Smirnov, Zh. Tekh. Fiz. **67**(2), 134 (1997) [Tech. Phys. **42**, 243 (1997)].
[27] O. Takai, M. Doyama, and Y. Hisamatsi, in *Point Defects and Defect Interactions in Metals*, edited by J.-I. Takamura, M. Doyama, and M. Kiritani (North-Holland, Amsterdam, 1982).

Translated by R. M. Durham

Description of force surfaces in atomic force microscopy with allowance for the mobility of the lattice atoms

E. V. Blagov and V. M. Mostepanenko

Modus Research and Development Enterprise, 125047 Moscow, Russia

G. L. Klimchitskaya

Northwest Correspondence Polytechnical Institute, 191186 St. Petersburg, Russia

(Submitted August 12, 1997; resubmitted October 20, 1998)
Zh. Tekh. Fiz. **69**, 111–117 (August 1999)

The surfaces of constant force and the profiles of the horizontal component of the force during scanning of the tip of an atomic force microscope above the surface of a close-packed lattice in the contact mode are calculated taking account of the mobility of the lattice atoms. It is shown that when the mobility is taken into account, the previously observed discontinuities on the surface of constant force arise at smaller scanning forces on the tip above the surface than in the immobile-atom approximation. The force surfaces arising when scanning above vacancies are obtained. The possibility of using atomic force microscopy data for diagnostics of point defects on a solid surface is discussed.
[S1063-7842(99)02008-5]

INTRODUCTION

Atomic force microscopy has been widely used in recent years for investigations of surface microstructure (see, for example, Refs. 1 and 2). The highest resolution of an atomic force microscope (AFM) in the horizontal direction with respect to the surface under study is obtained in the so-called contact mode,[3] where the force interaction between the tip and the surface is determined primarily by quantum-mechanical repulsion of an exchange nature, acting between the nearest atoms of the tip and surface. The typical distance between the apex of the tip and the surface is $d \approx (0.5–0.7)a$, where a is the equilibrium distance in the unperturbed crystal lattice.

A large number of works, some employing an approximation where the lattice atoms are immobile (see, for example, Refs. 4–8), and some taking into account the relaxation of the atoms during scanning,[4,9,10] have been devoted to the calculation of the constant-force surfaces along which the AFM tip is scanned. As was shown in Refs. 6–8, when a monatomic tip is scanned above a close-packed lattice at sufficiently small initial heights, the constant-force surface has discontinuities which are concentrated above the interatomic gaps in the crystal lattice. As a result of the presence of discontinuities on the constant-force surface, the AFM tip drops into the surface under study at the location of a discontinuity. This has been repeatedly observed by different authors.[9,11,12] This phenomenon is also discussed in Ref. 13. To ensure continuous scanning with any initial height a tip containing of a cluster of several atoms at the apex must be used.[4,5,7,8,14]

Since discontinuities of constant-force surfaces occur for quite large repulsive forces, it is of interest to study this phenomenon and to calculate the continuous force surfaces arising for a cluster apex of a tip with allowance for the mobility of the lattice atoms. The simple model proposed in Ref. 4 can be used to describe the displacements of the atoms during scanning without modification of the surface under study. The basic assumptions of this model are as follows. The potential energy U of the system AFM tip plus experimental sample can be represented in its most general form as a sum of the interaction potential energy U_{ts} between the tip and sample atoms and the interaction energies U_t and U_s of the tip and sample atoms with one another. During scanning the coordinates of the tip and sample atoms change as a result of atomic relaxation. For a fixed position of the apex of the tip relative to the surface (for fixed x,y) the constant-force surface $z=z(x,y,F_0)$ and the coordinates of individual atoms can be obtained by solving the system of equations

$$-\frac{\partial U}{\partial z}=F_0, \quad \operatorname{grad}_{\mathbf{r}_i} U=0, \tag{1}$$

where $\mathbf{r}_i$ are the radius vectors of all tip and sample atoms whose mobility is taken into account.

The potential U_{ts} can be obtained as a sum of Lennard-Jones repulsive pair potentials,[1] since in the contact mode all other interactions, specifically, the attractive forces,[5] can be neglected. Following Ref. 4, in the next Section the potentials U_t and U_s are described in the harmonic approximation.

CONSTANT-FORCE SURFACES ON A CLOSE-PACKED LATTICE WITH ALLOWANCE FOR ATOMIC RELAXATION

As noted in the introduction, the constant-force surfaces during scanning of an AFM tip above the sample surface can be found by solving the system of equations (1). The explicit form of the total potential energy U of the system tip plus sample is required for specific calculations. To obtain results which are independent of the semiempirical exchange interaction constant, we introduce the dimensionless coordinates,

force, and energy $\tilde{\mathbf{r}}=\mathbf{r}/a$, $\tilde{F}=F/F_0$, and $\tilde{U}=U(F_0a)$, respectively, where $F_0=12\alpha/d^{13}$ is the force acting on a monatomic tip located at an initial height d above a surface atom. Then

$$\tilde{U}_{ts}=\frac{d^{13}}{12a^{13}}\sum_{i,j}\frac{1}{\tilde{r}_{ij}^{12}}, \tag{2}$$

where the summation extends over all atoms of the tip and sample and $\tilde{r}_{ij}$ is the distance between the tip and sample atoms.

With allowance for the symmetry properties of a close-packed lattice, the potential energy U_s of the sample atoms, which in our approach is interpreted as the elastic vibrational energy of the lattice,[15] can be written as

$$\tilde{U}_s=\frac{1}{2}\sum_{i,k}\tilde{D}_s^{(i,k)}(\Delta\tilde{x}_{(s)i}\Delta\tilde{x}_{(s)k}+\Delta\tilde{y}_{(s)i}\Delta\tilde{y}_{(s)k}+\Delta\tilde{z}_{(s)i}\Delta\tilde{z}_{(s)k}). \tag{3}$$

Here $\Delta\tilde{x}_{(s)i}=\tilde{x}_{(s)i}-\tilde{x}_{(s)i}^{(\sigma)}$ and so on describe the displacements of the equilibrium positions $\tilde{x}_{(s)i}$ occupied by the lattice atoms during the scanning process as a function of their equilibrium positions $\tilde{x}_{(s)i}^{(0)}$ in a lattice which is not interacting with the tip. The dimensionless constants appearing in Eq. (3) can be expressed in terms of the matrix of force constants of the solid under study as

$$\tilde{D}_s^{(i,k)}=\frac{ad^{13}}{12\alpha}D_s^{(i,k)}. \tag{4}$$

The dimensionless potential energy $\tilde{U}_t$ of the tip atoms (with s replaced by t) can be represented similarly to Eq. (3).

Two models of an AFM tip are used below to calculate the force surfaces: a monatomic tip, whose apex contains only one atom, and a cluster tip, whose apex contains a cluster of seven atoms. In the first case, taking account of the displacement of a single atom at the apex relative to the other atoms in the tip, which actually do not interact with the atoms in the sample, has no effect on the results obtained and need not be taken into account at all. Then the summation in Eq. (2) extends only over the atoms in the sample. In the second case, one of the atoms in the cluster drops to a distance h relative to the plane of the other six atoms arranged in a hexagon with sides of length b. If it is assumed that the tip is a paraboloid of revolution with radius of curvature $R=5a/3$, then stable scanning for arbitrary initial heights obtains for $b=0.85a$ and $h=b^2/(2R)\approx 0.22a$.[7,8,16,17]

Two different approximations were used to solve the system (1). In the first one it was assumed that each atom of the surface layer of the sample is displaced only by the direction action of the tip, and this displacement does not lead to displacements of neighboring lattice atoms. This assumption, which is valid for sufficiently small displacements, means that the constants $\tilde{D}_s^{(i,k)}$ in Eq. (3) are zero for $i\neq k$. Since it also follows from symmetry considerations that these constants do not depend on the number of the atom, it is expedient to introduce the notation $\tilde{D}_s^{(i,j)}\equiv\tilde{D}_s$. It is convenient to write the parameter $\tilde{D}_s$ as $\tilde{D}_s=1/\Delta\tilde{l}_s$, where $\Delta\tilde{l}_s$ is the elastic displacement of an atom of the sample acted on by a force equal to the scanning force.

A two-dimensional image of a constant-force surface is presented in Fig. 1. This image was obtained by solving the system (1) using the first approximation with initial scan height $d=0.65a$ and mobility of lattice atoms $\Delta\tilde{l}_s=0.05$ (the scan force shifts an atom by $0.05a$ from its position of equilibrium). Here and below, the two-dimensional image is the result of sectioning the corresponding three-dimensional surfaces by 10 horizontal planes separated by the step $\Delta\tilde{h}=(\tilde{z}_{\max}-\tilde{z}_{\min})/10$. A lighter tone corresponds to a greater height, and a darker tone corresponds to a smaller height. Smooth sections of the constant-force surface in regions above the atoms of the surface under study and sharp peaks in the regions between atoms where the constant-force surface is discontinuous are clearly seen in Fig. 1. An attempt to solve the system (1) numerically near the discontinuities leads to a divergent iteration process. We note that when the mobility of the sample atoms is taken into account, the initial scanning heights for which the constant-force surfaces are discontinuous increase. Thus, in the approximation of immobile lattice atoms the condition that there be no discontinuities for a monatomic tip is $d>0.61a$.[7,8]

The second, and more accurate approximation used to solve the system (1) took account of the fact that the displacement of the surface atoms by the direct action of the AFM tip leads to displacements of their nearest neighbors also, including the neighbors in the second atomic layer of the sample. In other words, displacements of all atoms in the first coordination spheres of the atoms in the sample that are directly affected by the tip were taken into account. Just as above, it was assumed that the tip directly displaces three surface atoms located at the vertices of an equilateral triangle. To simplify the calculations, taking account of the smallness of the displacements of the atoms in a regime with no modification of the surface, it was also assumed that the displacements of neighboring atoms are independent of one another. Then the dimensionless potential energy (3) can be written as

$$\tilde{U}_s=\frac{1}{2}\tilde{D}_s\sum_i[(\Delta\tilde{x}_{(s)i})^2+(\Delta\tilde{y}_{(s)i})^2+(\Delta\tilde{z}_{(s)i})^2]+\tilde{D}_s^{(1)}\sum_{i=1}^{3}\left(\Delta\tilde{x}_{(s)i}\sum_j\Delta\tilde{x}_{(s)j}+\Delta\tilde{y}_{(s)i}\sum_j\Delta\tilde{y}_{(s)j}+\Delta\tilde{z}_{(s)i}\sum_j\Delta\tilde{z}_{(s)j}\right)+\tilde{D}_s^{(2)}\sum_{i=1}^{3}\left(\Delta\tilde{x}_{(s)i}\sum_k\Delta\tilde{x}_{(s)k}+\Delta\tilde{y}_{(s)i}\sum_k\Delta\tilde{y}_{(s)k}+\Delta\tilde{z}_{(s)i}\sum_k\Delta\tilde{z}_{(s)k}\right). \tag{5}$$

Here $\tilde{D}_s^{(1)}\equiv\tilde{D}_s^{(i,j)}$, where the indices i and j enumerate the nearest-neighbor atoms in the surface layer of a close-packed lattice, and $\tilde{D}_s^{(2)}\equiv\tilde{D}_s^{(i,k)}$, where i and k enumerate the

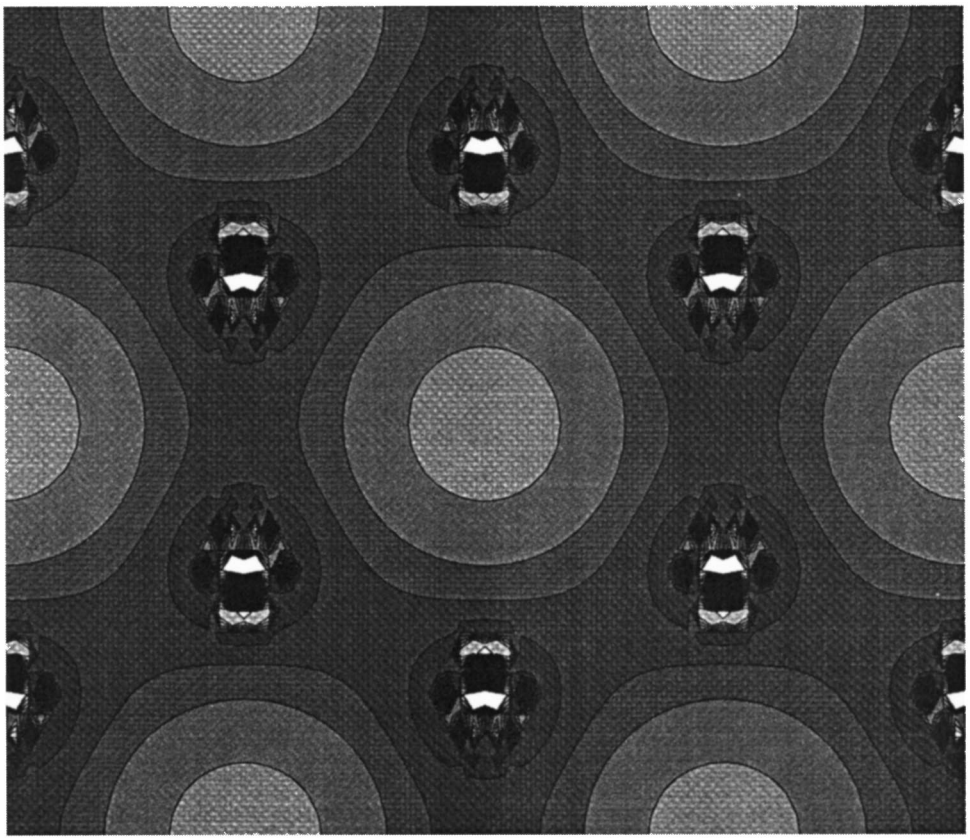

FIG. 1. Two-dimensional image of discontinuities on a constant-force surface with a monatomic tip scanned above the surface of a close-packed lattice.

nearest-neighbor atoms in the surface and second layers of the sample. From obvious geometric considerations we have $\tilde{D}_s^{(2)} = -\tilde{D}_s/6$ and $\tilde{D}_s^{(1)} = -\tilde{D}_s/4$.

The second, and more accurate, approximation for solving the system (1) will exacerbate even more the situation with discontinuities arising in the constant-force surfaces when a monatomic tip is scanned above a closed-packed lattice. Thus, for $d = 0.7a$ the first approximation gives continuous constant-force surfaces all the way up to atomic mobilities characterized by a value $\Delta\tilde{l}_s = 0.09$, whereas in the second case discontinuities are observed even for $\Delta\tilde{l}_s = 0.05$. This discontinuity, of width $\approx 0.05a$, occurs in the region above an interatomic gap. At the same time, the results are essentially identical in the region where both approximations give continuous constant-force surfaces. To illustrate this assertion, in Fig. 2 we show the force contours calculated in the first (dashed curve) and second (solid curve) approximations, with an initial tip height $d = 0.7a$ and $\Delta\tilde{l}_s = 0.04$.

Using the monatomic tip model, we shall now consider the scanning of a cluster tip above the surface of a close-packed lattice for values of the parameters such that continuous scanning is impossible. For simplicity, we shall use the first approximation to solve the system (1) (as illustrated above, this does not introduce any substantial errors when calculating continuous constant-force surfaces), and of the seven atoms we shall take into account the displacement of the atom located at the apex of the tip relative to the other six atoms. It is convenient to represent the dimensionless parameter $\tilde{D}_t$, related with the force constant of the tip material, in the form $\tilde{D}_t = 1/\Delta\tilde{l}_t$.

A three-dimensional image of the constant-force surface was calculated for the parameters $d = 0.5a$ (all results below are presented for this value of d), $\Delta\tilde{l}_s = 0.05$ and $\Delta\tilde{l}_t = 0.01$. The crystallographic axes of a cluster at the apex of the tip and the surface were assumed to be parallel and coincident with the scan direction. This surface differs only

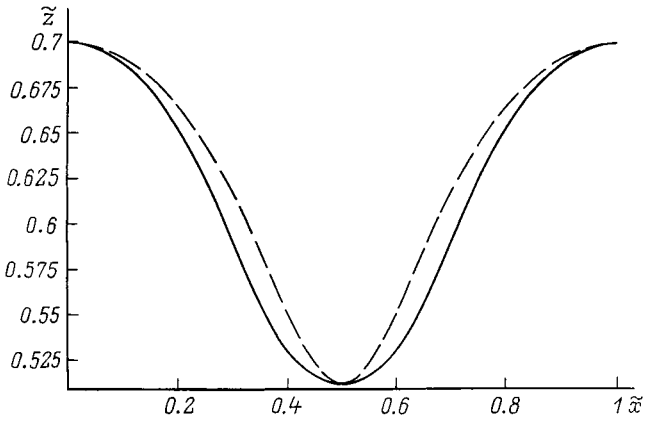

FIG. 2. Comparison of force contours with a monatomic tip scanned above a line of atoms in the sample using different approaches to describe the mobility of the atoms.

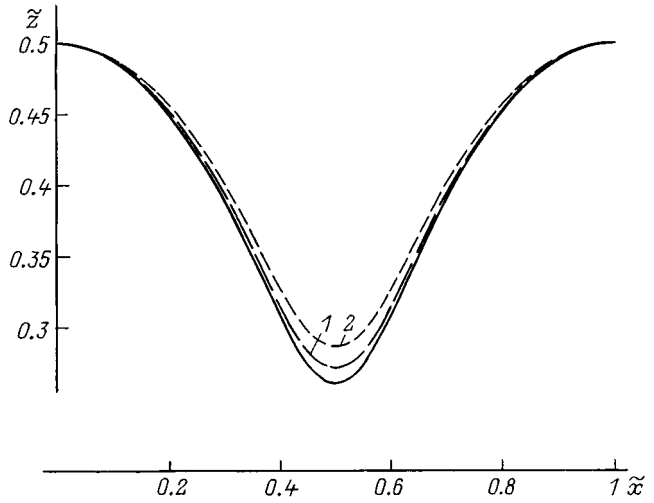

FIG. 3. Force contours with a cluster tip scanned along a line of atoms and different mobilities of atoms in the sample and in the tip.

