Neutralization of H^- ions in collisions with fast, multiply charged ions

A. B. Voĭtkiv and V. A. Pazdzerskiĭ

Institute of Electronics, Academy of Sciences of Uzbekistan, 700143 Tashkent, Uzbekistan
(Submitted October 13, 1997; resubmitted April 13, 1998)
Zh. Tekh. Fiz. **69**, 1–4 (July 1999)

The neutralization of H^- ions in collisions with fast, multiply charged ions is considered in the parameter region where the Born approximation applies. An analytical formula is obtained for the H^- neutralization cross section in such collisions. © *1999 American Institute of Physics.*
[S1063-7842(99)00107-5]

Atomic collisions involving negative ions have been the subject of intensive research in the past decades (see, e.g., Refs. 1–3 and the literature cited therein). The study of such collisions can have important practical applications (e.g., for obtaining beams of fast neutrals[1)]. For calculating the neutralization cross sections of negative ions in collisions with fast charged particles having relatively low charges ($Z \ll v$, where Z is the charge of the particle and v is the collision velocity; here and below we use the atomic system of units) one can use the Born approximation.[4] In the present paper we consider the neutralization of H^{-1} ions in collisions with fast, multiply charged ions (MCIs) in the parameter region $Z \gtrsim v \gg v_0$ (v_0 is the characteristic orbital velocity of the weakly bound electron in H^-), where the Born approximation is inapplicable. We note that for certain values of Z and v in this parameter region the H^- neutralization cross sections have been measured[2] for collisions with Ne ions ($Z \leqslant 4$) and with Ar and Xe ions ($Z \leqslant 8$). The neutralization of H^- in collisions with MCIs has been studied theoretically in Refs. 1–3, 5, and 6 by the method of classical Monte Carlo trajectories,[1] by a method proposed in Ref. 7 as a generalization of the Keldysh theory[8] for photoionization in a high field,[2,3] in a two-state model,[5] and by the coupled channels method.[6] In the present paper the problem of H^- neutralization is treated by the approach proposed in Refs. 9 and 10 for calculating the ionization cross sections of atoms in collisions with fast MCIs. Unlike the methods listed above, this approach can yield an analytical solution for the cross section, and the results can be used over a rather wide range of collision parameters Z and v, which will be specified below.

Let us examine the problem. In order to calculate the neutralization cross section one needs to know the electronic wave functions of the negative ion. The H^- ion has two electrons, which are usually treated theoretically under the assumption that one of them is found in an almost hydrogenlike $1s$ orbital, while the other is weakly bound, in a diffuse orbital of radius $\sim \varkappa^{-1} \sim 4$ ($\varkappa^2/2 = 0.0275$ is the electron affinity of H^-). In this paper we shall describe the negative ion in the one-electron approximation, assuming that the weakly bound active electron moves in the field of a ''frozen'' core of the H^- ion (the proton plus the tightly bound electron). It is known (see, e.g., Ref. 11) that the motion of an individual particle under the influence of other particles can be described as motion of this particle in an effective nonlocal potential. As in Refs. 5 and 12, we shall describe the interaction of the active electron with the core by the nonlocal separable Yamaguchi potential[13]

$$V(\mathbf{r},\mathbf{r}') = -\lambda |g(\mathbf{r})\rangle\langle g(\mathbf{r}')|, \tag{1}$$

where $g(\mathbf{r})$ is chosen in the form

$$g(\mathbf{r}) = g(r) = (\beta/2\pi)^{1/2} \exp(-\beta r)/r \tag{2}$$

with $\beta = 0.913$ and $\lambda = 0.659$ (Ref. 12); r is the distance between the active electron and the nucleus of the H^- ion.

With this potential the Schrödinger equation is easily solved for both the discrete spectrum and the continuum. For example, the bound state wave function is

$$\varphi_0(r) = N r^{-1}(\exp(-\varkappa r) - \exp(-\beta r)), \tag{3}$$

where $N = ((2\pi)^{-1}\varkappa\beta(\varkappa+\beta)\beta-\varkappa)^{-2})^{1/2}$ is a normalizing factor, and $\varkappa = 0.235$ is specified according to the known value of the electron affinity.

The wave function (4) has the correct asymptotic form for $r\varkappa \gg 1$:

$$\varphi_0(r) = 1.51((2\pi)^{-1}\varkappa)^{1/2}\exp(-\varkappa r)/r, \tag{4}$$

remaining finite as $r \to 0$. Below we will use the wave functions obtained with the use of the potential (1) to describe the eigenstates of the discrete and continuous spectra of H^-.

Let us consider a collision of a fast MCI with a negative hydrogen ion. We assume that the hydrogen ion is at rest at the origin of the coordinate system and that the MCI is moving along a classical rectilinear trajectory $\mathbf{R}(t) = \mathbf{b} + \mathbf{v}t$, where b is the impact parameter. It is known[14] that the main contribution to the cross sections of inelastic processes in collisions with fast MCIs comes from the region of impact parameters $b > r_0$, where r_0 is the characteristic dimension of the atomic system (in our case $r_0 \simeq \varkappa^{-1}$, the characteristic size of H^-). For $b > r_0$ the field of the MCI in the region where the hydrogen ion is found is assumed to be uniform,

$$W(\mathbf{r},t) = Z/|\mathbf{R}(t) - \mathbf{r}| \simeq Z/R(t) - Z\mathbf{R}(t)\cdot\mathbf{r}/R^3(t). \tag{5}$$

The first term in Eq. (5), which does not depend on the electron coordinates r, leads to a slight phase shift which is the same for all states, and we shall henceforth ignore it as it does not lead to electronic transitions. Electronic transitions

 © 1999 American Institute of Physics

are due to the dipole interaction $W_1(t) = -Z\mathbf{R}(t)\cdot\mathbf{r}/R^3(t)$, which can be regarded as acting over a finite time interval. In fact,

$$\int_{-\infty}^{+\infty} dt\, W_1(t) = W_1(t=0)\cdot T, \tag{6}$$

where $T = T(b) = 2b/v$ has the meaning of the effective time of interaction with the field of the MCI.

Following the approach proposed in Refs. 9 and 10, we divide the region of impact parameters $b > r_0$ into two subregions: 1) $r_0 < b \ll v\tau$, where $\tau \simeq r_0/v_0$ is the characteristic orbital time for the weakly bound electron in the H^- ion, and 2) $b \gg Z\varkappa/v$. Using the wave function (3), we can estimate the characteristic orbital velocity of the electron in this state as $v_0 \simeq (\varkappa\beta)^{1/2} \simeq \varkappa^{1/2}$, which gives $\tau \simeq \varkappa^{-3/2}$. For what follows it is important to note that subregion 1 exists under the condition $v \gg v_0 \sim \varkappa^{1/2}$ and that subregions 1 and 2 partially overlap under the condition $v^2 \gg Z\varkappa^{1/2}$, and we shall therefore assume below that both of these conditions hold.