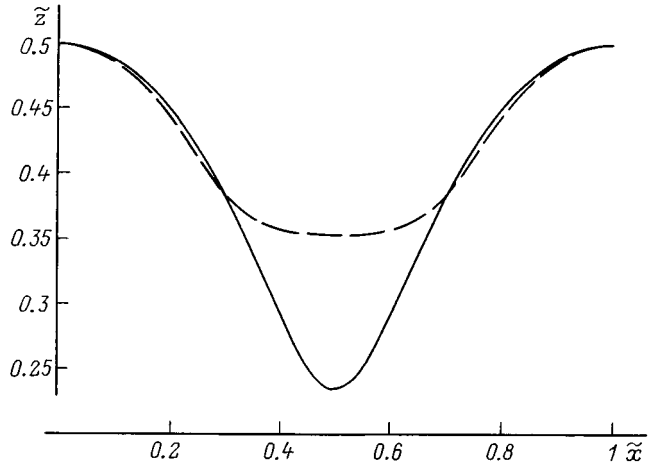

FIG. 4. Force contours versus the orientation of a cluster at the apex of the tip.

by the depth of the relief from the corresponding surface in the immobile-atom model[8] (in the present case, the differential of the relief is $\Delta\tilde{z}=0.411$ as opposed to $\Delta\tilde{z}=0.39$ assuming the atoms of the tip and samples to be immobile). It is interesting to note that when the mobility of the atoms is taken into account, the form of the relief also changes somewhat, the change consisting in the fact that on a force contour corresponding to scanning along a line of lattice atoms the relief differential decreases with increasing mobility. As an illustration, the force contours obtained with various sample and tip atom mobilities are presented in Fig. 3: $\Delta\tilde{l}_s=0.05$ and $\Delta\tilde{l}_t=0.05$ (solid curve), $\Delta\tilde{l}_s=0.1$ and $\Delta\tilde{l}_t=0.05$ (dashed curve *1*), and $\Delta\tilde{l}_s=0.1$ and $\Delta\tilde{l}_t=0.1$ (dashed curve *2*). In this case the relief differentials are, respectively, $\Delta\tilde{z}=0.239$, 0.288, and 0.213, but the change in relief remains beyond the sensitivity of the AFM.

We note that the computational results for the constant-force surfaces depend on the orientation of the tip. As an example, force contours obtained for $\Delta\tilde{l}_s=0.05$ and $\Delta\tilde{l}_t=0.01$ with the crystallographic axes of the tip parallel to those of the surface (solid curve) and with an angle of 30° between them (dashed curve) are presented in Fig. 4. As one can see from the figure, the maximum vertical difference between the two lines does not exceed $\Delta\tilde{z}=0.12$; this can only have a small effect on the experimental constant-force surfaces.

We shall now briefly discuss the computational results for constant-force surfaces with a cluster tip scanned above the surface of a close-packed lattice containing a point defect of the vacancy type. In accordance with what we have said above, in calculating continuous force surfaces the results obtained using the first and second approximations to solve the system (1) are essentially identical. For this reason, to

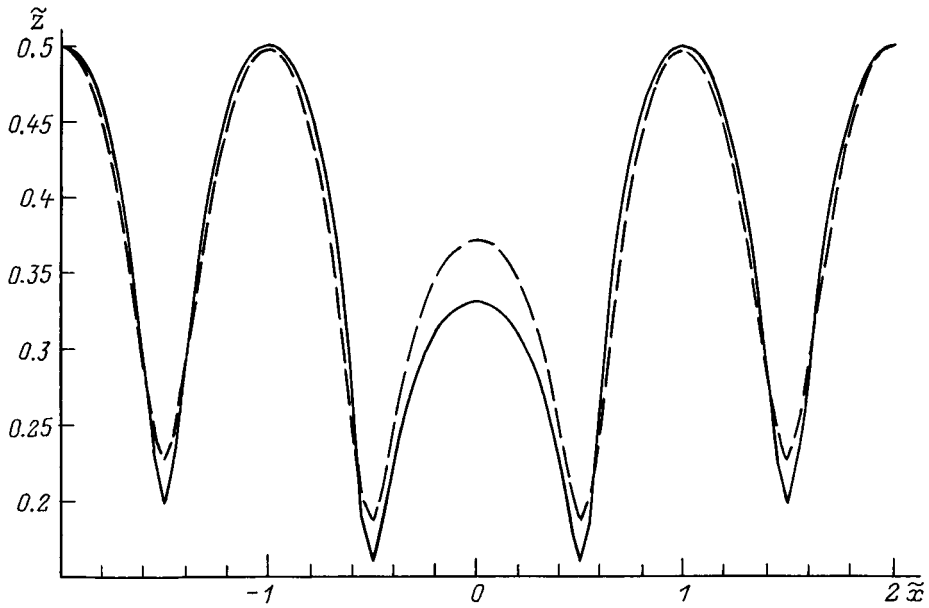

FIG. 5. Force contours with a cluster tip scanned above a vacancy for different mobilities of an atom at the apex of the tip.

simplify the calculations the first approximation was used here and in the next Section.

The calculations of the constant-force surface during scanning of a cluster tip with the mobilities of sample and tip atoms $\Delta l_s = 0.05$ and $\Delta l_t = 0.01$ showed that the total differential of the relative heights is $\Delta\tilde{z} = 0.411$, and the difference of the heights of the peaks above the surface atoms and above a vacancy is $0.13a$ (in the model with immobile lattice atoms these quantities are,[8] respectively, $\Delta\tilde{z} = 0.43$ and $0.15a$ (Ref. 8). We note that for scanning above a vacancy the results depend more strongly on the mobility of a tip atom than for scanning above an unperturbed close-packed lattice. As an example, force contours calculated for a rigid cluster at the apex of the tip ($\Delta\tilde{l}_t = 0$; solid curve) and for the mobility of an atom at the apex of the tip $\Delta\tilde{l}_t = 0.05$ (dashed curve) are presented in Fig. 5. At the same time, just as for an unperturbed lattice, taking account of the possible mobility of atoms in the scanning mode without the surface under investigation being modified gives only negligible changes in the constant-force surface, which lie outside the sensitivity of the AFM. Of course, this only strengthens the conclusion that a vacancy in a close-packed lattice can be discovered according to the change in the constant-force surface above the vacancy.[6–8]

EFFECT OF THE MOBILITY OF ATOMS ON THE HORIZONTAL COMPONENT OF THE FORCE

The measurement of the horizontal component of the force during scanning of an AFM tip along the surface of constant vertical component gives substantial additional information that can be used, specifically, for diagnostics of point surface defects by the AFM method. For this reason, it is of interest to determine the effect of the mobility of lattice atoms on the form of the computed profiles of the horizontal component of the force.

Assume that a point on the constant-force surface $\tilde{z}$, the coordinates of the sample atoms $\tilde{x}_{(s)i}$, $\tilde{y}_{(s)i}$, and $\tilde{z}_{(s)i}$ and the coordinates of the tip atoms $\tilde{x}_{(t)i}$, $\tilde{y}_{(t)i}$, and $\tilde{z}_{(t)i}$ have been found by solving the system of equations (1) for each posi-

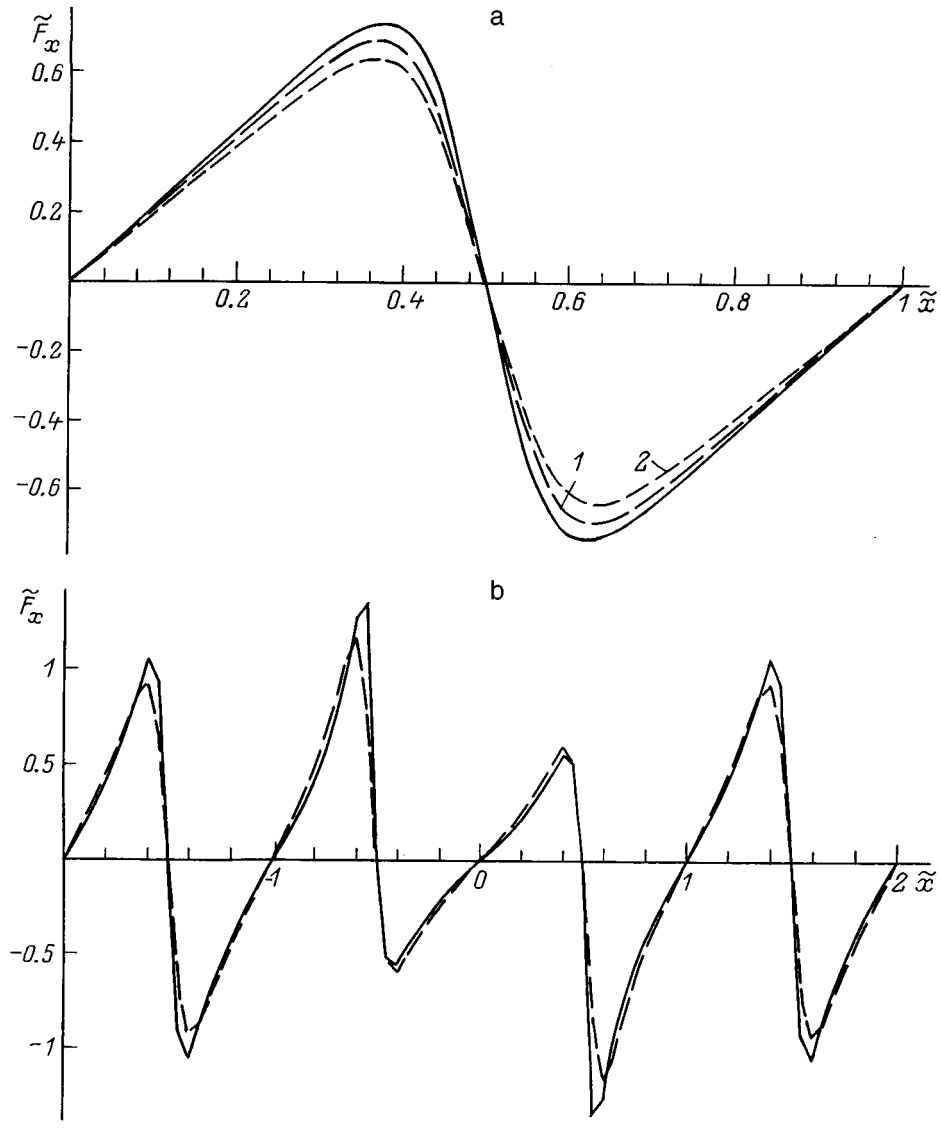

FIG. 6. Horizontal force with a cluster tip scanned along a line of atoms on the surface and different mobilities of the atoms.

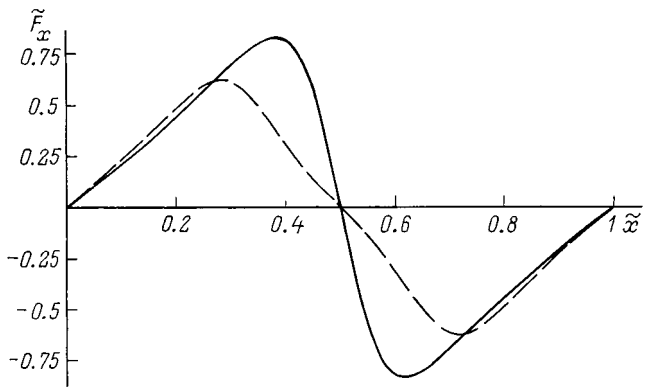

FIG. 7. Horizontal component of the force versus the orientation of a cluster at the apex of the tip.

tion of the tip relative to the surface $(\tilde{x},\tilde{y})$. Then the dimensionless horizontal force is given by the expression

$$\tilde{F}_x=\left(\frac{d}{a}\right)^{13}\sum_{ij}\frac{\tilde{x}+\tilde{x}_{(t)j}-\tilde{x}_{(s)i}}{\tilde{r}_{ij}^{14}} \tag{6}$$

and a similar expression for the y component.

Curves of the horizontal force component $\tilde{F}_x$ versus x for scanning along a line of atoms on a defect-free surface with various mobilities of the sample and tip atoms are presented in Fig. 6a: $\Delta\tilde{l}_s=0.05$ and $\Delta\tilde{l}_t=0.05$ (solid curve), $\Delta\tilde{l}_s=0.1$ and $\Delta\tilde{l}_t=0.05$ (dashed curve *1*), and $\Delta\tilde{l}_s=0.1$ and $\Delta\tilde{l}_t=0.1$ (dashed curve *2*); compare with Fig. 3 for the force contours. As is evident from Fig. 6a, the horizontal component of the force vanishes directly above the atoms of the surface under study and at points midway between them. Moreover, for an unperturbed close-packed lattice $|\tilde{F}_{x,\min}|=\tilde{F}_{x,\max}$, as should be on the basis of symmetry considerations. These results agree with the results obtained previously in the immobile-atom approximation.[16,17] The force differentials $\tilde{F}_{x,\max}-\tilde{F}_{x,\min}$ for the lines in Fig. 6a are, respectively, 1.482, 1.390, and 1.276. Therefore, as the mobility of the atoms increases, the force differential decreases, but this cannot affect the interpretation of the experimental data.

Figure 6b shows the horizontal force component $\tilde{F}_x$ versus x when a tip with a rigid cluster at the apex (solid curve) and with the mobility of an atom in the tip $\Delta\tilde{l}_t=0.05$ (dashed curve) is scanned along a line of atoms containing a vacancy. The mobility of the atoms in the sample is $\Delta\tilde{l}_s=0.05$. Peaks with various heights are clearly seen in Fig. 6b. Thus, point defects in the surface under study appreciably alter the profile of the horizontal component of the force. However, the mobility of a tip atom has little effect on the result, just as in the case of an unperturbed lattice.

Thus far it has been assumed that the crystallographic axes of the surface are parallel to those of the cluster at the apex of a tip. Figure 7 shows for comparison the horizontal force component $\tilde{F}_x$ versus $\tilde{x}$ for scanning above a defect-free surface for parallel axes (solid curve) and axes making an angle of 30° (dashed curve). The other parameters are $\Delta\tilde{l}_s=0.05$ and $\Delta\tilde{l}_t=0.01$. As one can see from Fig. 7, the lines of horizontal component of the force are much more sensitive to the tip orientation than the force contours (compare with Fig. 4). Therefore, in principle, the orientation of a cluster tip relative to the surface under study can be determined experimentally according to the profiles of the horizontal force.

DISCUSSION

In the present work the effect of the mobility of lattice atoms on the constant-force surface and the profiles of the horizontal component of the force for scanning of the AFM tip above a close-packed lattice was studied.

It was shown that the appearance of discontinuities on the constant-force surface depends strongly on the mobility of the atoms in the sample. In the immobile-atom approximation, discontinuities arise only if the initial scanning heights are sufficiently small $(d/a\leqslant 0.61)$, whereas discontinuities are also observed for $d\approx(0.7–0.8)a$ and realistic values of the mobility (this result was predicted qualitatively in Ref. 8). This makes the discontinuities more important from the standpoint of interpreting experimental results.

Constant-force surfaces were calculated taking into account the mobility of atoms in the sample and in the tip above a defect-free surface and above a vacancy. Different approximations were used to describe the mobility of the lattice atoms, and it was confirmed that in the continuous scan mode the results obtained are essentially identical.

A calculation of the horizontal component of the force taking account of the mobility of the atoms demonstrated that the profile of this force near a vacancy is substantially different from the force on a defect-free section of the surface. It was shown that profiles of the horizontal component of the force are more sensitive than a constant-force surface to the relative orientation of the surface under study and the cluster of atoms at the apex of the tip.

The results obtained can be used for diagnostics of point defects by atomic force microscopy (see Ref. 18, where a vacancy was first detected in this manner).

We thank V. I. Panov for his interest in this work and for fruitful discussions.

[1] D. Sarid, *Scanning Force Microscopy with Applications to Electric, Magnetic and Atomic Forces* (Oxford University Press, New York, 1991).
[2] C. F. Quate, Surf. Sci. **299/300**, 980 (1994).
[3] T. R. Albrecht and C. F. Quate, J. Vac. Sci. Technol. A **6**, 271 (1988).
[4] S. A. C. Gould, K. Burke, and P. K. Hansma, Phys. Rev. B **40**, 5363 (1989).
[5] S. Ciraci, A. Baratoff, and I. P. Batra, Phys. Rev. B **41**, 2763 (1990).
[6] E. V. Blagov, G. L. Klimchitskaya, A. A. Lobashev, and V. M. Mostepanenko, Pis'ma Zh. Tekh. Fiz. **21**(3), 73 (1995) [Tech. Phys. Lett. **21**, 121 (1995)].
[7] E. V. Blagov, G. L. Klimchitskaya, A. A. Lobashov, and V. M. Mostepanenko, Surf. Sci. **349**, 196 (1996).
[8] E. V. Blagov, G. L. Klimchitskaya, A. A. Lobashev, and V. M. Mostepanenko, Zh. Tekh. Fiz. **67**(6), 77 (1997) [Tech. Phys. **42**, 655 (1997)].
[9] F. F. Abraham and I. P. Barta, Surf. Sci. **209**, L125 (1989).
[10] U. Landman, W. D. Luedtke, and M. W. Ribarsky, J. Vac. Sci. Technol. A **7**, 2829 (1989).
[11] L. Howald, R. Lüthi, E. Meyer, and H.-J. Güntherodt, Phys. Rev. B **51**, 5484 (1995).

[12] R. Lüthi, E. Meyer, and L. Howald, Tribol. Lett. **1**, 129 (1995).
[13] B. S. Good and A. Banerjea, J. Phys. **8**, 1325 (1996).
[14] H. Tang, C. Joachim, and J. Devillers, Surf. Sci. **291**, 439 (1993).
[15] M. Born and K. Huang, *Dynamic Theory of Crystal Lattices* (Clarendon Press, Oxford, 1968).
[16] E. V. Blagov, G. L. Klimchitskaya, V. M. Mostepanenko, and M. Z. Smirnov, Zh. Tekh. Fiz. **67**(2), 134 (1997) [Tech. Phys. **42**, 243 (1997)].
[17] E. V. Blagov, G. L. Klimchitskaya, and V. M. Mostepanenko, Surf. Rev. Lett. **4**, 271 (1997).
[18] J. R. Hahn, H. Kang, S. Song, and I. C. Jeon, Phys. Rev. B **53**, R1725 (1996).