In collisions in the first subregion the effective collision time $T(b)$ is short ($T \ll \tau$), and to calculate the neutralization probability $w_{-0}(b)$ in this region we will use the zeroth-order sudden approximation[15]

$$\begin{aligned} w_{-0}(b) &= w_g(b) \\ &= \int d^3\mathbf{k} \left| \left\langle \varphi_{\mathbf{k}} \right| \exp\left(-i \int_{-\infty}^{+\infty} dt\, W_1(t) \right) \left| \varphi_0 \right\rangle \right|^2 \\ &= \int d^3\mathbf{k} \left| \langle \varphi_{\mathbf{k}} | \exp(i\mathbf{q}\cdot\mathbf{r}) | \varphi_0 \rangle \right|^2, \end{aligned} \tag{7}$$

where φ_0 and $\varphi_{\mathbf{k}}$ are the wave functions of the bound state and the states of the continuous spectrum of H^-, and $\mathbf{q} = 2Z\mathbf{b}/b^2v$ has the meaning of the average momentum transfer to the electron from the field of the fast MCI.

The probability (7) can be evaluated using the condition of completeness of the states of the negative ion,

$$|\varphi_0\rangle\left\langle \varphi_0 \right| + \int d^3\mathbf{k} \left| \varphi_{\mathbf{k}} \right\rangle \langle \varphi_{\mathbf{k}}| = 1,$$

which leads to the expression

$$\begin{aligned} w_g(b(q)) = 1 - \frac{16\pi^2 N^4}{q^2} &(\tan^{-1}(q/2\varkappa) + \tan^{-1}(q/2\beta) \\ &- 2\tan^{-1}(q/(\varkappa+\beta)))^2. \end{aligned} \tag{8}$$

The contribution to the neutralization cross section from the states in the region of impact parameters $b_1 \ll b \ll b_2$ (where $b_2 \ll v\tau$, and the value of b_1 will be determined below) has the form

$$\begin{aligned} \Delta\sigma_{-0}(b_1 \leqslant b \leqslant b_2) &= 2\pi \int_{b_1}^{b^2} db\, b w_g(b) \\ &= 8\pi \frac{Z^2}{v^2} \int_{q_2}^{q^1} \frac{dq}{q^3} p(q), \end{aligned} \tag{9}$$

where $p(q) = w_g(b(q))$, $q_{1,2} = 2Z/(vb_{1,2})$.

Conditions in the region of small impact parameters $b \leqslant r_0$ for $Z \gtrsim v$ are characterized by large (in comparison with the binding energy of the active electron in H^-) energy transfers to the electron. Indeed, for $b > r_0$ the average energy transfer can be estimated as $\varepsilon \cong 2Z^2/b^2v^2$. It follows that this energy is large even for $b \cong r_0$ (and, of course, becomes even larger in the region $b < r_0$). The large average energy transfer leads to detachment of the weakly bound electron with a probability close to unity. Here it is important to note the following. Even though expression (7) is formally applicable only for $b > r_0$, calculations using this formula gives reasonable values (close to unity) for the neutralization probability (for $Z \gtrsim v$) in the region $b \leqslant r_0$ as well. Therefore, for $Z \gtrsim v$ we shall use expression (7) to calculate the neutralization probability even in the region of small impact parameters $b \leqslant \varkappa^{-1}$. Since a calculation shows that the contribution (9) to the neutralization cross section for $q_1 > 1$ is practically insensitive to the choice of the value of q_1, we shall simply set $q_1 = \infty$ ($b_1 = 0$) in Eq. (9).

The integral over q in Eq. (9) is evaluated as follows. Since $p(q) \sim (q/\varkappa)^2$ for $q \ll \varkappa$, we make use of the relations $\varkappa \ll v$ and $q_2 \ll \varkappa$ (the latter holds for $Z/(\varkappa v) \ll b_2 \ll v/\varkappa^{3/2}$) to write

$$\begin{aligned} \Delta\sigma_{-0}(b_1 \leqslant b \leqslant b_2) &= 8\pi \frac{Z^2}{v^2} \int_{q_2}^{\infty} \frac{dq}{q^3} p(q) \\ &= 152.2 \frac{Z^2}{v^2} \left(\int_{q_2}^{\varkappa^{3/2}/v} \frac{dq}{q^3} q^2 \right. \\ &\quad \left. + 0.1652 \int_{\varkappa^{3/2}/v}^{\infty} \frac{dq}{q^3} p(q) \right). \end{aligned} \tag{10}$$

The second term in parentheses is independent of the charge Z of the MCI and depends on the velocity v only through the lower limit of integration. Since $\varkappa^{3/2}/v \ll 1$, this dependence is, to good accurac,y logarithmic. Indeed, we write the second integral in (10) in the form

$$\int_{\varkappa^{3/2}/v}^{q_1} \frac{dq}{q^3} w(q) = (0.1652)^{-1} \ln(Cv/\varkappa^{3/2}). \tag{11}$$

Numerical integration of the left-hand side of Eq. (11) for different values of v in the region $v/\varkappa^{3/2} \gg 1$ shows that $C = 0.46$ practically independently of the value of v in this region of velocities. Thus, using Eqs. (10) and (11), we obtain for the contribution (9) to the cross section

$$\Delta\sigma_{-0}(b \leqslant b_2) = 152.2 \frac{Z^2}{v^2} \ln\left(\frac{0.23v}{Z} b_2 \right). \tag{12}$$

In the region of impact parameters $b \gg Z/(\varkappa v)$ the neutralization probability is small. In fact, for $z/(\varkappa v) \ll b \ll v\tau$ we can use expressions (7) and (8) to obtain

$$w_{-0}(b) = 1.34 \frac{Z^2}{b^2v^2\varkappa^2} \ll 1; \quad Z/(\varkappa v) \ll b \ll v/\varkappa^{3/2}. \tag{13}$$

Therefore, to describe the ionization at $b \gg Z/(v\varkappa)$ we shall therefore use perturbation theory in the interaction $W_1(t)$. In the first order of perturbation theory we have for the ionization probability (see, e.g., Ref. 16)

$$w_{-0}(b) = w_p(b) = \frac{4Z^2}{v^4}\int_0^\infty dk\, k^2 \omega_{k1}^2 y_{k1}^2 \left(K_0^2\left(\frac{\omega_{k_1} b}{v}\right) + K_1^2\left(\frac{\omega_{k1} b}{v}\right)\right), \quad (14)$$

where $\omega_{k1} = (k^2+\varkappa^2)/2$ are the frequencies of the transitions, y_{k1}^2 is the mean square value of the component of the dipole matrix element for H^- ($y_{k1}^2 = x_{k1}^2 = z_{k1}^2 = r_{k1}^2/3$), and K_0 and K_1 are modified Bessel functions.