Translated by M. E. Alferieff

Method for controlled extraction of energy stored in a capacitor

D. I. Adeĭshvili and N. F. Shul'ga

Kharkov Physicotechnical Institute National Science Center, 310108 Kharkov, Ukraine

V. P. Kortkhondzhiya

Institute of Physics, Georgian Academy of Sciences, 380077 Tbilisi, Georgia
(Submitted July 30, 1998)
Zh. Tekh. Fiz. **69**, 118–121 (August 1999)

A method of controlled extraction of the energy stored in a capacitor with the required power over a long period of time is proposed. It is shown that capacitors can be used as autonomous source of electric power. © *1999 American Institute of Physics.* [S1063-7842(99)02108-X]

1. Capacitors are electric storage devices with unique properties. Their internal resistance is low, and they can have a specific energy capacity and are ecologically clean sources of electricity.[1–3] However, such sources are not widely used because of uncontrollable character of their discharge process. In the present paper a method is proposed for controlled, loss-free extraction of the energy stored in a capacitor with the required power on a specific load (user) over a quite long working time. The proposed method shows that capacitors can be used as efficient autonomous sources of electric power.

2. The essence and principle of operation of the method are as follows (Fig. 1). A small portion of the energy stored in a capacitor with a high capacitance C_1 (we call this the tank capacitor) is extracted with a definite frequency f and reciprocal duty factor $\varkappa$ by another capacitor (we call this the dispenser capacitor) with capacitance C_2, such that $C_2 \ll C_1$. Then, this small energy dose is transferred to a load R_l with the same frequency and reciprocal duty factor. Such a scheme with an appropriate choice of parameters (dispenser capacitance, duty factor, frequency, switches, and transfer devices) permits controlling the discharge of the capacitor C_1 and permits feeding energy uniformly to a user R_l for a long working time T with the required nominal power W_n.

In summary, the main idea of the method is that instead of direct, uncontrolled discharging of a tank capacitor C_1 on a load R_l, a dispenser C_2 is introduced to regulate the process of discharging the tank capacitor. In the absence of a dispenser, the energy stored in the tank capacitor, depending on the intrinsic discharge time $t^d = R_l C_1$ and the fixed working time T, will either be spilled out immediately at the start of the process with $t^d \ll T$ according to an exponential law or if $t^d \gg T$, for the same stored energy the required power W_n will not be released to the load. The dispenser permits the energy stored in the tank to be extracted in fixed portions, uniformly and with the required nominal power, by lengthening or shortening the discharge time t_E to the value of the required working time T. In the scheme presented, the main variable physical quantities are the voltage, duty factor, frequency, pulse duration, and current in a pulse. By varying these quantities from zero to their nominal values it is possible to regulate the delivery of energy to the load R_l and thereby to control the power delivery as required in a range from zero to the nominal value, i.e., $0 \leqslant W \leqslant W_n$. The nominal numerical values of the indicated physical quantities are determined by the specific problem. Evidently, switches will play a large role in this control process, since the above-indicated physical quantities all can be changed by using these switches exclusively. They can be implemented structurally either by a purely mechanical method or by a purely electronic method or by both methods simultaneously.

It should be noted especially that the proposed electric scheme functionally consists of two parts. The first part is a circuit that charges the dispenser from the charged tank, and the second part is a circuit that discharges the dispenser on a load R_l. The operation of the proposed scheme is based on the use of capacitors containing dielectrics with short polarization times (from 10^{-14} to 10^{-8} s). Then energy can be transferred rapidly (in $1-10^{-3}$ s) to the load without energy loss when charging the capacitor C_2 from C_1. On this basis the operation of the first part of the scheme satisfies the laws of electrostatics, since the dispenser is charged by connecting it directly to the tank, without a charging resistance, and for this reason it occurs without losses. This part of the scheme

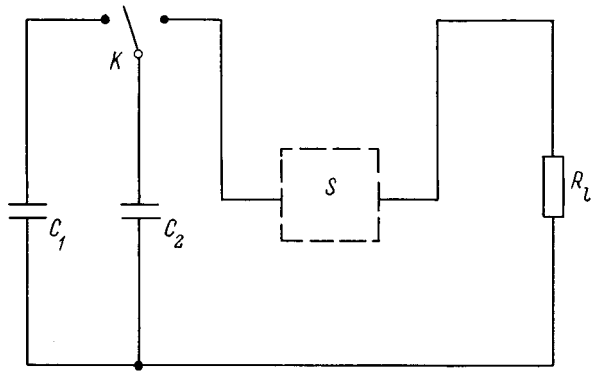

FIG. 1. Diagram of a scheme for controlled extraction of the energy stored in a capacitor: C_1 — tank capacitor; C_2 — dispenser capacitor; K — fast switch; S — all possible devices for transferring power to the load R_l.

contains only energy losses due to leakage currents on the enormous insulating resistance of the dielectric in the tank capacitor C_1. The second part of the scheme operates according to the laws of electrodynamics, since the dispenser is discharged across a discharge resistance (load). This part of the scheme is a dissipative system, and its operation is accompanied by some energy losses.

We note that the term ''dispenser'' is also used in Ref. 4. However, in our scheme the dispenser is charged from an autonomous source of electric power — the tank capacitor — and not from a charging device, as considered in Ref. 4, that is constantly switched on and is powered by an external electric circuit. Moreover, in our scheme the dispenser serves to transfer a definite portion of energy directly to the user (useful load) and not for charging a capacitor with a capacitance much greater than that of the dispenser. However, in practice the dispenser scheme of Ref. 4 can be used to recover electric energy.

3. We shall now give a quantitative treatment of the method. Let us assume that we are required to supply to an individual autonomous user R_l nominal electric power W_n for a quite long working time T with the possibility of continuously regulating the delivery of this power from zero to a nominal value, i.e., $0 \leqslant W \leqslant W_n$. This requires an amount of electricity

$$\Delta E = W_n T. \tag{1}$$

If a capacitor is chosen as an autonomous source of electric power, then a storage capacitor with the appropriate capacitance C_1 will be required. As is well known, the maximum electric energy that can be stored in capacitors with capacitance C is $E_0 = CV_0^2/2$, where V_0 is the maximum voltage between the capacitor plates. However, the electric energy stored in a capacitor, however large its capacity, cannot always be extracted with the required nominal power W_n during a long working time interval T. Indeed, if a maximally charged capacitor is connected directly to a load, then the energy of the capacitor is discharged according to an exponential law in time $t_E = R_l C/2$, and an amount of energy $\Delta E = 0.64E_0$ will be released from the capacitor in this time t_E. The fixed working time in which this energy is extracted from the capacitor must be $T = t_E$. Then the extracted power will be $W = 0.64CV_0^2/(2T)$.

In practice, however, it is by no means always possible to find a capacitor with parameters (dielectric, breakdown voltage, specific volume capacitance, and others) satisfying the conditions noted above: the release of the required nominal power W_n to a load R_l in a working time T. If the capacitance $C \ll 2T/R_l$, then the capacitor will be discharged in time $t_E \ll T$ and a high power $W \gg W_n$ will be released in the load (which is very undesirable). However, if $C \gg 2T/R_l$, then $t_E \gg T$ and a lower power $W \ll W_n$ will be released in the load (which does not satisfy the user). For this reason the discharging of the capacitor must be regulated: In the first case the discharge time t_E needs to be lengthened and in the second case the discharge time needs to be shortened to the working time T. The discharge process must be regulated almost with no losses of energy stored in the tank.

This result is illustrated in Fig. 2, where the relation

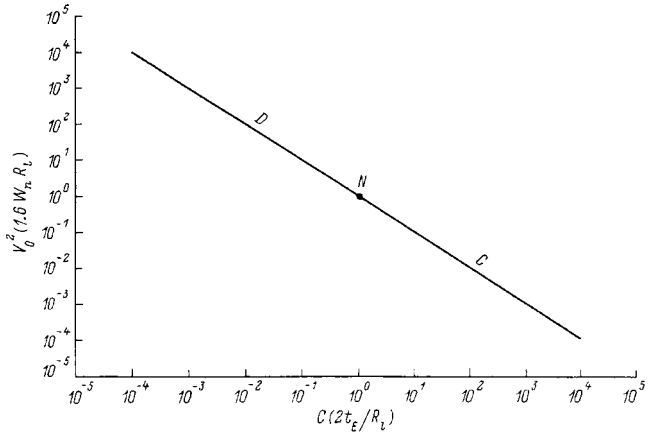

FIG. 2. Square V_0^2 of the voltage between the capacitor plates versus the capacitance C of the capacitor: W_n — nominal power, R_l — load, t_E — energy discharge time of the capacitor, $t_E = R_l C/2$; N — ''normal point'' — with coordinates $(1,1)$.

between the squared maximum voltage V_0^2 between the capacitor plates with the maximum accumulated energy E_0 and the capacitance C for the same dielectric is displayed on a logarithmic scale. The quantity $V_0^2/(1.6W_nR_l)$ is plotted along the ordinate, and the quantity $CR_l/(2t_E)$ is plotted along the abscissa. It is easy to verify that the condition $t_E = T$ holds on this plot only at the point with the coordinates $(1,1)$. If $C < 1$, then $t_E < T$ $(W > W_n)$ and the discharge process needs to be extended (the region CD in the figure). However, if $C > 1$, then $t_E > T$ $(W < W_n)$ and the discharge process needs to be compressed (the region C). We note that the plot presented is universal for the problems being considered, since it is normalized explicitly with respect to all fixed parameters E_0, W_n, T, and R_l and is valid for all dielectrics.

The point $(1,1)$ on the plot is a reference point. At this point not only is the method normalized with respect to all fixed parameters, but the required working time $T = t_E$ is matched with the required maximum energy $E_0 = 1.6\Delta E$ stored in the tank C_1, which is very important to do. At the point $(1,1)$ two conditions must hold simultaneously: $T = RC_1/2$ on the one hand and $E_0 = 1.6W_n/T$ on the other. This is possible, since the required maximum energy E_0 with fixed capacitance $C_1 = 2T/R_l$ can be provided by using an appropriate dielectric with the correct volume ν, according to the well-known formula $E_0 = \varepsilon_0 \varepsilon_r U^2 \nu/2$, where ε_0 is the permittivity of free space, ε_r is the dielectric constant, and U is the electric field in the dielectric. We shall call capacitors for which $t_E = T$ and $E_0 = 1.6W_nT$ hold simultaneously ''normal.'' In the plot they correspond to the point $(1,1)$. On this basis we shall also term the point $(1,1)$ ''normal.'' The coordinates on the C axis indicate simultaneously the required shortening or lengthening of the discharge time t_E to the value $t_E = T$ at the unit point $(1,1)$. The proposed method is based precisely on effectuating this process of bringing the discharge time t_E of the capacitor to the unit (normal) point $(1,1)$ on the plot, where $t_E = T$ and $E_0 = 1.6W_nT$.

The plot displayed in Fig. 2 makes it possible to determine whether or not the discharge time for the chosen tank

capacitor with the prescribed parameters E_0 and V_0^2 needs to be shortened or lengthened in order to deliver to the load R_l the required nominal power in the working time T.

We shall now discuss the requirements that the dispenser must meet. The dispenser C_2 extracts energy from the tank C_1 in short pulses τ_i and low doses ΔE_{τ_i} with a reciprocal duty factor $\varkappa$ and frequency f. The operating frequency of the dispenser is determined by the time period τ during which a complete cycle of charging and discharging of the dispenser occurs ($f=1/\tau$). The reciprocal duty factor of the process is $\varkappa=\tau/\tau_i$. It is assumed that $\tau_i \leqslant \tau$, so that $\varkappa \geqslant 1$.

The energy required per pulse in order to deliver the required nominal power to the user is

$$\Delta E_{\tau_i} = W_n \tau. \tag{2}$$

When the dispenser is discharged across the useful load, the energy changes by a factor e in time $\tau_i=\tau^d/2$, where $\tau^d=R_lC_2$ is the intrinsic discharge time of the dispenser C_2. Therefore the dispenser capacitance C_2 can be expressed in terms of the parameters τ, R, and κ as

$$C_2 = 2\tau/(R_l\varkappa). \tag{3}$$

4. We shall now consider the operation of the proposed scheme as a whole. If the discharge time t_E of the capacitor C_1 is comparable to the working time T, i.e., we have a normal capacitor, then the capacitor C_1 will be discharged across a load according to an exponential law. In this case a dispenser is required for uniform delivery of electric energy to the load. This can be done using a dispenser operating with a reciprocal duty factor $\varkappa=1$, since for large values of $\varkappa$ power will be lost in this case. In the ''normal'' regime the capacitance C_2 of the dispensing capacitor is determined by the relation

$$C_2 = \frac{2\tau}{R_l}. \tag{4}$$

If $t_E \ll T$, then a dispenser is required in order to lengthen the discharge time of the capacitor C_1 to the working time T and to delivery electric energy to the load R_l uniformly. For this the dispenser must operate with a reciprocal duty factor $\varkappa=T/t_E \gg 1$. In this case, its capacitance must be $\varkappa$ times smaller than that of the dispenser in the normal regime (4).

Finally, if $t_E > T$, then the discharging of the capacitor C_1 must be compressed to the required working time T. However, a single dispenser used for this purpose does not provide the required power on the load. This difficulty can be overcome by multiplying the voltage by using a chain of n auxiliary capacitors connected in series.[3,4] Specifically, if we charge n auxiliary capacitors simultaneously from the capacitor C_1 using a parallel arrangement and then connect them in series into a chain, then we obtain a single working dispenser with plate voltage n times greater than the plate voltage of the capacitor C_1. The number of such auxiliary dispenser capacitors is $n=\sqrt{t_E/T}$. The entire chain of capacitors as a whole will operate as a single dispenser with a reciprocal duty factor $\varkappa=1$. The capacitance of such a single dispenser should be $C_2=2\tau/R_l$, just as in Eq. (4).

We note that if $\sqrt{\tau_E/T}$ is very large, then it is possible to develop and use a procedure for multiplying the voltage by a number of auxiliary dispensers much less than $n=\sqrt{t_E/T}$. For this, the collection of auxiliary dispensers must be separated by voltage into several groups of capacitors. Capacitors in the first group must be charged simultaneously from C_1. Then these capacitors are connected in series into a chain, and the entire chain charges each capacitor in the second group separately. In turn, the charged capacitors in the second group are connected into a new group and such a chain is used to charge the capacitors in the next group, and so on. This voltage multiplication procedure can be performed without energy loss, since dispensers with a short charging time ($\tau_i \leqslant 10^{-8}$ s) will be used. We note that the capacitance of a single dispenser, consisting of the last chain of capacitors, should always be determined by Eq. (4) and total voltage on such a dispenser should be $V_2=\sqrt{t_E/T}V_1$, where V_1 is the voltage on the plates of the tank capacitor C_1.

We call attention to another important aspect of the operation of the scheme considered here. As energy is extracted from the capacitor C_1, the voltage on its plates will decrease exponentially from V_0 to $\sim 0.6V_0$ in a time t_E. The maximum dispenser voltage will also decrease to this value. This means that the amount of energy ΔE_{τ_i} per pulse transferred by the dispenser into the load will also gradually decrease in time by a factor e to $0.37\Delta E_{\tau_i}$. All this occurs for all three regimes considered above. This phenomenon can be prevented by adjusting the discharge time of the tank capacitor using additional dispensers. For this, it is sufficient to decrease the discharge time of the tank capacitor by a factor of 2. In practice this can be accomplished with two dispensers connected in series into a chain. The capacitance of each of these dispensers must be two times greater than the capacitance of the dispenser C_2 used to correct the discharge time of the capacitor C_1. The total capacitance of the chain obtained should once again be determined by the relation (3) or (4), depending on the ratio of t_E and T, i.e., depending on the regime used.

We thank A. A. Rukhadze and A. N. Dovbne for encouragement and their interest in this work.

[1] G. I. Skanavi, *The Physics of Dielectrics* (*Weak Fields*) [in Russian], Gostekhizdat, Moscow, 1949.
[2] V. T. Renne, *Capacitors* [in Russian], Énergiya, Leningrad, 1969.
[3] A. R. Von Hippel, *Dielectric Materials and Their Applications* [Technology Press of M. I. T. and Wiley, New York, 1954; Gosénergoizdat, Moscow, 1959].
[4] D. A. But, B.L. Alievskiĭ, S. R. Mizyurin, and P. V. Vasyukevich, *Energy Storage Systems* [in Russian], Énergoatomizdat, Moscow, 1991.

Translated by M. E. Alferieff

BRIEF COMMUNICATIONS

Effect of magnetic field pulses on the anelastic properties of nitrogen-containing steel

O. I. Datsko, V. I. Alekseenko, and A. L. Brusova

A. A. Galkin Donetsk Physicotechnical Institute, Ukrainian Academy of Sciences, 340114 Donetsk, Ukraine
(Submitted July 23, 1998)
Zh. Tekh. Fiz. Zh. Tekh. Fiz. **69**, 122–123 (August 1999)

The effect of weak magnetic field pulses on the dynamic properties of dislocations interacting with impurity–defect complexes is determined by the internal-friction method. The effect is characterized by an increase in the plasticity of a material under the conditions of microplastic deformation and slowing of strain hardening. © *1999 American Institute of Physics.* [S1063-7842(99)02208-4]

According to Ref. 1, treatment with weak magnetic field pulses ($H<10^6$ A/m) changes the state of impurity–defect complexes (IDCs) on dislocations. Therefore it should be possible to affect the anelasticity and microplastic deformation (MPD) and strain hardening (SH) by applying magnetic field pulses (MFPs) to a material.

We used nitrogen-containing chromium–manganese Kh14G10AS (0.1% C, 14% Cr, 10% Mn, and $<1.5\%$ each for N and Si) austenitic steel. Two different processes leading to a change in the structure of the material can occur in such steels as a result of plastic deformation. One process is the $\gamma-\alpha$ martensitic transformation in austenite. Preliminary x-ray analysis showed that a $\gamma-\alpha$ martensitic transformation is not observed after MPD. Moreover, it was found that the compounds Fe_2N, Fe_3N, and Fe_4N are not present in the initial material after MPD and post-MPD relaxation. The absence of a martensitic transformation during MPD and the absence of the aforementioned compounds is a comprise a necessary condition for the appearance and study of the SH process, which occurs in the given material because the mobile nitrogen atoms can segregate on dislocations.

The experimental samples were $3\times3\times60$ mm rectangular prisms. They were obtained mechanically from rolled material.

The change in the state of the dislocation–IDC system was monitored by measuring the low-frequency internal friction (IF Q^{-1}) with frequency 1 Hz (inverted torsional pendulum[2]) in an amplitude-independent range with a strain amplitude of the material $\varepsilon=4\times10^{-5}$. The investigations showed that in the initial samples the IF does not change with time; this attests to kinetic stability of the dislocation–IDC system under normal conditions (no MPD and no MFP).