For the contribution to the neutralization cross section from collisions with impact parameters $b_3 \leqslant b < \infty$, where the point b_3 lies in the range $Z/(v\varkappa) \ll b^3 \ll v/\varkappa^{3/2}$, we obtain

$$\Delta\sigma_{-0}(b \geqslant b_3) = 2\pi \int_{b_3}^\infty db\, b w_p(b) = 152.2 \frac{Z^2}{v^2} \ln\left(\frac{1.123 v}{\omega_{\text{eff}} b_3}\right), \quad (15)$$

where

$$\omega_{\text{eff}} = \exp\left(\int_0^\infty dk\, k^2 y_{k1}^2 \ln\omega_{k1} \Big/ \int_0^\infty dk\, k^2 y_{k1}^2 \right) = 0.081.$$

Because for $Z\varkappa^{1/2} \ll v^2$ the two investigated subregions of impact parameters partially overlap, we can set $b_2 = b_3$ and, summing the contributions (12) and (15), we obtain for the ionization cross section

$$\sigma_{-0} = 152.2 \frac{Z^2}{v^2} \ln\left(\frac{3.2 v^2}{Z}\right). \quad (16)$$

Let us refine the parameter region in which the approach taken above can be used. First, we have made use of the ''suddenness'' of the collision, $v \gg v_0 \simeq \varkappa^{1/2}$. Second, the partial overlap of the two regions of impact parameters considered can take place for $Z\varkappa^{1/2} \ll v^2$. Third, to minimize the error introduced by the method used to calculate the contribution to the cross section from collisions with small impact parameters ($b \lesssim r_0$), it is necessary to have $Z \gtrsim v$. Consequently, this approach is applicable in the parameter region $v \lesssim Z \ll v^2/\varkappa^{1/2}$, $v \gg \varkappa^{1/2}$. Here we should also note the following. Since the active electron in H^{-1} has a very low binding energy and a low orbital velocity on atomic scales, even collisions with particles with $Z \sim 1$ and $v \sim 1$ can be treated as collisions with fast, ''multiply charged'' ions, and the neutralization cross section of H^- in such collisions can be estimated using formula (16).

Figure 1 shows a comparison of the neutralization cross sections calculated using formula (16) (the solid curve) with the experimental data of Ref. 2 on the cross sections for neutralization of H^- ions by Ne ions ($Z \leqslant 4$) and by Ar and Xe ions ($Z \leqslant 8$) at a collision energy of 200 keV in the center-of-mass frame. The results of the calculation using formula (16) are in good agreement with the available results of numerical calculations of other authors[2,5,6] in this region of parameters Z and v.

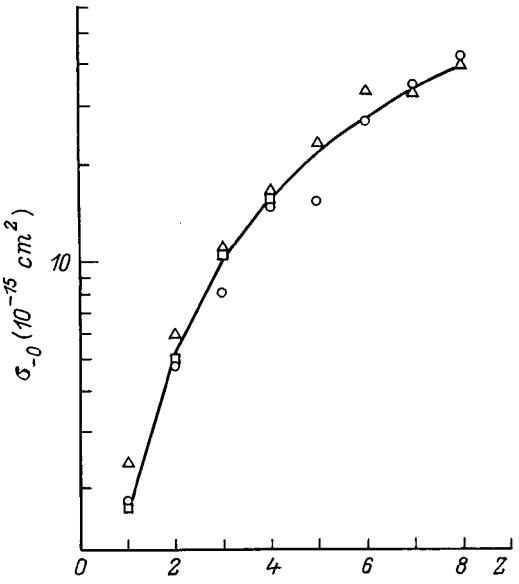

FIG. 1. Cross sections for the neutralization of H^- at collision energy $E = 200$ keV. The symbols are the experimental data of Ref. 2: □—Ne^{2+} ($Z \leqslant 4$), ○—Ar^{2+} ($Z \leqslant 8$), △—Xe^{2+} ($Z \leqslant 8$); the solid curve shows the results of a calculation using formula (16).

In this paper, as in Ref. 5, a bound-state wave function of the form (3) was used in the calculation of the H^- neutralization cross section. At the same time, in Refs. 2 and 3 this cross section was calculated for a wave function taken in the form (4) (but for all r, $0 \leqslant r < \infty$). We should therefore like to conclude with a brief discussion of the influence of the form of the bound-state wave function on the calculated value of the cross section in the region $\varkappa r \lesssim 1$. For this purpose we carried out an additional calculation of the H^- neutralization cross section using the asymptotic form (5) of the wave function for all r ($0 \leqslant r < \infty$). We found that the main contribution to the cross section still comes from the region of impact parameters $b > \varkappa^{-1}$, and the calculated value of the cross sections exceed those calculated using (16) by approximately 20%. An uncertainty of ~20% in the theoretical data, while not exceeding the experimental error in the neutralization cross section,[2] is still appreciable. Therefore a realistic choice of the wave function for $\varkappa r \lesssim 1$ is one of the main ways of improving the accuracy in calculating the cross sections for neutralization of H^- in collisions with fast, multiply charged ions. We note in this regard that the use of $\varkappa_0(r)$ in the form (3) is preferable to form (4), since the latter has clearly the wrong behavior at small r.

[1] F. Melchert, W. Debus, M. Liehr *et al.*, Europhys. Lett. **9**, 433 (1989).
[2] F. Melchert, M. Beuner, S. Krudener *et al.*, Phys. Rev. Lett. **74**, 888 (1995).
[3] D. V. Uskov, in *The Physics of Electronic and Atomic Collisions. XIX International Conference*, Whistler, Canada (1995), p. 687.
[4] Y. K. Kim and M. Inokuti, Phys. Rev. A **3**, 665 (1971).
[5] M. H. Cherkani, D. Fussen, M. I. Chibisov, and F. Broillard, Phys. Rev. A **54**, 1445 (1996).
[6] J. T. Lin, T. F. Jiang, J. Y. Kuang, and C. D. Lin, Phys. Rev. A **56**, 2020 (1997).
[7] L. P. Presnyakov and D. V. Uskov, Sov. Phys. JETP **59**, 515 (1984).
[8] L. V. Keldysh, Sov. Phys. JETP **20**, 1307 (1965).

[9] A. B. Voitkiv and V. A. Pazdzersky, J. Phys. **21**, 3369 (1988).
[10] A. B. Voĭtkiv and A. V. Koval', Zh. Tekh. Fiz. **64**(3), 188 (1994) [Tech. Phys. **39**, 335 (1994)].
[11] M. E. Goldberger and K. M. Watson, *Collision Theory* [Wiley, New York (1964); Mir, Moscow (1967)].
[12] V. Sidis, C. Kubash, and D. Fussen, Phys. Rev. A **27**, 2431 (1983).
[13] Y. Tamaguchi, Phys. Rev. **95**, 1628 (1954).
[14] L. P. Presnyakov, V. P. Shevel'ko, and R. K. Yanev, *Elementary Processes Involving Multiply Charged Ions* [in Russian], Énergoatomizdat, Moscow (1986). 200 pp.
[15] A. M. Dykhne and G. L. Yudin, Usp. Fiz. Nauk **125**, 377 (1978) [Sov. Phys. Usp. **21**, 549 (1978)].
[16] Z. Flyugge, *Practical Quantum Mechanics* [Springer-Verlag, Berlin (1971); Vol. 2, Mir, Moscow (1974)].

Translated by Steve Torstveit

Investigation of an unstable diffusion process in isothermal three-component gaseous mixtures under steady conditions

Yu. I. Zharin, V. N. Kosov, and S. A. Krasikov

Scientific-Research Institute of Experimental and Theoretical Physics, Al-Farabi Kazakh National University, 480078 Alma-Ata, Kazakhstan

(Submitted April 4, 1996; resubmitted October 6, 1998)
Zh. Tekh. Fiz. **69**, 5–9 (July 1999)

The diffusional transport of a ternary gaseous mixture of helium, Freon-12, and argon is studied under isothermal conditions by a steady-flow method. The experiments were done at atmospheric pressure and a temperature of 298 K. The data are compared with the theory. It is shown that the mass transfer of the components is influenced by convective flows.
© *1999 American Institute of Physics.* [S1063-7842(99)00207-X]

INTRODUCTION

Experimental studies of isothermal diffusion in certain ternary gaseous mixtures have shown that under certain conditions the difference in the diffusion velocities of the components can lead to stratification of the gaseous mixture according to density, which is followed in a gravitational field by macroscopic motions forming distinct structures.[1–3] The process wherein the convective flows are superposed on the molecular transport is defined as diffusionally unstable. The detailed investigation[4–7] of this effect has revealed a complicated dependence on many parameters and conditions but at the same time has made it possible to formulate the following physical model for the diffusional instability and the subsequent dynamics of the process. In the case of multicomponent diffusion, differences in the mobility of the components give rise to stratification into regions of different densities. If the change in concentration over the characteristic length of this region is small, then the mass transfer will occur at the microscopic level, and no convective flow of the gaseous mixture will be observed. However, the influence of such parameters as the pressure, temperature, viscosity, etc., on the system can lead to a decrease in the linear size of the stratified region. Then the gradients can increase to certain critical values, whereupon a sudden transition (bifurcation) to macroscopic flows occurs. The unstable regime can develop further only if the size of the ''convection cells''[1] is the same as the geometric dimensions of the diffusion channel. It is clear that in that case the conventional study of multicomponent diffusion, both in the experimental and computational/theoretical arenas, will not permit a sufficiently correct estimate of the true contribution to the total mass transfer from the convective flows in the unstable regime. Previous experiments involving the study of the diffusional instability were carried out in closed apparatus by quasisteady or unsteady methods[1–7] and could not eliminate the effect of circulation of the gaseous mixture between the upper and lower parts of the apparatus, as has been shown to exist for the mass transfer in ternary systems with a diluent gas.[8] The standard methods of studying both in the thermodynamic[9,10] and kinetic[11,12] approximations permit one to determine the boundaries of the domain of applicability of the diffusional representations with the help of stability theory,[13,14] as has been done by the authors (in analogy with Ref. 13) in the linear approximation for a plane-layer model.[15] However, the relations obtained in Ref. 15 cannot be used to obtain information about the contribution of molecular diffusion and convection to the total mass transfer. The behavior of diffusionally unstable systems under steady conditions becomes a question of fundamental importance, the answer to which will permit estimation of the magnitude of various effects accompanying mass transfer in closed apparatus.

EXPERIMENTAL RESULTS AND PROCEDURE

The experiments were carried out on the steady-flow apparatus schematically illustrated in Fig. 1. The idea of the method is to achieve conditions of binary diffusion by equalizing the total volume gas flows in opposite directions. This was achieved by adjusting the hydrodynamic drag of one of the gases at the outlet from the pipe. The procedure was analogous to that of Ref. 16. The initial mixtures of gases from bottles *1* were passed through pressure regulators to dibutylphthalate-filled manostats *2* and then to buffer tanks *3* and capillaries *4* for smoothing out pulsations, and then to heat exchangers *5*. The volume flow rates of the mixtures were measured by inclined rheometers *6*. Then the steady gas flows entered the diffusion cell *7*, which consisted of two slot channels connected by a diffusion cuvette having a height of 7.8 cm and a cross section of 5.5×1.05 cm. In the central part of the cuvette there was a 2.0×5.0 cm inspection port into which flat panes of glass were cemented; this permitted observation of the dynamics of the unstable process in a shadow device *12* (Ref. 17).

From the diffusion cell the gaseous mixtures were passed through containers for selecting the gases for analysis *8*, capillaries *9*, and fine-control valves *10* and then to soap-film flow meters *11* with a graduation of 0.1 ml/div. The velocity of the film was determined by a timer to an accuracy of ±0.1 s. The temperature was regulated at the necessary

 © 1999 American Institute of Physics

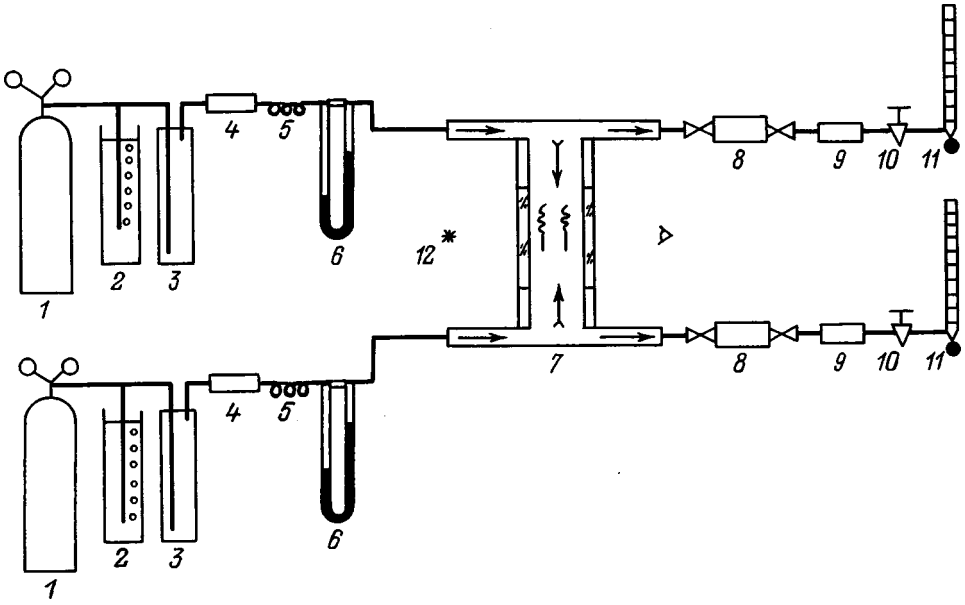

FIG. 1. Diagram of the experimental apparatus implementing the steady-flow method.

places in the apparatus to an accuracy of ± 0.1 K and equalled 298.0 K at atmospheric pressure.