Microplastic deformation was conducted during measurements of the amplitude-dependent IF by straining the material in the amplitude range $\varepsilon=4\times10^{-5}-16\times10^{-5}$. To initiate SH the material was subjected to MPD (by twisting) to a relative deformation $\varepsilon=2\times10^{-3}$.

The experimental samples were treated with MFPs according to the following regime: The amplitude of the MFP intensity was $H=3\times10^5$ A/m, the pulse repetition frequency was 10 Hz, the duration of the leading edge of the pulse was 10^{-4} s, and the treatment times were 2 min (this is the minimum time for determining the statistical average IF) and 100 min. The basic assumptions of the method are described in Ref. 3.

The investigations showed that when a MFP is switched on, the IF increases abruptly, and when the MFP is switched off, the IF remains constant for 60 min and then decreases monotonically, approaching the initial value (Fig. 1, curve *1*). With the MFP switched on the IF remains at the level reached after increasing abruptly (curve *2*).

The observed behavior of the IF could be due to a change in the interaction of dislocations with IDCs as a result of impurity atoms being transferred into a state with a different value of the interaction potential, and it attests to the fact that MFPs can change the energy state of the dislocation–IDC system, plasticizing the experimental material.

The study of the effect of a MFP on MPD shows that a magnetic field increases the general level of the amplitude-dependent IF relative to the initial value of the IF (Fig. 2), thereby facilitating depinning of dislocations from stops. Therefore an MFP influences equally effectively a pinned dislocation (amplitude–independent IF), depinning it, and the very process that leads to depinning of dislocations from stops (amplitude-dependent IF).

The kinetics of the IF level during SH (with and without the application of a MFP), occurring after MPD is displayed in Fig. 1 (curves 3 and 4). As one can see, in these cases the level of the amplitude-independent ($\varepsilon=4\times10^{-5}$) dislocation IF decreases monotonically. Comparing these IF curve shows that SH with the application of an MFP is slower and terminates at higher values of the IF (curve *4*). This shows that less strongly pinned IDCs form on dislocations. In other words, the application of an MFP tends to decrease the degree of pinning of dislocations by impurities and increase the plasticity of the material.

The time dependences of the IF during SH of a material can be described in the Harper model[4], taking account the

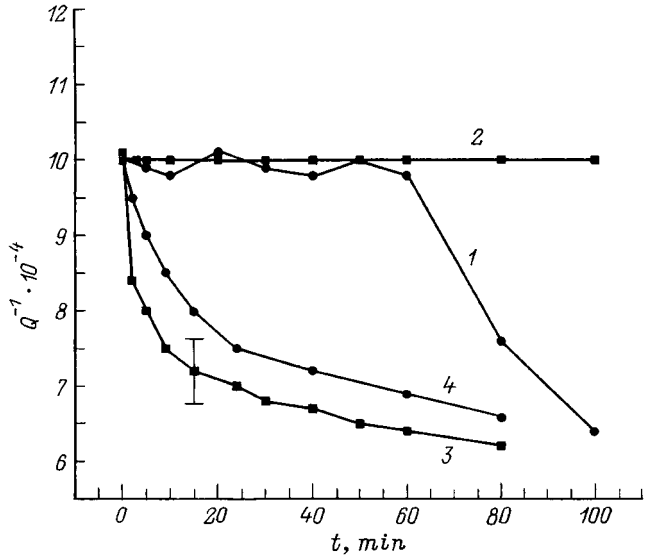

FIG. 1. Time dependence of the internal friction in Kh14G10AS steel: *1* — after the MFP is switched off, *2* — after the MFP is switched on, *3* — after MPD, *4* — after MPD in the presence of a MFP.

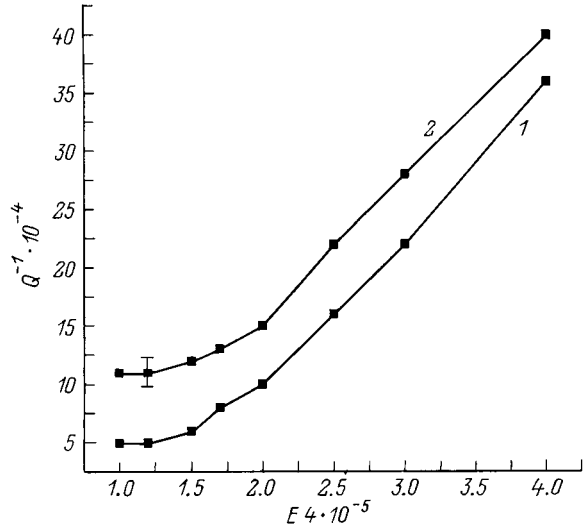

FIG. 2. Amplitude dependence of the internal friction in Kh14G10AS steel: *1* — no MFP, *2* — in the presence of a MFP.

relation $Q^{-1}\sim L_c^4$ (where L_c is the distance between pinning points), as

$$Q_t^{-1}=Q_\infty^{-1}+[Q_0^{-1}-Q_\infty^{-1}][\exp(-t/\tau)^{2/3}]^4,$$

where Q_0^{-1} and Q_∞^{-1} are the IF at the start and end of the process under study, t is the time, and τ is the relaxation time constant.

The analytic expression presented was used to estimate the relaxation time of SH under ordinary and MFP conditions: $\tau_1=70$ min in the first case and $\tau_2=110$ min in the second case.

In summary, it has been shown experimentally in this work that MFPs can effectively influence the dynamics of pinned linear defects and the processes leading to depinning and pinning of such defects. A material containing linear and point defects plasticizes as a result of these effects.

[1] O. I. Datsko, V. I. Alekseenko, and A. D. Shakhova, Fiz. Tverd. Tela (St. Petersburg) **38**, 1799 (1996) [Phys. Solid State **38**, 992 (1996)].
[2] V. S. Postnikov, *Internal Friction in Metals* [in Russian], Metallurgiya, Moscow, 1974, 324 pp.
[3] O. I. Datsko and V. I. Alekseenko, Fiz. Tverd. Tela (St. Petersburg) **39**, 1234 (1997) [Phys. Solid State **39**, 1094 (1997)].
[4] S. Harper, Phys. Rev. **83**, 709 (1951).

Translated by M. E. Alferieff

Nanotubes and force interactions in an atomic force microscope

G. V. Dedkov and S. Sh. Rekhviashvili

Kabardino-Balkar State University, 360004 Nalchik, Russia
(Submitted September 29, 1998)
Zh. Tekh. Fiz. **69**, 124–127 (August 1999)

The conditions for nanotubes to be used as atomic force microscope (AFM) probes are analyzed. It is shown theoretically for the first time that single- and multilayer tubes with diameters ranging from 0.5 to 5 nm give atomic-level resolution of the surface. The presence of cylindrical symmetry makes each surface atom of a nanotube ''imaging.'' For a definite ratio of the diameter of a single nanotube and the period of the surface structure, the atomic resolution vanishes. Such nanotubes are of special interest for probing the details of the large-scale relief and for investigations in nanotribology. In contrast to silicon and other (solid) probes, nanotubes are not blunted on contact with the surface, but rather they bend and their initial shape is restored when the stress is removed. The critical loads for an AFM to function in the repulsive regime are determined.
[S1063-7842(99)02308-9]

The problem of searching for new probes for AFMs remains pressing, because existing structures (based on silicon and other solid materials) do not meet adequately the increasing requirements of this important field of physical research. The main problems are due to the destruction or blunting of probes coming into hard contact with surfaces and to poor shape control. For this reason, there is no reliable relation between the measured forces and theoretical models.

One way to solve this problem is to use nanotubules or C_{60} fullerene molecules.[1,2] Experiments show that an AFM probe tip formed by a nanotube can provide atomic-level resolution.[3] Specifically, it has been noted that such structures possess a very high elastic modulus ($\sim$1–5 TPa) and resonance frequency ($>$200 kHz).

The present work is a continuation of Ref. 2. Our objective is to investigate theoretically the resolution of an AFM by calculating the interaction forces acting between single- and multilayer nanotubes (different lengths and diameters) and a solid surface. The minimum distance to which a nanotube approaches a surface is estimated on the basis of the calculations, and the operating regimes of an AFM are determined.

Two different models are used. In the first one the surface is treated as a semi-infinite medium with volume density n_2 and a single-layer nanotube is treated as a hollow cylinder with a uniform surface density n_1 of atoms, which is located at a distance h from the surface. The second model takes into account the specific atomic structure of the sample surface.

The nanotube is assumed to be attached to a cantilever, in the form of a rectangular plate with area S and thickness d (Fig. 1). This is the configuration employed in Ref. 3.

The interaction potential between a single atom of the nanotube and a surface atom is chosen as a combination of a repulsive potential at short distances that is obtained by approximating the computational results obtained with the electron-gas model[2] a Lennard–Jones potential at medium distances, and a retarded Casimir potential at large distances:

$$U(r)=\begin{cases}\dfrac{\beta}{r}\exp(-\alpha r), & 0\leqslant r\leqslant r_1,\\[2ex] -D\left(\dfrac{1}{r^6}-\dfrac{r_0^6}{2}\dfrac{1}{r^{12}}\right), & r_1\leqslant r\leqslant \lambda_0/2\pi,\\[2ex] -\dfrac{23\hbar c\alpha_1\alpha_2}{2\pi r^7}, & r>\lambda_0/2\pi,\end{cases} \tag{1}$$

where r is the distance between the atoms; $D=1.49\times10^{-5}$ eV·nm^6; $r_0=0.381$ nm (for carbon atoms in graphite layers);[4] λ_0 is the characteristic wavelength of the absorption spectrum; c and $\hbar$ are speed of light in vacuum and Planck's constant; $\alpha_{1,2}$ are the polarizabilities of a carbon atom and a surface atom; and, β, α, and r_1 are adjustable parameters.

The coefficient in the Casimir potential is exact for the interaction of identical atoms, while for different atoms the geometric-mean is used.

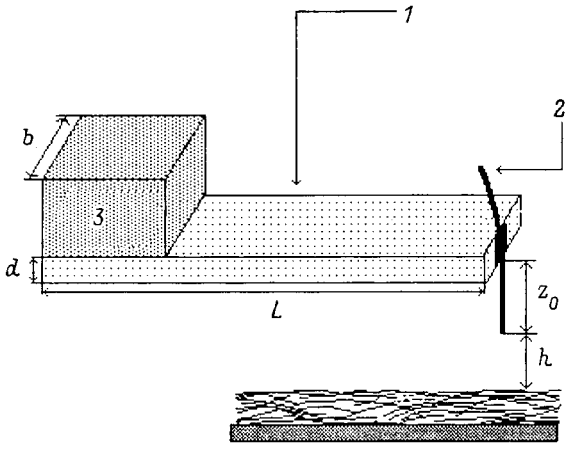

FIG. 1. Diagram of the cantilever of an AFM with a nanotube: *1* — cantilever, *2* — nanotube, *3* — piezoelectric drive.

TABLE I. Analytical expressions for the interaction forces in an AFM in the continuum approximation (model 1).

System	Interaction force
Nanotube–sample (van der Waals forces)	$F(h)=-\frac{AR}{3}\frac{z_0(3h^2+3hz_0+z_0^2)}{(h+z_0)^3h^3}$
Nanotube–sample (Casimir forces)	$F(h)=-\frac{23\hbar c\alpha_1\alpha_2 n_1 n_2\pi R}{10}\frac{z_0(4h^2+6z_0h^2+4hz_0^2+z_0^3)}{(h+z_0)^4h^4}$
Cantilever–sample (Casimir forces)	$F(x)=-\frac{23\hbar c\alpha_1\alpha_2 n_1 n_2 S}{20}\frac{d(4x^3+6dx^2+4xd^2+d^2)}{(x+d)^4x^4}$, $x=h+z_0$
Nanotube–sample (repulsive forces) $r_1\leqslant r\leqslant\lambda_0/2\pi$	$F(h)=\frac{ARr_0^6}{45}\frac{1}{h^9}$
Nanotube–sample (repulsive forces) $r<r_1$	$F(h)=\left(\frac{2\pi}{\alpha}\right)^2 n_1 n_2 R\beta\exp(-\alpha h)$

Note: In the formulas, z_0 and R are the nanotube length and radius, respectively, and $A=\pi^2 n_1 n_2 D$.

In standard AFM structures with solid probes, it is ordinarily assumed that the interaction of the cantilever with the sample is negligibly weak and is due to the repulsion or attraction of a small group of atoms at the probe tip. For an AFM operating with a nanotube, the tube can be fractions of a micron long, so that the contribution of the upper part of the tube and cantilever, which are located within the range of the retarded Casimir forces, to the force could be substantial.

Using the continuum approximation (model 1) and the potential (1) it is easy to find all components of the force interactions of an AFM with the surface. The resulting formulas are presented in Table I.

To estimate the effect of the cantilever and the top part of the nanotube, calculations of the resulting attractive force were performed for the system silicon cantilever–nanotube–graphite surface as a function of the gap width h. The results are shown in Fig. 2 and correspond to the following geometric dimensions: $z_0=d=1$ μm, $S=30\times100$ μm, and $R=5$ nm. Curve *1* was obtained taking into account the Casimir forces arising between the sample, the cantilever, and the top part of the nanotube; curve *2* was obtained neglecting these forces and neglecting the contribution of the cantilever. It follows from Fig. 2 that the contributions of the force interactions with the cantilever and the top part of the nanotube must be taken into account even at distances $h>1.5$ nm. As the length of the nanotube decreases, this distance becomes even smaller.

It is of interest to estimate the critical distance to which a nanotube approaches the surface. In the contact regime, stable operation of an AFM is possible if the interaction force does not exceed the Euler critical value[5]

$$F_E=\frac{\pi^2 EJ}{\mu z_0^2}, \tag{2}$$

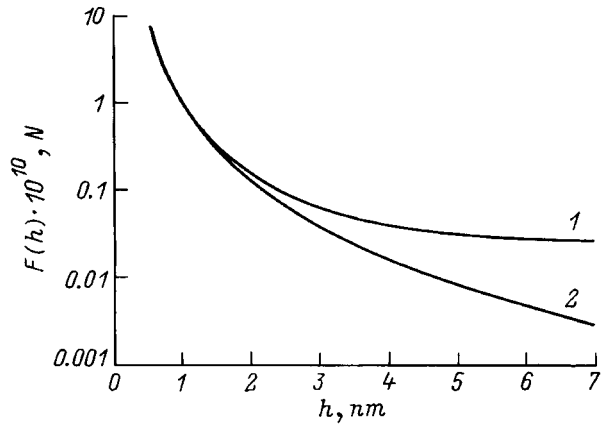

FIG. 2. Interaction force versus the distance: *1* — with the Casimir forces and the contribution of the cantilever taken into account; *2* — calculation in the van der Waals approximation, neglecting the contribution of the cantilever.

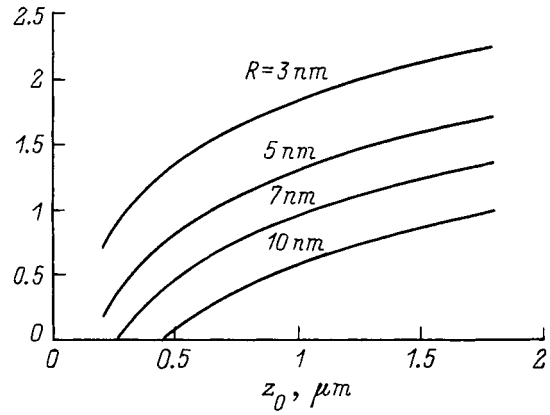

FIG. 3. Critical distance versus the nanotube length and radius.

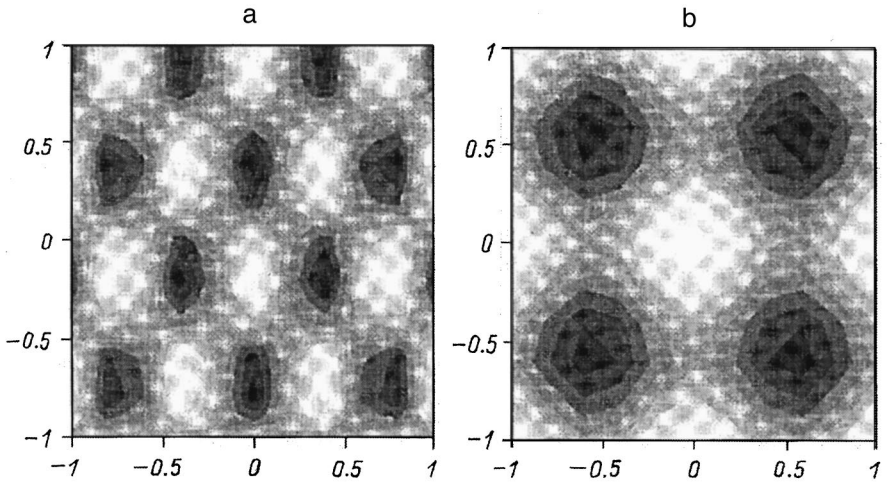

FIG. 4. Surfaces obtained with a single-layer nanotube ($R=5$ nm): a — graphite, force differential $\Delta F=6.382\times10^{-9}$ N; b — arbitrary square lattice of carbon atoms with period 0.246 nm, $\Delta F=5.177\times10^{-9}$ N.

where E is the elastic modulus; J is the moment of inertia of the transverse cross section (for a circular section $J=\pi R^4/4$); and, $\mu=0.7$ in the mounting shown in Fig. 1.

Using Eqs. (1) and (2) we obtain for the corresponding critical gap width

$$h_{\min}=\begin{cases}\sqrt[9]{\dfrac{ARr_0^6\mu z_0^2}{45\pi^2 EJ}}, & h_{\min}\geqslant r_1,\\ \dfrac{1}{\alpha}\ln\left(\dfrac{n_1n_2\beta R\mu z_0^2}{\alpha^2 EJ}\right), & h_{\min}\leqslant r_1.\end{cases} \tag{3}$$

It is evident from Eq. (3) that the critical distance for probing a surface with a nanotube is determined by the dimensions and the physical characteristics of the tube. The dependence of $h_{\min}$ on the nanotube radius and length is presented in Fig. 3.