The steady flows of the initial gaseous mixtures, which were maintained constant and equal in volume flow rate, passed into the upper and lower inlets of the diffusion cell, the lower-density mixture into the upper inlet and the higher-density mixture into the lower inlet. The fine-control valves were used to equalize the volume flow rates of the gases at the outlet from the cell, a condition that was monitored by means of the soap-film flow meters. Thus we modeled an binary diffusion process in which the observed volume flow rates of the gases in opposite directions were equal. The initial gas mixtures were analyzed on an ITR-1 interferometer with an error of not more than 0.1%, while the mixtures from the samplers after diffusion were analyzed on a Khrom-4 chromatograph with an error of 0.3%.

The experiments were carried out for the system helium (He) + Freon-12 ($R12$) – argon (Ar). This system was chosen because, for the given geometric dimensions of the diffusion cuvette, the unstable process occurred in it at atmospheric pressure over a wide range of Freon-12 concentrations in the binary mixture.[2,5]

In the experiment conditions were always maintained such that the density of the mixture (pure gas) in the upper cell was less than the density of the pure gas (mixture) in the lower cell. Fluctuations of atmospheric pressure during the course of the experiment did not exceed 0.4% of the value 91.9 kPa, which was taken as the mean value.

Figure 2 shows the concentrations of helium and Freon-12 that have crossed over as functions of the Freon-12 concentration in the binary mixture (the argon concentration is equal to the sum of the concentrations of helium and Freon-12): △ — He, ○ — $R12$ (experimental values); the solid curves are an approximation of the experimental data. Also shown in Fig. 1 are the concentrations (the dashed curve for He, the dotted curve for $R12$) calculated from the Stefan–Maxwell equations[18] on the assumption of a diffusion process. The following binary diffusion coefficients were used in the calculation for the gas pairs at $P=0.101$ MPa, $T=298.0$ K: $D_{\text{He-Ar}}=0.750$, $D_{\mathbf{He}-R12}=0.385$, $D_{\text{Ar}-R12}=0.067$ $\text{cm}^2\cdot\text{s}^{-1}$. The calculations were done for the equilibrium values of the mole fractions and initial differences of the concentrations of the components, referred to mean atmospheric pressure, as in the experiment. One notices that the stable process is characterized by a monotonic dependence of the diffused amount of the components on the Freon-12 concentration in the mixture. Then, beginning at a concentration of 0.04 mole fraction Freon-12 in the binary mixture in the experiments, a completely different picture is observed. When the mixture is above and the pure gas is

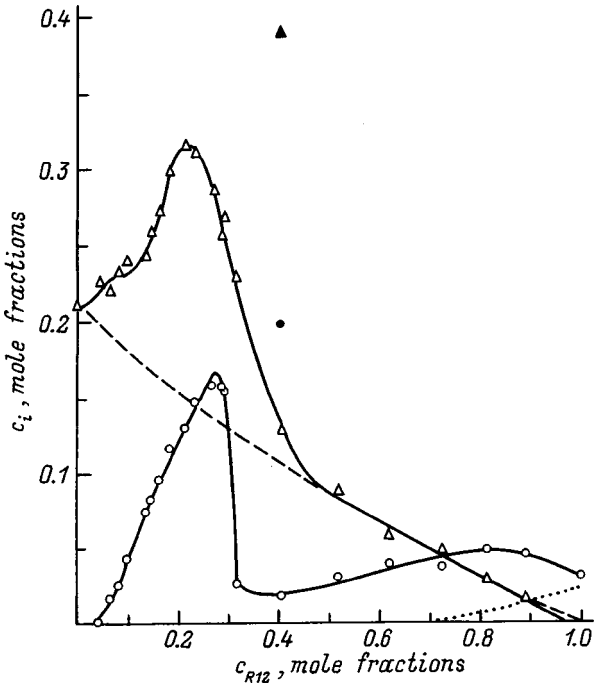

FIG. 2. Helium and Freon-12 concentrations versus the concentration of $R12$ in a binary mixture with helium, diffusing in argon at $T=298.0$ K.

below, a maximum enrichment of the pure gas in helium and freon occurs. The unstable regime is manifested most strongly for systems in which the density of the mixture is equal to or somewhat lower than the density of the pure component (argon). For the opposite orientation of the system, when the mixture flows along the lower channel of the diffusion cell and the argon flows along the upper, the influence of convection resulting from the buoyancy forces is weaker. Visual observations of the unstable regime by the technique of Ref. 17 revealed the presence of structural formations moving along the diffusion channel in opposite directions. In certain situations one can observe relatively immobile strata in the channel, which, as the Freon-12 is ''accumulated,'' transform into convective flows which lead to perturbations of the hydrodynamic type.

If conditions of hydrostatic instability are realized in the system, i.e., when a heavier mixture, say 0.6 mole fraction helium and 0.4 mole fraction Freon-12, flows in the upper part of the channel and a lighter gas (here argon) flows in the lower part, the convection mechanism loses its anomalous character. The partial flows Q_i of the components will obey the condition $Q_{\mathrm{He}}/Q_{R12}=c_{\mathrm{He}}/c_{R12}$, where c_i is the initial concentration of the ith component. The symbols ▲, ● in Fig. 2 illustrate this situation.

According to our studies, the concentration data can be separated into the following regions: 1) for a Freon-12 content in the binary mixture in the range $0<c<0.04$ mole fraction the system is diffusionally stable; 2) if the Freon-12 concentration lies in the range $0.04<c<0.5$ mole fraction, then the diffusion process is unstable independently of the orientation of the gases in the cell; 3) in the interval $0.50<c<1.00$ the convective mechanism of diffusion of the light component is quenched. This is evidenced by the experimental data and the results of a calculation for helium. A comparison with Freon-12 gives poorer agreement, a circumstance that can be attributed to the presence of a diffusion ''gate'' in such mixtures.[19]

INTERPRETATION OF THE EXPERIMENTAL RESULTS

The results attest to a significant influence of convective flows on the mass transfer in the unstable region under steady-state conditions. In this case the data obtained by the steady-flow method will differ from the data obtained in closed devices.

Let us examine this question for the transport of Freon-12 in the case when the binary mixture is above the pure gas. The Freon-12 that ''falls'' down is carried by the counterflowing argon to the outlet of the diffusion cell. It can therefore be asserted that whether the gases have crossed over from upper to lower or vice versa, there is practically no chance for them to return. The picture is completely different for the unstable process in closed systems, e.g., in a two-bulb diffusion apparatus.[5] There one also observes enrichment of the lower part of the apparatus in the heavy component, but convective flows which transport the gas from the upper part of the apparatus to the lower part encounter an upward backflow, and this mixture can also contain some amount of the heavy component. A so-called circulation of the gas arises, which is most often manifested in an unstable process in systems with a ballast gas.[8]

The domain of applicability of Fick's laws for the steady case can be estimated in the framework of stability theory, which is widely used in problems of thermal convection.[13,14] The macroscopic motion of an isothermal ternary gaseous mixture is described by the Navier–Stokes equation, which is supplemented with the equations of mass transfer of the components:[13]

$$\rho\left[\frac{\partial \mathbf{u}}{\partial t}+(\mathbf{u}\cdot\nabla)\mathbf{u}\right]=-\nabla p+\eta\nabla^2\mathbf{u}+\left(\frac{\eta}{3}+\xi\right)\nabla(\nabla\cdot\mathbf{u})+\rho\mathbf{g},$$

$$\frac{\partial \rho}{\partial t}+\nabla\cdot(\rho\mathbf{u})=0,\quad \rho\left(\frac{\partial c_i}{\partial t}+\mathbf{u}\nabla c_i\right)=\nabla\cdot\mathbf{j}_i,$$

$$\mathbf{j}_1=-\rho(D_{11}^*\nabla c_1+D_{12}^*\nabla c_2),$$

$$\mathbf{j}_2=-\rho(D_{21}^*\nabla c_1+D_{22}^*\nabla c_2), \tag{1}$$

where $\mathbf{u}$ is the velocity, ρ the density, p the pressure, η and ξ the coefficients of shear and bulk viscosity, $\mathbf{g}$ the acceleration of gravity, c_i the concentration of the ith component, $\mathbf{j}_i$ the diffusional flux density of the ith component, and D_{ij}^* the practical diffusion coefficients, which are defined in terms of the binary diffusion coefficients.

Equation (1) must be supplemented with the equation of state of the medium,

$$\rho=\rho(c_1,c_2,p),\quad T=\mathrm{const}, \tag{2}$$

which relates the thermodynamic parameters in (1).

Since the unsteady perturbations of mechanical equilibrium are small, we neglect terms quadratic in the perturbations, and, assuming that the concentrations are distributed linearly and choosing scale units of measurement (distances are measured in units of the characteristic linear dimension d of the cavity, the time in units of d^2/ν, the frequency in units of D_{22}^*/d, the concentration in the ith component in units of $A_i d$, and the pressure in units of $\rho_0\nu D_{22}^*/d^2$), for a plane vertical layer (the z axis is directed upward and perpendicular to the plane formed by the x and y axes) we can rewrite Eq. (1) in the form

$$P_{22}\frac{\partial c_1}{\partial t}-u=\tau_{11}\frac{\partial^2 c_1}{\partial x^2}+\frac{A_2}{A_1}\tau_{12}\frac{\partial^2 c_2}{\partial x^2},$$

$$P_{22}\frac{\partial c_2}{\partial t}-u=\frac{A_1}{A_2}\tau_{21}\frac{\partial^2 c_1}{\partial x^2}+\frac{\partial^2 c_2}{\partial x^2},$$

$$\frac{\partial u}{\partial t}=\frac{\partial^2 u}{\partial x^2}+(R_1\tau_{11}c_1+R_2c_2), \tag{3}$$

where $P_{ii}=\nu/D_{ii}^*$ is the Prandtl number, $R_i=g\beta_i A_i d^4/\nu D_{ii}^*$ is the Rayleigh number, $\tau_{ij}=D_{ij}^*/D_{22}^*$, $u=u_z$,

$$\beta_i=-\frac{1}{\rho_0}\left(\frac{\partial\rho}{\partial c_i}\right)_{P,T},$$

$A_i\gamma=-\nabla c_{i0}$, $\nu=\eta/\rho$, γ is a unit vector, the average quantities are denoted by a subscript 0, and the perturbed quantities are written without a subscript.

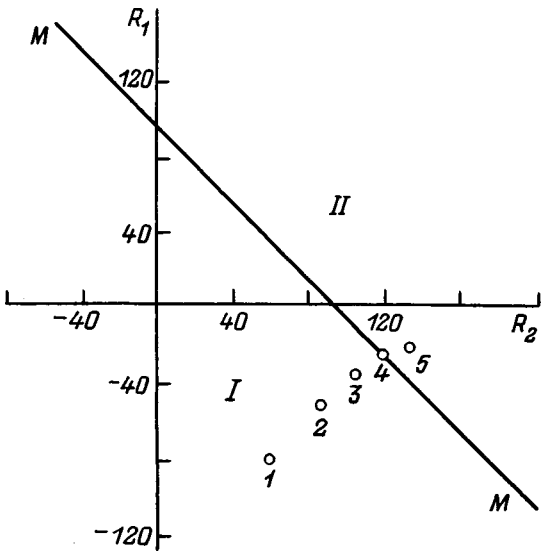

FIG. 3. Regions of diffusion *I* and diffusional instability *II*: line *MM* is the neutral line of monotonic perturbations; ○ are experimental data points.

We seek a solution of equations (3) in the form

$$\{c_1, c_2, u\} = \{c_1^0, c_2^0, u^0\} \sin\left[(n+1)\frac{\pi}{2}x\right] \exp[-\lambda t], \quad (4)$$

where $n = 1,3,5,\ldots$ are the characteristic odd modes of the perturbations.