To study the resolving power of nanotubes as AFM probes computer simulation of the images of several surfaces was performed using model 2, taking into account the atomic structure of the surface. The calculations employed the nanotube–surface atom potential

$$U_t(h,\rho)=2Rn_1\times\int_0^{\pi}\int_0^{z_0}U\left(\sqrt{(h+z)^2+\rho^2+R^2-2\rho R\cos\varphi}\right)dz\,d\varphi, \tag{4}$$

where ρ is the distance from the point of projection of the nanotube axis to the surface atom and $U(r)$ is the pair potential (1).

The interaction with all surface atoms was found using Eq. (4) by direct summation. A procedure for smoothing the numerical data using the direct and inverse Fourier transforms was used to improve image contrast.

Images of an arbitrary square lattice of carbon atoms (with edge length $d=0.246$ nm) and a graphite surface, which were obtained by a single nanotube with a radius of 5 nm, are shown in Figs. 4a and 4b. It was assumed that $h=0.3$ nm.

Figures 5a and 5b show images of the same surfaces witha three-layer nanotube with layer radii 1, 1.35, and 1.7 nm. It is evident from the images presented that single- and

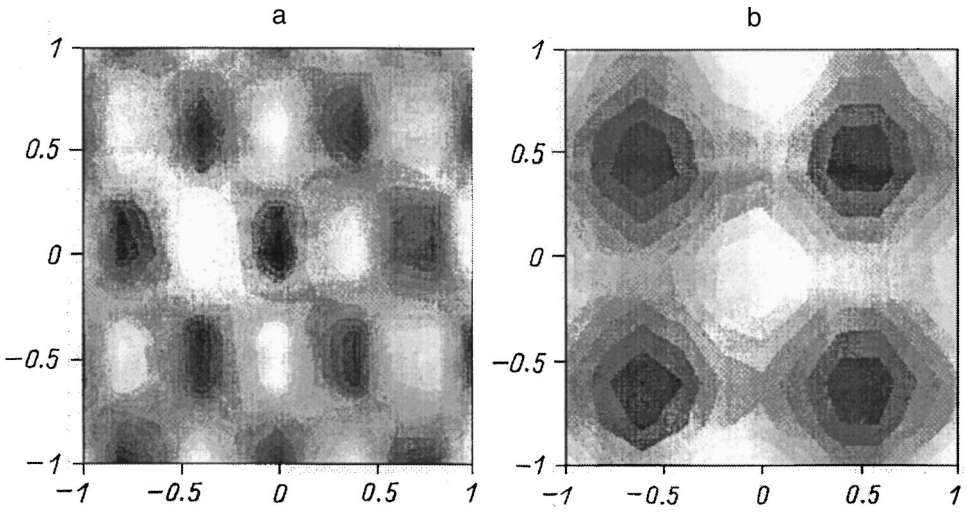

FIG. 5. Surfaces obtained with a multilayer nanotube: a — graphite, $\Delta F=4.418\times10^{-9}$ N; b — arbitrary square lattice of carbon atoms with period 0.246 nm, $\Delta F=2.327\times10^{-9}$ N.

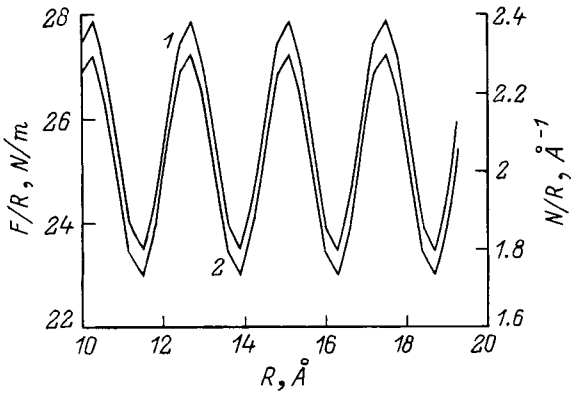

FIG. 6. For explanation of image inversion for a square lattice (period 0.246 nm): *1* — F/R; *2* — N/R.

multilayer nanotubes give a quite high contrast.

In the simulation process, it was found that for certain combinations of the nanotube radii and periods of the surface structure the image contrast is inverted. This geometric effect is due to a change in the effective number of atoms in the surface structure that fall within the ''field of view'' of the carbon atoms localized on the nanotube surface. This effect is illustrated in Fig. 6, which shows the vertical force applied to a nanotube with a fixed point of projection of the axis on the graphite surface as a function of the radius R. Since Eq. (4) gives a directly proportional increase of the force with increasing R, the ratio F/R, obtained after Fourier smoothing (curve *1*), is plotted along the ordinate. Curve *2* shows N/R, where N is the number of surface atoms in the ''zone of visibility'' of the nanotube. The corresponding zone was ring-shaped with radii $R \pm \Delta R$ ($\Delta R = 0.1$ nm). Figure 6 shows that the periodic variation of the force contrast is due to oscillations of the ''effective number'' of atoms probed.

For certain values of the nanotube radius and lattice period, the effective number of interacting atoms can change negligible at each step as the image is formed. In this case the atomic resolution vanishes. Such a combination of tube size and lattice period is of special interest for studying the mechanism of friction interactions in nanotribocontacts formed by a nanotube and a surface. It is also important to use nanotubes for these purposes because the radius of the contact zone in this case is clearly determined, while for ordinary contacts it depends on the clamping force.

[1] David L. Wilson, Pranov Dalal, and Kenneth S. Kump, J. Vac. Sci. Technol. A **14**, 2407 (1996).
[2] G. V. Dedkov, Pis'ma Zh. Tekh. Fiz. **23**(12), 37 (1997) [Tech. Phys. Lett. **23**, 469 (1997)].
[3] Dai Hongjie, H. Hafner Jason, and G. Rinzler Andrew, Nature (London) **384**, 147 (1996).
[4] L. A. Girifalco and R. A. Lad, J. Chem. Phys. **25**, 693 (1956).
[5] L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* [Pergamon Press, New York, 1986; Nauka, Moscow, 1987].

Translated by M. E. Alferieff

Critical conditions for spontaneous relaxation in self-organizing dissipative film systems

V. P. Khan, V. Ya. Kogaĭ, and E. V. Aleksandrovich

Institute of Applied Mechanics, Urals Branch of the Russian Academy of Sciences, 426001 Izhevsk, Russia
(Submitted November 27, 1998)
Zh. Tekh. Fiz. **69**, 128–129 (August 1999)

It is established that the necessary conditions for spontaneous relaxation of elastic strain energy in a copper–$As_{60}Se_{40}$ self-organizing dissipative heterostructure is that the elastic deformation energy and the temperature must reach their threshold values. It is shown that in the temperature range 270–340 K the spontaneous relaxation of elastic deformation energy is accompanied by structural–chemical ordering and anomalous diffusion of copper into the glassy chalcogenide semiconductor layer. The maximum concentration of copper dissolved in the films is 40 at. %. Conductivity inversion from *p* to *n* type is observed in doped layers obtained by this method. © *1999 American Institute of Physics.* [S1063-7842(99)02408-3]

A great deal of attention has been devoted to relaxation processes in oxide and metallic glasses.[1,2] However, these processes have been little studied in glassy chalcogenide film structures.[3]

In the present work we investigated the spontaneous relaxation of elastic deformation energy in a copper–$As_{60}Se_{40}$. self-organizing thin-film heterostructure.

A distinguishing feature of films of glassy chalcogenide semiconductors (GCSs), as compared with their crystalline analogs, is that they contain a high concentration of structural and chemical defects. This promotes the accumulation of high elastic deformation energies in films.

Kinetic curves of the variation of the surface resistivity of Cu–$As_{60}Se_{40}$ heterolayers with different thickness ratios are displayed in Fig. 1. We produced a Cu–GCS heterostructure at temperature 293 K. It follows from the figure that the maximum rates of diffusion of copper into the GCS layers are attained with a GCS–copper thickness ratio of 8, for which the elastic deformation energy density in the GCS film corresponds to the threshold value. In the process, the dissipative heterostructure Cu–GCS transforms into a new ordered state. The internal parameter controlling this transition is the elastic deformation energy density.

Figure 2 shows the kinetic curves of the surface resistivity of Cu–GCS heterolayers with a GCS–copper thickness ratio of 8 at various temperatures. As the temperature increases, the rate of copper diffusion into the GCS layer increases, and at $T=340$ K the rate changes abruptly. The maximum concentration of copper dissolved in GCS films is 40 at. %. Structural-chemical ordering occurs during the process of spontaneous relaxation of the elastic deformation energy. The copper diffusing in the form of a positively charged Cu^+ ions into the GCS layer becomes uniformly distributed in the layer and interacts chemically with the components of the glass, producing glass-forming structural units consisting of the unstable chemical compound $CuAsSe_2$. These structural units and the structural units $As_2Se_{3/2}$ and $As_2Se_{4/2}$ (Ref. 4) join together into polymer formations, where arsenic atoms are copolymerizers. This expands the region of glass formation. As the copper concentration in the films increases further, molecular fragments with intercluster boundaries appear. The granular structure of amorphous As_2Se_3 films has been observed in Ref. 5. When the copper concentration exceeds 30 at. %, the molecular clusters of the unstable compound $CuAsSe_2$ decompose according to the reaction

$$3\ CuAsSe_2 = Cu_3AsSe_4 + 2\ AsSe, \tag{1}$$

forming the stable crystal compound Cu_3AsSe_4. All chemical ordering processes in the system occur according to the law of mass action.

Since at temperatures below the critical temperature, which is 270 K (Fig. 3), the acoustic modes are ''frozen'' and the process of spontaneous relaxation of the elastic deformation energy is not observed. Hence it follows that spontaneous relaxation of elastic deformation energy and diffusion of copper into the GCS layer are possible only at temperatures above the critical temperature and when the elastic deformation energy density reaches its threshold value.

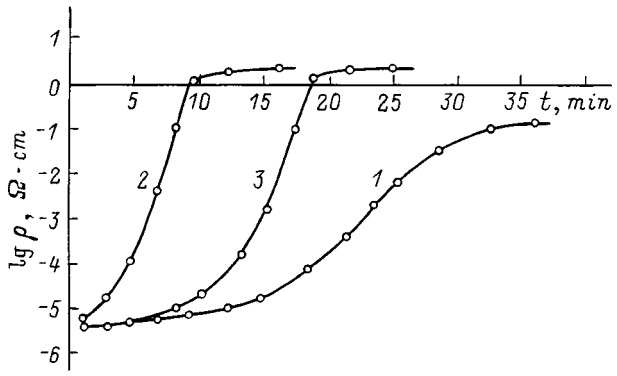

FIG. 1. Kinetic curves of the surface resistivity of Cu–GCS heterolayers at $T=293$ K for different thickness ratios: $d_{GCS}/d_{Cu}=5$ (*1*), 8 (*2*), and 10 (*3*).

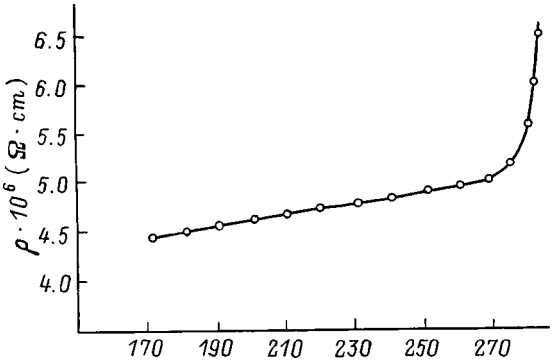

FIG. 2. Kinetic curves of the surface resistivity of heterolayers with $d_{GCS}/d_{Cu}=8$ for temperatures 310 (*1*) and 340 K (*2*).

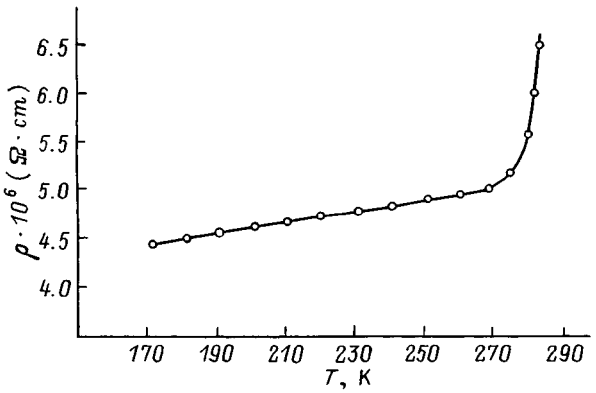

FIG. 3. Surface resistivity of Cu–GCS heterolayers versus the temperature for $d_{GCS}/d_{Cu}=8$.

Inversion of conductivity from *p* to *n* type was observed in doped layers obtained by this method. It is due to the presence of zero-valence quasiatoms Cu^0 with weakly bound valence electrons in the system. These atoms appear as a result of a disproportionation reaction of Cu^+ ions according to the scheme

$$Cu^+ \rightarrow Cu^{2+} + Cu^0. \quad (2)$$

To check the existence of heterovalent copper ions Cu^{2+}, Cu^+, and Cu^0 in the system, we investigated the x-ray-electron spectra of the internal lines $Cu2p_{3/2}$, $As3d_{5/2}$, and $Se3d_{5/2}$ and the Auger line Cu(*LMM*), which were obtained on the ÉS-2401 electronic spectrometer. The shape and position of the copper spectral lines (E_t $Cu2p_{3/2}$ $=933.0$ eV, $E_{kin}Cu(LMM)=917.4$ eV) indicated that Cu^0 and Cu^+ are present in the system, while the satellite at a shift of 8.5 eV from the main peak attested to the presence of Cu^{2+}. The electrophysical properties of the doped layers will be studied in further investigations.

We thank O. M. Kanunnikova for performing the x-ray electron analysis of the samples.

[1] G. M. Bartenev and D. S. Sanditov, *Relaxation Processes in Glassy Systems* [in Russian] (Nauka, Novosibirsk, 1986), 256 pp.
[2] O. V. Mazurin, *Vitrification* [in Russian], Nauka, Leningrad, 1986, 158 pp.
[3] M. L. Trunov, A. G. Anchugin, A. A. Tarnaĭ *et al.*, in *Materials and Systems for Recording Holograms* (Physicotechnical Institute, Leningrad, 1986), 24 pp.
[4] Z. U. Borisova, *Chalcogenide Semiconductor Glasses* [in Russian], Leningrad State University Press, Leningrad, 1983, 344 pp.
[5] J. Phillips, in *Physics Abroad* (Mir, Moscow, 1983), 178 pp.

Translated by M. E. Alferieff

Requirements on dielectrics for a capacitor-based autonomous source of electric power

D. I. Adeĭshvili and N. F. Shul'ga

Kharkov Physicotechnical Institute National Science Center, 310108 Kharkov, Ukraine

V. P. Kortkhondzhiya

Institute of Physics, Georgian Academy of Sciences, 380077 Tbilisi, Georgia
(Submitted July 30, 1998)
Zh. Tekh. Fiz. **69**, 130–132 (August 1999)

The requirements on dielectrics for solving the problem of accumulating, storing, and regulating the extraction of electric energy by means of electric capacitors with the required nominal power over a long time are examined and established. © *1999 American Institute of Physics.* [S1063-7842(99)02508-8]

1. In Ref. 1 we called attention to the possibility of using capacitors as autonomous sources of electric power. To solve this problem a method was proposed for controlled extraction of energy stored in a capacitor that makes it possible to release a prescribed nominal power in a load over a long period of time. The electric energy accumulated in a capacitor depends strongly on the dielectric used in the capacitor.[2–5]

In the present paper we examine the requirements on dielectrics to solve the problems of the accumulation, storage, and controlled extraction of electric energy in a long period of time with the required nominal power using capacitors.

2. The insulating medium for the purpose indicated must satisfy the following two requirements above all. First, it must possess a high specific volume energy capacity. This will make it possible to produce a compact source of electric energy. Second, it must possess a high volume resistivity, i.e., it must be a good dielectric. Since the leakage currents are small, this will make it possible to preserve for a long time the energy stored in a capacitor. The possibility of large dissipative power losses in the dielectric during the discharging of the capacitor on a load is also of concern here. We shall discuss these requirements in greater detail.

3. A capacitor C_b making it possible to accumulate the maximum energy $E_0=1.6W_HT$ is required to supply nominal power W_H to a user for a time T.[1] The specific volume electric capacity of a capacitor is given by the formula $\rho_\nu=\varepsilon_0\varepsilon_r U_0^2/2$ (ε_0 is the permittivity of empty space, ε_r is the relative permittivity, and U_0 is the breakdown field strength), where $U_0=V_0/d$ (V_0 is the breakdown voltage and d is the thickness of the dielectric layer between the capacitor plates). The smallest volume ν of the dielectric layer between the capacitor plates that is required for this case is determined from the expression $E_0=\rho_\nu\nu$.[3,6–8]

As insurance against breakdown of the dielectric, we shall take as the maximum voltage on the capacitor plates half the breakdown voltage $V_0/2$. In this case the minimum spatial volume of the dielectric in the capacitor will be

$$\nu=\frac{12.8QW_HT}{\varepsilon_0\varepsilon_r U_0^2}. \tag{1}$$

4. We shall now consider the energy losses in the dielectrics used in storage capacitors. Energy losses due to volume leakage currents and discharge currents on a load in the external electric discharge circuit of the capacitor occur in the dielectric in a capacitor. In the first part of the electric arrangement[1] in the tank capacitor circuit the energy losses are due only to leakage currents. There are no other losses here, since the dispenser is charged from the tank without any energy losses. This is because dielectrics with rapid polarization ($10^{-14}-10^{-8}$ s) are used in both capacitors. In the second part of the arrangement[1], as the dispenser is discharged on a load, dissipative power losses occur in the dispenser because of powerful discharge currents in the load. There are virtually no other energy losses here. Therefore the total energy losses in the method proposed in Ref. 1 for controlled extraction of the energy stored in a capacitor consist of the energy losses due to the leakage currents of the tank capacitor in the dispenser charging circuit and the energy losses due to dissipative power losses in the discharge circuit of the dispenser.

The energy losses due to volume leakage currents are characterized by the time constant $\tau_c=\varepsilon_0\varepsilon_r\rho$ of the capacitor, and they are determined only by the electric parameters of the dielectric (ρ — volume resistivity). The ratio T/τ_c is proportional to the relative energy losses due to volume leakage currents of the tank capacitor. However, since the factor of e energy discharge time of the capacitor is half the intrinsic discharge time of the capacitor, $\tau_c/2$ should be used instead of τ_c. Then the relative energy losses due to volume leakage currents of the tank capacitor will be determined by the relation

$$P_y=\frac{2T}{\tau_c}. \tag{2}$$

Let us now consider the dissipative power losses in the dispenser C_d. We note that the dissipative power losses W_{nd} in the dispenser are compensated by corresponding power

losses W_n in the tank capacitor. The relation between W_n and W_{nd} can be determined as follows. The dissipative specific volume dielectric power losses are determined by the relation[2] $P_n = U_0^2/(2\rho)$. Using the formulas $C_b = \varepsilon_0 \varepsilon_r / d^2$, $E_0 = 1.6\Delta E$ ($\Delta E = W_H T$), and $E_0 = C_b V_0^2/2$, we can determine the total dielectric dissipative power losses of the tank capacitor as $W_n = P_n \nu$ or $W_n = 1.6\Delta E/\tau_c$. We note that for a dispenser, similarly, $W_{nd} = 1.6\Delta E \tau_i / \tau_c$, where $\Delta E \tau_i = W_H \tau$ is the amount of energy transferred by the dispenser to the load in one pulse. The nominal power is determined by the formula $W_H = \Delta E \tau_i / \tau$, so that $\Delta E = N \Delta E \tau_i$ and therefore $W_n = N W_{nd}$.

Since the nominal power is $W_H = \Delta E/T$, comparing W_n and W_H we obtain, using Eq. (2), the relative dissipative power losses in our method as

$$P_0 = 0.8 P_y. \tag{3}$$

The relation (3) shows that in the present method the relative dissipative power losses can be expressed in terms of the relative energy losses due to leakage currents. Therefore the total relative energy losses in our method will be $P = P_y + P_d$, i.e.,

$$P = 3.6 \frac{T}{\tau_c}. \tag{4}$$

The expression (4) shows that the dielectrics in the capacitors C_b and C_∂ will always be in states that are not overloaded, so that there is no need to fear loss of the dielectric properties of the dielectrics and other problems arising when dielectrics overheat.

5. The results obtained make it possible to formulate the requirements on dielectrics. To this end we introduce the parameters (ε, ρ) and $(\varepsilon_r U_0^2)$, which are convenient for describing the electric characteristics of a dielectric in the problem at hand.

Since $\tau_c = \varepsilon_0 \varepsilon_r \rho$, we find, taking Eq. (4) into account, that $(\varepsilon_r \rho) = 3.6T/(\varepsilon_0 P) = 4 \times 10^{11} T/P$ (we have used here the value $\varepsilon_0 = 8.854 \times 10^{-12}$ F/m). At present, relative losses $P \leqslant 10^{-2}$ are considered to be low, and they are very good for wide practical applications. In this case $\varepsilon_r \rho = 4 \times 10^{13} T$. This relation shows the values of the parameters ε_r and ρ required to preserve the accumulated energy for the required working time with losses not exceeding $P \leqslant 10^{-2}$. For convenience we call the product $\varepsilon_r \rho$ the ''figure of merit'' of the dielectric. Its dimensions in the SI system are $\Omega \cdot$m. For example, if the user requires nominal power for working time $T = 30$ h (10^5 s) with losses not exceeding $P \leqslant 10^{-2}$ (1%), then a dielectric with figure of merit $\varepsilon_r \rho = 4 \times 10^{13}$ $\Omega \cdot$m is required.

To clarify the specific volume energy capacity of capacitors we shall use Eq. (1). Substituting into it the numerical value of ε_0 we obtain $\varepsilon_r U_0^2 = 4 \times 10^{11} W_H T/\nu$.

This relation shows the values of the parameters ε_r and U_0^2 required to accumulate for a given dielectric volume the required amount of electric energy. For convenience we call the parameter $\varepsilon_r U_0^2$ the ''measure of strength'' (with respect to volume). Its dimensions in the SI system are $(\text{V/m})^2$.

Therefore the dielectric must satisfy the requirements

$$(\varepsilon_r U_0^2) = 4 \times 10^{11} \frac{W_H T}{\nu}, \tag{5}$$

$$(\varepsilon_r \rho) = 4 \times 10^{11} \frac{T}{P}. \tag{6}$$

6. We shall now consider a specific typical situation where the user requires electric energy with nominal power $W_H = 100$ kW in time $T = 30$ h, and the user admits a power source with volume $\nu \leqslant 10^{-1}$ m^3 and energy losses $P \leqslant 10^{-2}$. This requires a tank capacitor with capacitance C_b that is capable of storing the maximum energy $E_0 = 4.8 \times 10^3$ kWh. Then, according to Eqs. (5) and (6), the dielectric in the tank capacitor must satisfy

$$(\varepsilon_r U_0^2) \geqslant 4 \times 10^{22} (\text{V/m})^2, \tag{7}$$

$$(\varepsilon_r \rho) \geqslant 4 \times 10^{18} \, \Omega \cdot \text{m}. \tag{8}$$

These requirements are quite stringent. Considering the progress made in insulator technology,[11] there is hope that dielectrics satisfying the requirements (7) and (8) will be produced in the near future. For other cases where the volume is not very critical, existing dielectrics[9–11] which possess two orders of magnitude lower values of the measure of strength (7) and figure of merit (8) could be quite suitable for this purpose.

In closing, we express our sincere appreciation to A. A. Rukhadze and A. N. Dovbne for their active support and their interest in this work.

[1] D. I. Adeĭshvili, V. P. Kortkhondzhiya, and N. H. Shul'ga, Zh. Tekh. Fiz. **69**(8), 118 (1999) [Tech. Phys. **44**, 977 (1999)].
[2] G. I. Skanavi, *The Physics of Dielectrics (Weak Fields)* [in Russian], Gostekhizdat, Moscow, 1949.
[3] G. I. Skanavi, *The Physics of Dielectrics (Strong Fields)* [in Russian], Fizmatgiz, Moscow, 1958.
[4] B. M. Tareev, *The Physics of Dielectric Materials* [in Russian], Énergoatomizdat, Moscow, 1982.
[5] V. T. Renne, *Capacitors* [in Russian], Énergiya, Leningrad, 1969.
[6] W. Franz, in *Handbuch der Physik*, Vol. 17, edited by S. Flüge [Springer-Verlag, Berlin (1956), p. 155; Inostr. Lit., Moscow, 1961, 208 pp.].
[7] A. A. Vorob'ev and E. K. Zavadovskaya, *Electric Strength of Solid Dielectrics* [in Russian], GTTL, Moscow, 1956.
[8] A. A. Vorob'ev and G. A. Vorob'ev, *Electric Strength and Fracture of Dielectrics* [in Russian], Vysshaya Shkola, Moscow, 1966.
[9] *Reference Data on Physical Quantities*, edited by I. K. Kikoin (Atomizdat, Moscow, 1976).
[10] *Handbook of Physical Quantities* [in Russian], edited by I. S. Grigor'ev and E. Z. Meĭlikhov (Énergoatomizdat, Moscow, 1991).
[11] *Handbook of Electrotechnical Materials* [in Russian], edited by Yu. V. Koritskiĭ, V. V. Pasynkov, and B. M. Tareev, Vols. 1 and 2 (Energoatomizdat, Moscow, 1986, 1987); Vol.3 (Énergoatomizdat, Moscow, 1988).

Translated by M. E. Alferieff

A double-gap planar magnetic lens

V. A. Zhukov and E. N. Kotlikov

St. Petersburg State University of Aerocosmic Instrument Building, 190000 St. Petersburg, Russia

V. D. Gelever

Institute of Analytical Instrument Building, Russian Academy of Sciences, 198103 St. Petersburg, Russia
(Submitted July 2, 1998)
Zh. Tekh. Fiz. **69**, 133–134 (August 1999)

An analytical model and equations for describing the field of a planar magnetic lens with two concentric nonmagnetic gaps are developed. It is demonstrated that the ''half-width'' of the magnetic field of the lens can be controlled by varying independently the currents in the excitation coils of the two gaps. This makes it possible to control the focal length and the spherical aberration of the lens. It is shown that when oppositely directed currents are used in the excitation coils, a planar magnetic lens that does not rotate the image can be obtained. © *1999 American Institute of Physics.* [S1063-7842(99)02608-2]

INTRODUCTION

Planar magnetic lenses with a single ring-shaped nonmagnetic gap are being increasingly used in standard and scanning electron microscopy as electronic objective lenses.[1–3] Because of their geometric characteristics these lenses make it possible easily to place the experimental object in the region of maximum magnetic field. Such an arrangement of the object decreases the spherical aberration and makes it possible to place a secondary-electron detector outside the nonmagnetic gap of the lens. To speed up and simplify the analysis of the electron-optic characteristics of such lenses we have proposed[3] a simple analytical model of the magnetic field of a planar magnetic lens, taking into account the main geometric features of the lens, and we have checked this model experimentally. The on-axis magnetic induction $B_z(0,0,z)$ in our model is described by the formula

$$B_z(0,0,z)=\mu_0 NI((\rho_1^2+z^2)^{-1/2}-(\rho_2^2+z^2)^{-1/2}) \times(\ln(\rho_1/\rho_2))^{-1}, \quad (1)$$

where $\mu_0=4\pi\times10^{-7}$ H/m is the magnetic constant in the SI system, ρ_1 and ρ_2 are, respectively, the inner and outer radii of the ring-shaped nonmagnetic gap, Z is the coordinate along the axis of the electron-optic system, and NI is the number of ampere-turns in the excitation coil. The ''half-width'' a of the axial magnetic induction distribution $B_z(0,0,z)$, described by Eq. (1), in the field of the lens was found to be strictly related with the outer ρ_2 and inner ρ_1 radii of the nonmagnetic gap of a planar lens as

$$a=\sqrt{2}\rho_1\rho_2(\rho_1^2+\rho_1\rho_2+\rho_2^2)^{-1/2}. \quad (2)$$

However, for certain applications, for example, the investigation of VLSI integrated circuits, it would be helpful to have a magnetic objective lens such that the half-width of the on-axis magnetic induction distribution $B_z(0,0,z)$ is a smooth function of the shape of the object. This can be accomplished with a planar magnetic lens possessing not one but two ring-shaped nonmagnetic gaps. In such a lens, each nonmagnetic gap will have its own excitation coil in which the current can be varied independently.

Our aim in the present work is to extend the analytical model developed for planar magnetic lenses in Ref. 3 to lenses with two nonmagnetic gaps.

STATEMENT OF THE PROBLEM

Just as in Ref. 3, we shall assume that the magnetic scalar potential in the model adopted for the magnetic lens is given on the plane containing the flat edge of the central pole tip and the edges of two outer pole tips concentric with the central one. Once again we shall assume, following Glaser,[4] that the magnetic field inside the narrow nonmagnetic gaps is orthogonal to the inner surfaces of the gaps. The magnetic scalar potential $\Phi_k(\rho)$, corresponding to such a field, inside the nonmagnetic gaps will satisfy the logarithmic law

$$\Phi_k(\rho)=A_k\ln(\rho)+B_k, \quad (3)$$

where $\rho=(x^2+y^2)^{1/2}$ is the distance from the axis of the lens to the point of interest; $K=1,2$ is the number of the nonmagnetic gap; and, A_k and B_k are constants, which according to Glaser[4] are determined from the boundary conditions on the neighboring pole tips.

We shall assume the potential within the flat boundaries of the magnetic poles to be constant, the constants for neighboring poles differing by the amount $\mu_0(NI)_k$, where $(NI)_k$ is the number of ampere-turns in the excitation coil, corresponding to the given nonmagnetic gap with number K, where $k=1,2$. Finally, the magnetic scalar potential on the entire boundary plane is

$$\Phi(\rho)=\sum_{k=1}^{2} C_k(\Theta(\rho)-\Theta(\rho-\rho_{2k-1})) + \frac{\ln(\rho/\rho_{2k})}{\ln(\rho_{2k-1}/\rho_{2k})}(\Theta(\rho-\rho_{2k-1})-\Theta(\rho-\rho_{2k})), \tag{4}$$

where $C_k=\mu_0(NI)_k$ and $\Theta(\rho)$ is the Heaviside θ function, defined as

$$\Theta(\rho-\rho_i)=\begin{cases}0 & \rho<\rho_i,\\ 1 & \rho\geqslant\rho_i.\end{cases} \tag{5}$$

Therefore we have a Dirichlet problem for the Laplace equation in the half space $z\geqslant 0$ with a boundary condition of the form (4) on the plane $z=0$.

METHOD OF SOLUTION

Using the Green's function for the Dirichlet problem for a half space,[5] we obtain the following expression for the on-axis distribution of the magnetic scalar potential for $Z\geqslant 0$:

$$\Phi(x,y,z)|_{x=y=0}=\Phi(0,0,z)=z\int_0^\infty \frac{\Phi(\rho)\rho d\rho}{(\rho^2+z^2)^{3/2}}, \tag{6}$$

where the potential $\Phi(\rho)$ is given by Eq. (4).

Using Eq. (6) for the magnetic scalar potential on the axis of the system we obtain for the on-axis magnetic induction distribution

$$B_z(0,0,z)=-\frac{\partial\Phi}{\partial z}(0,0,z)=-\mu_0\sum_{k=1}^{2} D_k((\rho_{2k-1}^2+z^2)^{-1/2} -(\rho_{2k}^2+z^2)^{-1/2})^{-1}. \tag{7}$$

Here

$$2_k=(NI)_k(\ln(\rho_{2k-1}/\rho_{2k})).$$

Continuing the field in the form (7) symmetrically to negative values of z, as in Ref. 3, gives a bell-shaped field, which, according to Ref. 4, can be approximated as

$$B_z(0,0,z)=B_0/(1+(z/a)^2),$$

where $B_0=B_z(0,0,0)$ is the maximum field on the z axis at $z=0$, and the ''half-width'' of the field, according to Ref. 4, is given by

$$a=(-2B_0/B'')^{1/2}, \tag{8}$$

where

$$B''\equiv\frac{\partial^2}{\partial z^2}B_z(0,0,z)$$

at $z=0$. Substituting into Eq. (8) the values of B_0 and B'' calculated using the field (7), we obtain for the half-width a

$$a=\left(-2\sum_{k=1}^{2} D_k(\rho_{2k-1}^{-1}-\rho_{2k}^{-1})/\sum_{k=1}^{2} D_k(\rho_{2k-1}^{-3}-\rho_{2k}^{-3})\right)^{1/2}. \tag{9}$$

Let us now examine the behavior of the half-width a of the field for different ratios between the currents in the first and second excitation coils. Let both nonmagnetic gaps be narrow, i.e., $\rho_{2k}=\rho_k^*+\Delta$, $\rho_{2k-1}=\rho_k^*-\Delta$, and $\rho_{2k}-\rho_{2k-1}=2\Delta\ll\rho_k^*$, where ρ_k^* is the average radius of the kth ring-shaped gap. Since the coefficients $D_k\sim(NI)_k$, setting $D_1\gg D_2$ in Eq. (9) we obtain $a_{\min}\approx\rho_1^*\sqrt{2/3}$, and setting $D_2\gg D_1$ we obtain, correspondingly, $a_{\max}\approx\rho_2^*\sqrt{2/3}$.

When the excitation currents in the two coils are approximately the same, intermediate values of the half-width a are obtained. Thus, the half-width of the field can be regulated continuously in the range $\rho_1^*\sqrt{2/3}\leqslant a\leqslant\rho_2^*\sqrt{2/3}$ by varying the relative magnitude of the currents in the first and second excitation coils of a flat double-gap magnetic lens.

Analyzing Eq. (7) for the on-axis magnetic induction distribution, we note that if the currents in the excitation coils of the two nonmagnetic gaps flow in opposite directions, then the condition

$$\int_{z_0}^{0} B_z(0,0,z)dz=0, \tag{10}$$

which corresponds to no rotation of the image of the object when the object is imaged using a planar magnetic lens, can hold. The object is located in the plane $z=z_0$, and its image is located in the plane $z=0$.

[1] P. W. Hawkes (ed.), *Magnetic Electron Lenses* (Springer, West Berlin, 1982), pp. 359–412.
[2] M. Szilgyi, *Electron and Ion Optics* (Plenum Press, New York, 1988) 639 pp.
[3] A. B. Abraamyants, V. D. Gelever, and V. A. Zhukov, in *Abstracts of the 14th All-Union Conference on Electron Microscopy* [in Russian], USSR Academy of Sciences Press, Moscow, 1990, 286 pp.
[4] W. Glaser, *Grundlagen Electronenoptik* (Springer-Verlag, Wien, 1952) 763 pp.
[5] V. M. Smirnov, *Course in Higher Mathematics*, Vol. 2 [in Russian] (GITTL, Moscow, 1957) 628 pp.

Translated by M. E. Alferieff

On the critical distance in heavy-ion collisions

V. I. Matveev, D. U. Matrasulov, and Kh. Yu. Rakhimov

Division of Thermal Physics, Uzbekistan Academy of Sciences, 700135 Tashkent, Uzbekistan
(Submitted June 10, 1998)
Zh. Tekh. Fiz. **69**, 135–137 (August 1999)

An analytical expression is obtained for the electron energy term by solving a two-center Dirac problem. This expression has the correct asymptotic form in two limiting cases — the Popov approximation and the nonrelativistic problem. © *1999 American Institute of Physics.* [S1063-7842(99)02708-7]

The two-center problem (i.e., the problem of the motion of an electron in the field of two fixed Coulomb centers with charges Z_1 and Z_2 separated by a distance R) is a classic problem of nonrelativistic quantum mechanics with applications in the theory of the chemical bond, in the physics of μ-meson processes, and so on. An extensive literature is devoted to it (see, for example, Refs. 1–3). The corresponding problem for the Dirac equation exhibits a number of new features that make solution difficult: 1) the variables in the Dirac equation with the potential

$$V=-\alpha\left(\frac{Z_1}{r_1}+\frac{Z_2}{r_2}\right)$$

are not separable in any orthogonal coordinate system; 2) for large Z falling into a center occurs; and, 3) the wave function is a multicomponent function, and for $Z\alpha\approx 1$ all components are of the same order of magnitude.