The boundary conditions presuppose the vanishing of the velocities and the perturbations of the component concentrations c_i on the vertical planes bounding the layer of the gaseous mixture,

$$u = c_1 = c_2 = 0, \quad x = \pm 1. \quad (5)$$

Substituting (4) into the system of equations (3) and taking (5) into account, we obtain for the steady case the equation of the neutral line separating the diffusion region and the region of monotonic instability:

$$\tau_{11}\left(1 - \frac{A_2}{A_1}\tau_{12}\right)R_1 + \left(\tau_{11} - \frac{A_1}{A_2}\tau_{21}\right)R_2 = \left[(n+1)\frac{\pi}{2}\right]^4 \left[\tau_{11} - \tau_{12}\tau_{21}\left[(n+1)\frac{\pi}{2}\right]^2\right]. \quad (6)$$

Figure 3 shows a plot in the coordinates R_1, R_2 (R_1 is helium and R_2 is Freon-12) of the region of molecular transport *I* and the region of the diffusional instability *II* for $n = 1$ in the limiting case in which there are no ''cross'' effects ($\tau_{ij} = \tau_{ji} = 0$). Points *1–5* correspond to systems with the following Freon-12 concentrations in the binary mixture: *1* — 0.05, *2* — 0.15, *3* — 0.22, *4* — 0.22, *5* — 0.25 mole fraction. In this system at Freon-12 concentrations of 0.2 mole fraction in the binary mixture a transition occurs from the state of diffusion to the region of the monotonic instability; this agrees qualitatively with the experimental data. The discrepancy between the experimental data and the results calculated using Eqs. (3)–(6) is apparently due to the nonlinear distribution of the component concentrations in the diffusion channel; this happens for systems in which the binary diffusion coefficients of the components obey the condition $D_{23} \ll D_{12} \ll D_{13}$ (Ref. 19). We also note that the spectrum of critical Rayleigh numbers in the steady diffusion problem is determined by the wave numbers (modes of the perturbations) characterizing the set of convective forms of motion.[13]

CONCLUSIONS

In summary, these studies have shown that in the steady regime one can determine the existence regions of convective flows arising under the influence of the buoyancy forces. The proposed calculation technique permits estimation of the critical parameters of the transition to the regime of concentrational convection.

This study was done with the financial support of the Science Foundation of the Republic of Kazakhstan (Project No. 171-96 FN).

[1] L. Miller and E. A. Mason, Phys. Fluids **9**, 711 (1966).
[2] L. Miller, T. H. Spurling, and E. A. Mason, Phys. Fluids **10**, 1809 (1967).
[3] Yu. I. Zhavrin, N. D. Kosov *et al.*, Zh. Tekh. Fiz. **54**, 943 (1984) [Sov. Phys. Tech. Phys. **29**, 561 (1984)].
[4] Yu. I. Zhavrin, V. N. Kosov *et al.*, Inzh.-Fiz. Zh. **55**, 92 (1988).
[5] V. N. Kosov and Yu. I. Zhavrin, *Thermophysical Properties of Substances and Materials* [in Russian], Izd. Standartov, Moscow (1989), No. 28, pp. 112–122.
[6] Yu. I. Zhavrin and V. N. Kosov, Pis'ma Zh. Tekh. Fiz. **19**(10), 18 (1993) [Tech. Phys. Lett. **19**(5), 301 (1993)].
[7] Yu. I. Zhavrin, A. Z. Aĭtkozhaev, V. N. Kosov, and S. A. Krasikov, Pis'ma Zh. Tekh. Fiz. **21**(6), 7 (1995) [Tech. Phys. Lett. **21**(3), 206 (1995)].
[8] A. Z. Aĭtkozhaev, Yu. I. Zhavrin, and N. D. Kosov, Izv. AN KazSSR Ser. Fiz., No. 6, pp. 88–92 (1991).
[9] S. R. de Groot and P. Mazur, *Nonequilibrium Thermodynamics* [North-Holland, Amsterdam (1962); Mir, Moscow (1964), 456 pp.].
[10] L. S. Kotousov, *Thermodiffusion—a Method for Studying Nonideal Systems* [in Russian], Nauka, Leningrad (1973), 199 pp.
[11] J. Hirschfelder, C. F. Curtiss, and R. B. Bird, *Molecular Theory of Gases and Liquids*, [Wiley, New York (1954); IL, Moscow (1961), 931 pp.].
[12] K. M. Aref'ev, *Transport Phenomena in Gases and Plasmas*, Énergoatomizdat, Leningrad (1983), 112 pp.
[13] G. Z. Gershuni and E. M. Zhukhovitskiĭ, *Convective Instability of an Incompressible Liquid* [in Russian], Nauka, Moscow (1972), 392 pp.
[14] D. D. Joseph, *Stability of Fluid Motions* [Springer-Verlag, New York (1976); Mir, Moscow (1981), 638 pp.].
[15] Yu. I. Zhavrin, V. N. Kosov *et al.*, Inzh.-Fiz. Zh. **60**, 419 (1991).
[16] Yu. I. Zhavrin and N. D. Kosov, *Diffusion in Gases and Liquids* [in Russian], KazGU, Alma-Ata (1974), pp. 24–29.
[17] N. D. Kosov, A. G. Bychkov, Yu. I. Zhavrin, and V. N. Kosov, Teplofiz. Aeromekh. **1**, 87 (1994).
[18] Yu. I. Zhavrin, N. D. Kosov, and Z. I. Novosad, *Diffusion in Gases and Liquids* [in Russian], KazGU, Alma-Ata (1974), pp. 12–19.
[19] V. D. Seleznev and V. G. Smirnov, Zh. Tekh. Fiz. **51**, 795 (1981) [Sov. Phys. Tech. Phys. **26**, 471 (1981)].

Translated by Steve Torstveit

Critical conditions for instability of a highly charged oblate spheroidal drop

A. I. Grigor'ev and S. O. Shiryaeva

Yaroslavl State University, 150000 Yaroslavl, Russia
(Submitted February 9, 1998)
Zh. Tekh. Fiz. **69**, 10–14 (July 1999)

The spectrum of capillary oscillations of a charged oblate spheroidal drop is calculated in neglect of the interaction between modes by means of a perturbation expansion in the small deviation of the equilibrium shape of the drop from spherical. The critical conditions for instability of its nth mode with respect to the self-charge are calculated in the form of an analytical function describing how the dimensionless Rayleigh parameter characterizing the stability of the drop depends on the value of the spheroidal deformation. © *1999 American Institute of Physics.* [S1063-7842(99)00307-4]

The problem of studying capillary oscillations, stability, and self-dispersal of a highly charged drop is encountered in diverse physical problems (see, e.g., Refs. 1–3 and the literature cited therein). In this regard the case of spherical and prolate spheroidal drops have been studied repeatedly.[3] At the same time, the stability of charged oblate spheroidal drops has hardly been studied, although there have been a few published reports on this subject.[4,5] The problem of studying the stability of charged oblate spheroidal drops is of interest in connection with the physics of storm clouds, in which, according to observational data, a certain fraction of the drops have oblate spheroidal form,[6–8] and also for problems of liquid-metal epitaxy and inkjet printing, where charged drops landing on the substrate with a nonzero velocity have for a certain interval an oblate spheroidal form,[3] and also in connection with the problem of St. Elmo's fire, which is due to the instability in an external electric field of water drops that have precipitated on the objects around which St. Elmo's fire is observed.[9,10] In Ref. 4, where a charged oblate spheroidal drop was investigated for stability against axisymmetric deformations, it was found that such a drop is stable. Moreover, it was asserted[4] that the stability of a charged spheroidal drop against its self-charge increases with increasing degree of oblateness (with increasing eccentricity), and it was concluded that the concentration of such drops in clouds should be high. In Ref. 5 it was shown that the conclusion of Ref. 4 that oblate charged drops have a heightened stability was reached prematurely and that including possible nonaxisymmetric deformations destroys the stability of such drops. Since the calculations of Ref. 5 were of a preliminary, qualitative nature and their correctness can be disputed, it is advisable to do them more rigorously.