Interest in the two-center relativistic problem arose after it was indicated in Ref. 4 that quantum electrodynamics can be checked in experiments on heavy-ion collisions. As is well known,[6,7] for charges $Z\approx Z_{cr}=170$ the lowest energy level of the single-center Dirac equation sinks into the lower continuum and spontaneous positron creation starts. Since nuclei with $Z\approx 170$ do not exist, in Ref. 4 it was suggested that these fields could be obtained in heavy-ion collisions. To obtain the positron production cross section it is necessary to known the energy of quasimolecular states (Z_1,Z_2,e^-) as a function of distance, i.e., the energy term. In this connection, numerical[5] and analytical[6,7] calculations of the energy term and the critical distance (i.e., the distance for which the term sinks into the lower continuum) were performed. In Ref. 5, which contains the results of extensive numerical calculations, the Dirac equation for a two-center Coulomb potential was solved by diagonalizing the Dirac Hamiltonian in a two-center basis. Approximate analytical calculations of the energy term and the critical distance were performed in a series of works by V. S. Popov.[6–8] But the formulas obtained in these works permit calculating the energy term and the critical distance only numerically or for

$$\frac{Z_1+Z_2-Z_{cr}}{Z_{cr}}\ll 1$$

(the condition of small supercriticality) and for small intercenter distances.

In the present work, the Dirac equation for a two-center problem is solved by a method similar to the method of linear combination of atomic orbitals (LCAO), widely used to solve the nonrelativistic two-center problem.[2] As is well known, the LCAO method in application to the hydrogen molecular ion and the hydrogen molecule makes it possible to calculate the energy term in analytical form. However, similar calculations for the Dirac electron have not yet been performed. The analytical formula obtained by applying this method to the relativistic two-center problem permits calculating the energy term in a wide range of total nuclear charges and intercenter distances.

The system of units $\hbar=c=m_e=1$ is used throughout; R is the distance between nuclei; r_1 and r_2 are the distances of the electron from the nuclei; and, $Z_1=Z_2=Z$.

The motion of a relativistic electron in the field of two Coulomb centers is described by the stationary Dirac equation

$$H\Psi=E\Psi, \tag{1}$$

where $H=\boldsymbol{\alpha}\cdot\mathbf{p}+\beta+V$ is the Dirac Hamiltonian, and $\boldsymbol{\alpha}$ and β are Dirac matrices.

We shall solve Eq. (1) by the LCAO method, choosing the wave function in the form

$$\Psi=d_1\Psi_1+d_2\Psi_2,$$

where Ψ_1 and Ψ_2 are the wave functions of an electron moving in the field of the first and second centers, respectively.

As is well known,[2] the symmetry of the problem ($Z_1=Z_2$), the normalization conditions $\langle\Psi|\Psi\rangle=1$ and $\langle\Psi_j|\Psi_i\rangle=1$ $(j=1,2)$, and the fact that the ground state should not have any nodes give

$$d_1=d_2=d=\frac{1}{\sqrt{2(1+S)}},$$

where $S=\langle\Psi_1|\Psi_2\rangle$ is the overlap integral.

The electron energy can be calculated as the matrix element

$$E=\langle\Psi|H|\Psi\rangle, \tag{2}$$

where

$$\langle\Psi|=[\varphi\chi],\quad |\Psi\rangle=\begin{bmatrix}\varphi\\ \chi\end{bmatrix}$$

is a bispinor in the standard representation.

As mentioned above, we take for the functions Ψ_1 and Ψ_2 the relativistic wave functions of a hydrogen-like atom[9] with effective charge $Q\alpha$

$$\varphi_j=Ar_j^{\gamma-1}e^{-Q\alpha r_j}\begin{bmatrix}1\\0\end{bmatrix}=Ag_j\begin{bmatrix}1\\0\end{bmatrix},$$

$$\chi_j=iABr_j^{\gamma-1}e^{-Q\alpha r_j}\begin{bmatrix}\cos\theta\\ e^{i\varphi}\sin\theta\end{bmatrix}=iABg_j\begin{bmatrix}\cos\theta\\ e^{i\varphi}\sin\theta\end{bmatrix},$$

where $g_j=r_r^{\gamma-1}e^{-Q\alpha r_j}$; $j=1,2$; and,

$$A\frac{(2Q\alpha)^{\frac{3}{2}}}{\sqrt{4\pi}}\sqrt{\frac{1+\gamma}{2\Gamma(1+2\gamma)}}(2Q\alpha)^{\gamma-1},$$

$$B=\frac{1-\gamma}{Q\alpha},\quad \gamma=\sqrt{1-Q^2\alpha^2}.$$

Substituting these wave functions into Eq. (2) we obtain for the energy an expression containing five integrals which can be expressed analytically in terms of the complete $\Gamma(x)$ and incomplete $\Gamma(x,y)$ gamma functions[10] as

$$E=\frac{2\pi A^2bR^{2\gamma}}{1+S}\left[Q\alpha(I_1+I_2)+\frac{\alpha\gamma}{2Q\alpha}(I_3+I_4)-2Z\alpha(I_2+I_5)\right],\qquad(3)$$

$$S=2\pi A^2R^{2\gamma+1}bI_4,$$

where the integrals I_1–I_5 have the form

$$I_1=\frac{1}{\alpha^{2\gamma}}2\Gamma(2\gamma),$$

$$I_2=\frac{1}{a^{2\gamma}}\left[\left(2-\frac{a^2}{3(2\gamma-1)}\right)\Gamma(2\gamma,a)+\left(\frac{1}{3}+\frac{a}{3(2\gamma-1)}\right)a^{2\gamma}e^{-a}\right],$$

$$I_3=\frac{1}{a^{2\gamma+1}}4\gamma\Gamma(2\gamma),$$

$$I_4=\frac{1}{a^{2\gamma+1}}\left[\left(4\gamma-\frac{2a^2\gamma}{3(2\gamma-1)}\right)\Gamma(2\gamma,a)+\left(2+\frac{2a\gamma}{3(2\gamma-1)}\right)a^{2\gamma}e^{-a}\right],$$

$$I_5=\frac{1}{a^{2\gamma+1}}\left[(a-\gamma)\Gamma(2\gamma,2a)+(a+\gamma)\Gamma(2\gamma)\frac{1}{2}(2a)^{2\gamma}e^{-2a}\right],$$

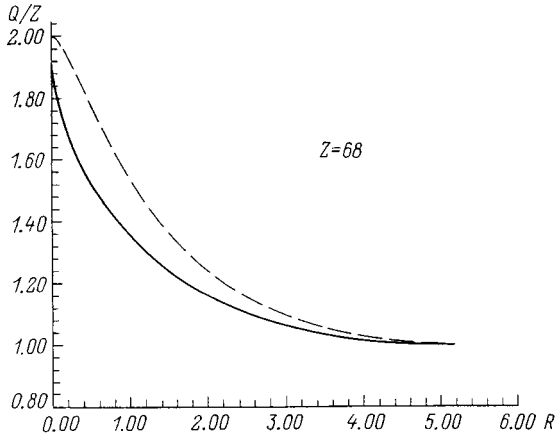

FIG. 1. Q/Z for $z=68$: dashed curve — dependence obtained by minimizing the expression for the nonrelativistic term; the solid curve is obtained by minimizing the expression for the relativistic term.

$$a=2Q\alpha R,\quad b=\frac{2}{1+\gamma}.$$

Thus, we have obtained the energy term of the electron as a function of the distance R, the charge Z, and the test charge Q. For $Q\alpha\ll 1$ ($\gamma\simeq 1$) Eq. (3) assumes the form

$$E=Q^2\alpha^2F_1(a)+Q\alpha F_2(a),\qquad(4)$$

where

$$F_1(a)=\frac{1}{2}\frac{1+e^{-a}\left(1+a-\frac{a^2}{3}\right)}{1+e^{-a}\left(1+a+\frac{a^2}{3}\right)},$$

$$F_2(a)=-Z\alpha\frac{1+2e^{-a}(1+a)+\frac{1}{a}-\left(\frac{1}{a}+1\right)e^{-2a}}{1+e^{-a}\left(1+a+\frac{a^2}{3}\right)}.$$

The expression (4) is identical to that obtained in the nonrelativistic LCAO method for the molecular hydrogen ion.[2]

The charge Q is, generally speaking, a function of R and Z, i.e.,

$$Q=Q(R,Z),$$

where, as is well known,[2]

$$Q|_{R\to 0}=2Z,\qquad(5)$$

$$Q|_{R\to\infty}=Z.\qquad(6)$$

Figure 1 shows the function Q/Z versus R for relativistic and nonrelativistic systems $Er-Er$ with charge $Z=68$, obtained by numerically minimizing expressions (3) and (4), respectively. As one can see, in the limits $R\to\infty$ and $R\to 0$ the effective charge approaches the charge of the isolated and unified atoms, respectively.

Figure 2 shows the relativistic and nonrelativistic energy terms for charge $Z=68$. It is evident that the relativistic corrections become substantial for smaller values of R.

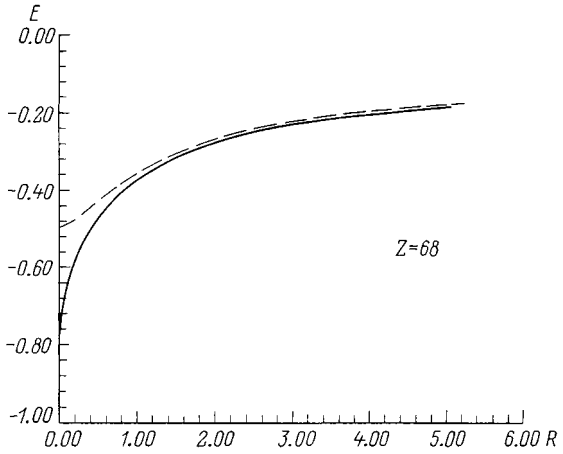

FIG. 2. Energy term for $Z=68$. The dashed curve is for a nonrelativistic system and solid curve is for a relativistic system.

Calculating the limit of expression (3) as $R\to 0$ taking account of Eq. (4), we find (unified atom)

$$E(R\to 0)=\sqrt{1-4Z^2\alpha^2}.$$

and in the limit $R\to\infty$ (isolated atom), we have taking account of Eq. (5),

$$E(R\to\infty)=\sqrt{1-Z^2\alpha^2}.$$

We note that in the cases considered above, the total charge of the nuclei is less than 137. For $Z_1+Z_2>137$, generally speaking, the finite sizes of the nuclei must be taken into account, i.e., regularization of the Coulomb potential is required at short distances. However, for $Z_1+Z_2>137$ Popov's formula (for the critical distance at which the ground state level sinks into the lower continuum), obtained in Ref. 6 by the method of joining asymptotic expressions at large distances from the nuclei, can be obtained directly from Eq. (3) under the same assumptions as in Ref. 6. Indeed, since in this case $\gamma=i\sqrt{4Z^2\alpha^2-1}$, $a\ll 1$, and $Q=2Z$, from the condition of sinking into the lower continuum

$$E(R_{\rm cr})=-1$$

we have

$$[(1-\gamma)4^{\gamma}+2\gamma]a^{2\gamma}=-4\gamma\Gamma(2\gamma+1)$$

or

$$(2R)^{-2\gamma}\frac{4\gamma}{(1-\gamma^2)^{\gamma}((\gamma-1)4\gamma-2\gamma)}\Gamma(2\gamma+1)=1\equiv e^{2\pi i}.$$

Simple transformations convert this expression into the form

$$R_{\rm cr}C\,\exp\left(-\frac{\pi}{\sqrt{4Z^2\alpha^2-1}}\right).$$

This expression is identical to that obtained previously in Ref. 8.

In summary, we have obtained an analytical expression (3) for the energy term of a relativistic electron moving in the field of two stationary Coulomb centers. This expression is valid in a wide range of total nuclear charges and intercenter distances. If the total charge of the nuclei is close to the critical value $(Z_1+Z_2\sim 170)$, Popov's well-known formula[6–8] for the critical distance between the nuclei at which the level sinks into the lower continuum and electron–positron pair creation occurs, obtains from this formula.

[1] H. A. Bethe, *Quantum Mechanics of Very Simple Systems*, ONTI, Moscow, 1935.
[2] J. C. Slater, *The Electronic Structure of Atoms and Molecules* [MIT Press, Cambridge, Mass., 1953; Mir, Moscow, 1965].
[3] I. V. Komarov, L. I. Ponomarev, and S. Yu. Slavyanov, *Spheroidal and Coulombic Spheroidal Functions* [in Russian], Nauka, Moscow, 1976.
[4] S. S. Gershteĭn and Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **57**, 654 (1969) [Sov. Phys. JETP **30**, 358 (1970)].
[5] B. Müller, J. Rafelski, and W. Greiner, Phys. Lett. B **47**, 5 (1973).
[6] V. S. Popov, Yad. Fiz. **17**, 621 (1973) [Sov. J. Nucl. Phys. **17**, 322 (1974)].
[7] V. S. Popov, Yad. Fiz. **19**, 155 (1974) [Sov. J. Nucl. Phys. **19**, 81 (1974)].
[8] A. A. Grib, S. G. Mamaev, and V. M. Mostapanenko, *Vacuum Quantum Effects in Strong Fields* [in Russian], Énergoatomizdat, Moscow, 1988.
[9] J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* [McGraw-Hill, New York, 1964; *Relativistic Quantum Field Theory*, Nauka, Moscow, 1978, Vol. 1].
[10] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. Stegan [Dover, New York, 1965; Nauka, Moscow, 1979].

Translated by M. E. Alferieff

Electric fields in dielectric structures with nonuniformly distributed surface charges

A. S. Yastrebov

M. A. Bonch-Bruevich St. Petersburg State University, 191186 St. Petersburg, Russia
(Submitted June 30, 1998)
Zh. Tekh. Fiz. **69**, 138–140 (August 1999)

The distribution of the potential and intensity of the electric field in dielectric structures with a nonuniform surface charge is analyzed. Computational formulas are proposed and computational results are presented. © *1999 American Institute of Physics.*
[S1063-7842(99)02808-1]

Nonuniform structures whose physical properties vary in a regular manner in space and give devices and systems new qualities have been increasingly used in recent years for information transmission and processing.[1–8] For example, an electret electroacoustic transducers with nonuniformly charged polymer-film surfaces are much more reliable and have a much longer service life than transducers based on uniformly charged films.[9]

We shall examine a method for calculating the electric field above nonuniformly charged electrets. Let a regular periodic distribution of either the potential V or surface charge density σ be given on a flat interface between two dielectric media with relative permittivities ε_1 and ε_2. Let the interface be the (x,y) coordinate plane, and let us represent $V(x,y)$ or $\sigma(x,y)$ as Fourier expansions.

We consider first the case where the potential on the (x,y) plane is constant, V_s, inside zero outside rectangular spots (Figs. 1a and b).

We represent the distribution $V(x,y)$ as a product of Fourier series

$$V(x,y)=V_sF1(x)\cdot F2(y)=V_sF1\cdot F_2, \tag{1}$$

where

$$F1=d/a+(2/\pi)\sum_{m=1} m^{-1}\sin(\pi md/a)\cos(2\pi mx/a), \tag{2}$$

$$F2=h/b+(2/\pi)\sum_{n=1} n^{-1}\sin(\pi nh/b)\cos(2\pi ny/b). \tag{3}$$

Let the potential $\varphi(x,y,z)$ above the (x,y) plane be expressed as

$$\varphi(x,y,z)=V_sF1(x)F2(y)F3(z).$$

Solving the Laplace equation $\Delta\varphi(x,y,z)=0$ we obtain the potential distribution above the (x,y) plane, placed on the interface between the two media with relative permittivities ε_1 and ε_2, the first of which fills the bottom half-space and the second the top half-space:

$$\begin{aligned}\varphi(x,y,z)=-E_cz+V_s\Big\{&(dh/ab)+(2/\pi)\Big[(h/b)\\&\times\sum_{m=1} m^{-1}\exp(-2\pi mz/a)\\&\times\sin(\pi md/a)\cos(2\pi mx/a)+(d/a)\\&\times\sum_{n=1} n^{-1}\exp(-2\pi nz/b)\sin(\pi nh/b)\\&\times\cos(2\pi ny/b)\Big]+(4/\pi^2)\sum_{m=1}\sum_{n=1}(mn)^{-1}\\&\times\exp(-\lambda_{mn^z})\sin(\pi md/a)\cos(2\pi mx/a)\\&\times\sin(\pi nh/b)\cos(2\pi ny/b)\Big\},\end{aligned} \tag{4}$$

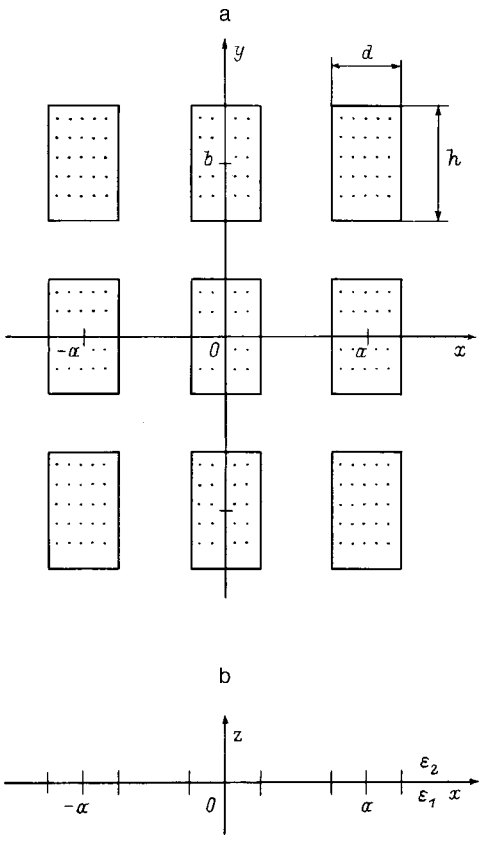

FIG. 1. Nonuniformly charged structure: a — plan view, b — section view.

where E_c is the average value of E_z, the z component of the electric field above the (x,y) plane;

$$\lambda_{mn}=2\pi[(m/a)^2+(n/b)^2]^{0.5}. \tag{5}$$

The electric field components $E_x=E_x(x,y,z)$, $E_y=E_y(x,y,z)$, and $E_z=E_z(x,y,z)$ above a periodically charged plane can be obtained by differentiating the expression for the potential $\varphi(x,y,z)$ with respect to the corresponding coordinates: The z component of the field intensity above the plane calculated according to the formula

$$\begin{aligned}E_z(x,y,z)=-\partial\varphi(x,y,z)/\partial z=E_c+V_s\Bigg\{4\Bigg[(h/ab)&\\ \times\sum_{m=1}\exp(-2\pi mz/a)\sin(\pi md/a)&\\ \times\cos(2\pi mx/a)+(d/ab)&\\ \times\sum_{n=1}\exp(-2\pi nz/b)\sin(\pi nh/b)&\\ \times\cos(2\pi ny/b)\Bigg]+(4/\pi^2)&\\ \times\sum_{m=1}\sum_{n=1}\lambda_{mn}(mn)^{-1}&\\ \times\exp(-\lambda_{mn}z)\sin(\pi md/a)\cos(2\pi mx/a)&\\ \times\sin(\pi nh/b)\cos(2\pi ny/b)\Bigg\},&\end{aligned} \tag{6}$$

is shown in Fig. 2.

The electric field component $E_z(x,y,0)$ on the (x,y) surface can be obtained from Eq. (6), setting $z=0$.

The average value E_c of the field normal to the (x,y) surface depends on the electric charge distribution in space. If the source of the electric field are only charges in the (x,y) plane, then E_c is related with the average surface charge density σ_c as

$$E_c=\sigma_c/[\varepsilon_0(\varepsilon_1+\varepsilon_2)]. \tag{7}$$

If the total charge on a plane is zero, $\sigma_c=0$, we must set $E_c=0$ in Eqs. (4) and (6).

If the surface charge density in the (x,y) plane (Figs. 1a and b) is constant, $+\sigma_s$, inside and zero outside the rectangular spots, then

$$\sigma_c=\sigma_s(dh/ab) \tag{8}$$

and the distribution $\sigma(x,y)$ can be represented as a product of Fourier series

$$\sigma(x,y)=\sigma_s F1(x)F2(y), \tag{9}$$

where $F1(x)$ and $F2(y)$ are determined from Eqs. (2) and (3).

The potential distribution is determined by the expression

$$\begin{aligned}\varphi(x,y,z)=\sigma_s/[\varepsilon_0(\varepsilon_1+\varepsilon_2)]\Bigg\{-(dh/ab)z+(1/\pi^2)\Bigg[(ha/b)&\\ \times\sum_{m=1}m^{-2}\exp(-2\pi mz/a)\sin(\pi md/a)&\\ \times\cos(2\pi mx/a)+(db/a)\sum_{n=1}n^{-2}&\\ \times\exp(-2\pi nz/b)\sin(\pi nh/b)\cos(2\pi ny/b)\Bigg]&\\ +(4/\pi^2)\sum_{m=1}\sum_{n=1}(mn\lambda_{mn})^{-1}\exp(-\lambda_{mn}z)&\\ \times\sin(\pi md/a)\cos(2\pi mx/a)\sin(\pi nh/b)&\\ \times\cos(2\pi ny/b)\Bigg\},&\end{aligned} \tag{10}$$

whence we can obtain the z component of the field, shown in Fig. 3, as

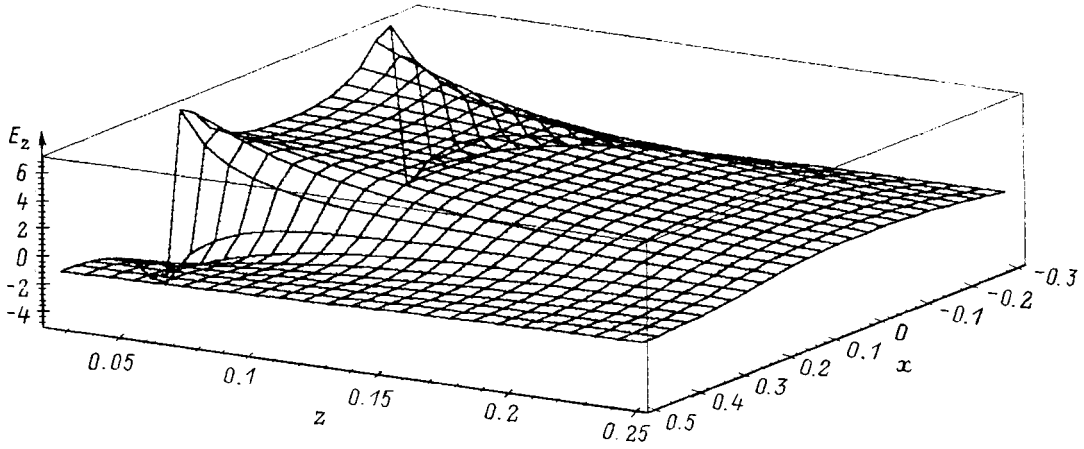

FIG. 2. Distribution of the z component E_z of the electric field above a nonuniformly charged surface with a fixed distribution of the potential V_s.

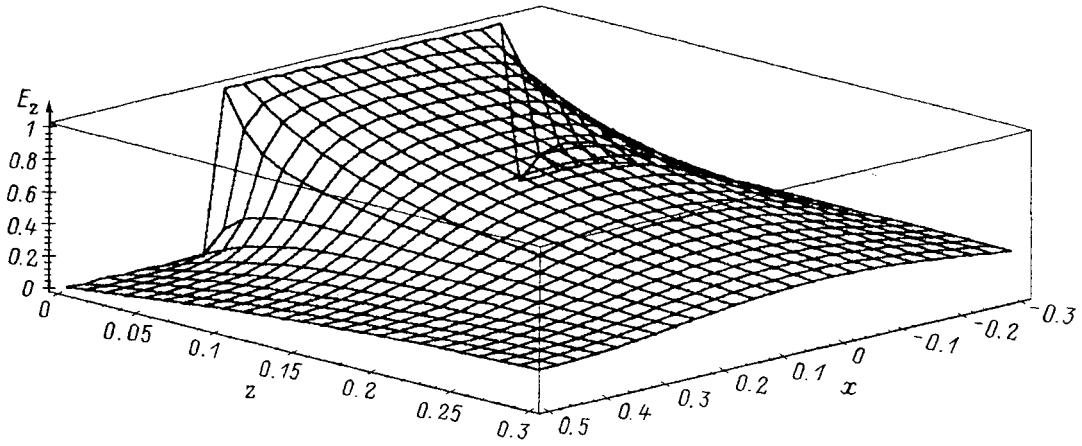

FIG. 3. Distribution of the field E_z above a nonuniformly charged surface for the case where the surface charge density σ_s is given within the rectangles.

$$E_z(x,y,z)=\sigma_s/[\varepsilon_0(\varepsilon_1+\varepsilon_2)]\Bigg\{(dh/ab+(2/\pi))\times\Bigg[(h/b)\sum_{m=1} m^{-1}\exp(-2\pi mz/a)\times\sin(\pi md/a)\cos(2\pi mx/a)+(d/a)\times\sum^{n=1} n^{-1}\exp(-2\pi nz/b)\sin(\pi nh/b)\times\cos(2\pi ny/b)\Bigg]+(4/\pi^2)\sum_{m=1}\sum_{n=1}(mn)^{-1}\times\exp(-\lambda_{mn}z)\sin(\pi md/a)\cos(2\pi mx/a)\times\sin(\pi nh/b)\cos(2\pi ny/b)\Bigg\}. \quad (11)$$

The method considered above can be used to calculate the electric field in electroacoustic transducers with electrets consisting of nonuniformly charged dielectric films.

The proposed method for calculating electrets with a regular surface charge distribution was used to develop software for a control and measurement system intended for use in operations of controlling the parameters of electroacoustic transducers.

[1] *Optics and Communications: Optical Transmission and Processing of Information* (Mir, Moscow, 1984), 468 pp.
[2] A. W. Snyder and J. D. Love, *Optical Waveguide Theory* [Chapman and Hall, New York-London, 1983; Radio i Svyaz', Moscow, 1987, 656 pp.].
[3] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. [Wiley, New York, 1981; Mir, Moscow, 1984, 456 pp.].
[4] A. Blicher, *Field-Effect and Bipolar Power Transistor Physics* (Academic Press, New York, 1981; Énergoatomizdat, Leningrad, 1986, 248 pp.].
[5] J. Kaufhold, A. Jastrebov, and M. Volokobinsky, Jahresbericht 1995 der Deutsche Telekom FH Dieburg, 1995, pp. 103–113.
[6] Y. Tarui (ed.), *VLSI Technology: Fundamentals and Applications* (Springer-Verlag, Berlin–New York, 1986; Radio i Svyaz', Moscow, 1985, 480 pp.].
[7] I. Broudaĭ and Dzh. Mereĭ, *Physical Principles of Microtechnology*, edited by A. V. Shal'nov (Mir, Moscow, 1985, 496 pp.).
[8] E. Dieulesaint and D. Royer, *Elastic Waves in Solids* [Wiley, New York, 1981; Mir, Moscow, 1982, 424 pp.].
[9] G. Aĭzenbletter, Author's Abstract of Candidate's Dissertation [in Russian], Leningrad, 1982, 21 pp.

Translated by M. E. Alferieff

On the nature of the mobile charge in passivating coatings based on lead–borosilicate glasses

S. I. Vlasov, P. B. Parchinskiĭ, A. A. Nasirov, and B. A. Olmatov

Tashkent State University, 700095 Tashkent, Uzbekistan
(Submitted June 18, 1998)
Zh. Tekh. Fiz. **69**, 141–142 (August 1999)

The temperature dependences of the mobile charge in lead–borosilicate glass passivating coatings are investigated. Mechanisms are proposed for the formation of the mobile charge in the structure of the glass. It is shown that the temperature dependence of the mobile charge becomes stronger with increasing content of the crystal phase in the interior volume of a passivating coating. © *1999 American Institute of Physics.* [S1063-7842(99)02908-6]

Mobile charge in insulator coatings used for passivating semiconductor surfaces strongly influences the characteristics of the passivated devices.[1] In Ref. 1 it was observed that mobile charge is present in the interior volume of passivating coatings based on lead–borosilicate glasses. Our objective in the present work is to determine the mechanism leading to the formation of this charge.

The composition and fabrication method of the coatings investigated are similar to those described in Ref. 3. The experimental structures were divided into two groups: group-1 coatings melted at $T=680\,^{\circ}\mathrm{C}$ and group-2 coatings melted at $T=700\,^{\circ}\mathrm{C}$.

The method of high-frequency capacitance–voltage characteristics was used to investigate the charges localized in the interior volume of the insulator.[4] The measurements were performed by the bridge compensation method in the dark in the temperature range from -30 to $+80\,^{\circ}\mathrm{C}$ with the holding time $t_0=2$ min at each point. The frequency of the test signal was 80 kHz–1.5 MHz, which satisfies the criteria for the capacitance–voltage characteristics of metal–insulator–semiconductor structures (MIS structures) to be high-frequency characteristics.[5] The typical capacitance–voltage dependences for coatings in both groups are displayed in Fig. 1. Hysteresis due to the presence of mobile charge Q_m is observed for all samples. The width of the hysteresis loop is independent of the duration of the enriching voltage. The mobile charge was determined as

$$Q_m=C_0\Delta V, \tag{1}$$

where C_0 is the capacitance of the insulator and ΔV is the width of the hysteresis loop.

For group-1 coatings the charge $Q_m=(2.2\pm0.2)\times10^{-9}\,\mathrm{C/cm^2}$ remains constant in the entire experimental temperature range. The observed shift of the capacitance–voltage characteristics on the voltage axis is due to the temperature dependence of the charge Q_{ss} of the surface states.[6]

For group-2 coatings the values of Q_m increase with decreasing temperature. The broadening in the lower part of the hysteresis loop, reaching a maximum at $C/C_0=0.7$, is interesting. The temperature dependence of Q_m for group-2 coatings at the characteristic points $C/C_0=0.7$ and 0.9 is presented in Fig. 2. As the temperature increases to $T=40\,^{\circ}\mathrm{C}$, the broadening in the lower part of the hysteresis loop vanishes, and the mobile charge becomes $Q_{m_2}=(2.4\pm0.2)\times10^{-9}\,\mathrm{C/cm^2}$ and remains constant with further increase in temperature. This behavior of the capacitance–voltage characteristics indicates the following. For both groups of structures, for an enriching voltage electrons are

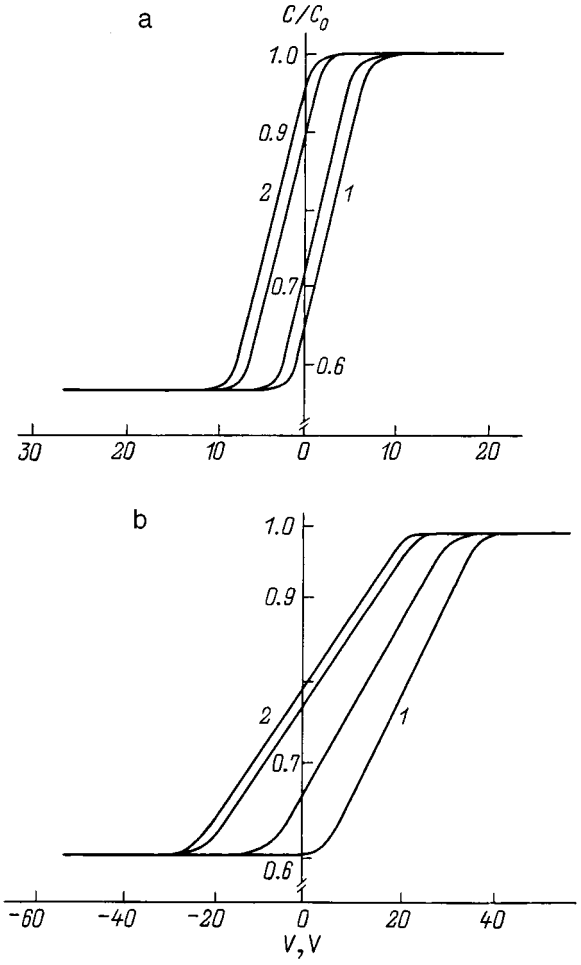

FIG. 1. Typical capacitance–voltage characteristics for group-1 (a) and group-2 (b) coatings: $T=-20$ (*1*) and 40 °C (*2*).

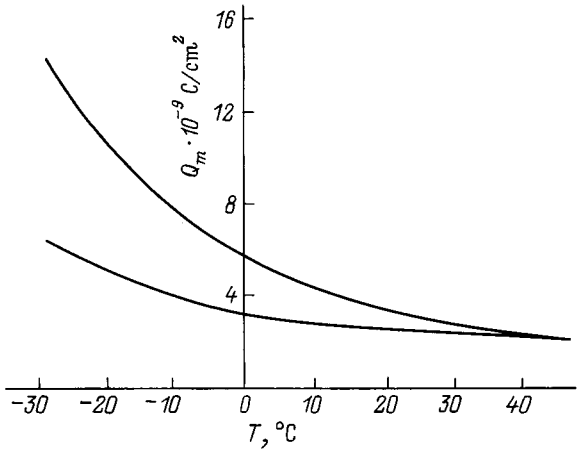

FIG. 2. Mobile charge versus the temperature for $C/C_0=0.95$ (*1*) and 0.7 (*2*).

injected from the semiconductor into the interior volume of the glass. The glass contains Pb^{2+} ions, which are well known to have a high polarizability.[7] The localization of electrons near the ions of the easily polarizable elements results in the formation of polarons, which drift between heterovalent Pb ions as the external field varies.[8] Polaron drift shifts the center of charge in the insulator and shifts the capacitance–voltage characteristics along the voltage axis.[9] In the experimental temperature range the polaron mobility is independent of the activation energy,[7] which explains the independence of Q_m from T in group-1 coatings, as well as in group-2 coatings at temperatures $T>40\,^\circ$C. In this case, Q_m in structures from both groups depends only on the number of electrons trapped by lead ions and is determined by the composition of the glass. This is confirmed by the fact that the mobile charge in the group-2 coatings at temperatures $T>40\,^\circ$C becomes equal to the mobile charge in group-1 coatings.

The temperature dependence of the mobile charge observed in group-2 coatings at temperatures $T<40\,^\circ$C is explained as follows. As the temperature of formation of the coatings increases from 680 to 700 °C, the number and volume of inclusions of the crystal phase in the glass structure increase sharply.[2] The potential barriers formed in the process prevent electron extraction from the interior volume of the glass. Then Q_m is determined by the condition

$$Q_m = Q_0 - \int_0^{t_0} j(t)\,dt + Q_{m_1}, \tag{2}$$

where Q_0 is the charge initially trapped on the potential barriers, Q_{m_1} is the charge due to polaron drift, $j(t)$ is the thermionic current, and t is the measurement time at a fixed voltage.

The thermionic emission current increases with temperature (the dependence of $j(t)$ on T is determined by the emission mechanism[1]), and for $T>40\,^\circ$C the relaxation time of the charge trapped on the barriers becomes less than t_0. In group-1 coatings, where the number of phase nonuniformities is much smaller,[2] the charge localized on them relaxes in a time less than t_0 in the entire experimental temperature range.

In summary, the presence of a mobile charge in passivating coatings based on lead–borosilicate glasses is due to electron localization near the easily polarizable lead ions and relaxation of the charge accumulated on the potential barriers. As the content of the crystal phase in the glass volume increases, the mobile charge becomes temperature-dependent and as a result the characteristics of the passivated devices show temperature instabilities.

[1] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. [Wiley, New York, 1981; Mir, Moscow, Vol. 1, 1984, 456 pp.].
[2] P. B. Parchinskiĭ, S. I. Vlasov, and A. A. Nasirov, Neorg. Mater. **36**, 1278 (1996).
[3] P. B. Parchinskiĭ, S. I. Vlasov, and A. A. Nasirov, Mikroélektronika **26**(1), 62 (1997).
[4] V. G. Karpechenko, S. A. Inozemtsev, and É. N. Loĭtsker, Élektron. Tekh. Ser. 2, No. 2(45), 84 (1969).
[5] A. S. Grove, E. H. Snow, and B. E. Deal, J. Appl. Phys. **35**, 2458 (1964).
[6] Yu. S. Chistov and V. F. Syporov, *Physics of MIS Structures* [in Russian], Voronezh State University, Voronezh, 1989.
[7] S. M. Brekhovskikh, Yu. M. Viktorova, and A. M. Landa, *Radiation Effects in Glasses* [in Russian], Énergoatomizdat, Moscow, 1988.
[8] P. T. Oreshkin, *Physics of Semiconductors and Insulators* [in Russian], Vysshaya Shkola, Moscow, 1977.
[9] V. S. Vasilenko, Ya. O. Roĭzin, and V. I. Solomenko, Poverkhnost', No. 7, 151 (1990).

Translated by M. E. Alferieff