1. Let us consider the problem of stability against nonaxisymmetric disturbances of the surface of a highly charged oblate spheroidal drop of an ideal incompressible, perfectly conducting liquid immersed in an ideal incompressible nonconducting medium. We shall assume that the spheroidal shape of the drop is due to ''outside'' forces of a nonelectrical nature, e.g., forces of acoustic (ultrasonic) pressure $P_a = P_0 Y_{20}(\cos\Theta)$, as implemented in Ref. 11, where P_0 is a constant and Y_{20} is the axisymmetric spherical harmonic $Y_{20}(\cos\Theta) \equiv \mathfrak{P}_2(\cos\Theta)$, where $\mathfrak{P}_2$ is a normalized Legendre polynomial. We assume that the drop carries a charge Q and that its volume is that of a spherical drop of radius R. We will denote the coefficient of surface tension of the interface by σ and the densities of the drop and medium by ρ_1 and ρ_2, respectively. We shall solve the problem by the method applied previously in Refs. 12 and 13.

The equation of the surface of an oblate spheroidal perturbed by thermal capillary wave motion in the linear approximation in the square of the eccentricity is written as follows in a spherical coordinate system with origin at the center of the drop:

$$r = r(\Theta) + \xi(\Theta,\varphi,t) \approx R[1 - e^2 \cdot h(\Theta) R^{-1} + \xi(\Theta,\varphi,t) E^{-1}];$$

$$r(\Theta) = \frac{R(1-e^2)^{1/3}}{(1-e^2\sin^2\Theta)^{1/2}}; \quad h(\Theta) = \frac{R}{6}(3\cos^2\Theta - 1). \tag{1}$$

Here $e = (1 - a^2/b^2)^{1/2}$ is the eccentricity of the spheroid, a and b are its semiminor and semimajor axes, $\xi(\Theta,\varphi,t)$ is the perturbation of the equilibrium spheroidal surface of the drop due to capillary oscillations occurring on account of the thermal motion of the molecules and having an amplitude $\sim\sqrt{kT/\sigma}$ (k is Boltzmann's constant and T is the absolute temperature). We also note that for the majority of liquids the amplitude of such thermal capillary oscillations is of the order of tenths of a nanometer.

Under the stated conditions the wave motions in the drop and surrounding medium will be of a potential character, with velocity potentials Ψ_1 and Ψ_2, respectively, which for an incompressible liquid are harmonic functions:[14]

$$\nabla^2 \Psi_i = 0 \quad (i = 1,2),$$

$$r \to 0: \quad \Psi_1(\mathbf{r},t) \to 0;$$

$$r \to \infty: \quad \Psi_2(\mathbf{r},t) \to 0 \tag{2}$$

and at the interface (for $r = r(\Theta) + \xi(\Theta,\varphi,t)$) the following boundary conditions are satisfied: equality of the velocity components normal to the boundary,

 © 1999 American Institute of Physics

$$\frac{\partial \Psi_1}{\partial n_1} = -\frac{\partial \Psi_2}{\partial n_2} \equiv \frac{\partial \Psi}{\partial n}, \tag{3}$$

the kinematic boundary condition

$$\frac{\partial \xi}{\partial t} \approx \frac{\partial \Psi}{\partial n}, \tag{4}$$

and the dynamical boundary condition

$$\Delta P = -\rho_1 \frac{\partial \Psi_1}{\partial t} + \rho_2 \frac{\partial \Psi_2}{\partial t} + P_E - R_\sigma - P_a, \tag{5}$$

where ΔP is the pressure differential between the drop and medium, P_E and P_0 are the pressures exerted by the electric field and surface tension, respectively, and $\mathbf{n}_1$ and $\mathbf{n}_2$ are the outward and inward normals to the surface of the drop ($\mathbf{n} \equiv \mathbf{n}_1 = -\mathbf{n}_2$).

To determine the electrical potential Φ produced by the perturbed charged drop in the surrounding space, which should be a harmonic function, we have the boundary-value problem[15]

$$\nabla^2 \Phi = 0, \tag{6}$$

$$r \to \infty: \quad \Phi \to 0, \tag{7}$$

$$r = r(\Theta) + \xi: \quad \Phi = \text{const}. \tag{8}$$

In the exposition below, in the linear approximation in $|\xi|/R$, all the derivatives in the boundary conditions (3)–(5) will be referred to the unperturbed surface of the drop $r = r(\Theta)$, as is standard procedure in the theory of waves of infinitesimal amplitude.[14]

We also require that the volume of the drop remain constant,

$$\int_v dV = \frac{4}{3}\pi R^3, \tag{9}$$

and that its center of mass does not move,

$$\int_v \mathbf{r} dV = 0. \tag{10}$$

In Eqs. (9) and (10) the integration is over the volume of the drop.

2. The subsequent analysis is carried out in the framework of perturbation theory by means of an expansion in the small parameters e^2 and ξ to terms of order $\sim e^2$, ξ, and $e^2\xi$, i.e., in the linear approximation in each of these. We note that the small parameters e^2 and ξ are independent, and that $e^2 \gg \xi$. For this reason it might seem that if we are keeping terms of order $\sim e^2\xi$, we should also take the term $\sim e^4$ into account. Although this is a completely valid conclusion, we shall see below that the contribution to the dispersion relation for the capillary oscillations of a charged spheroidal drop has only terms $\sim \xi$ and $e^4\xi$, the terms $\sim e^2, e^4$ vanishing when when the kinematic boundary condition is imposed (it contains a partial derivative with respect to time). Since this dispersion relation is what we are after, it would serve no purpose to complicate the mathematical formulas by keeping terms $\sim e^4$ below.

The time dependence of the velocity field $\mathbf{U}(\mathbf{r},t)$, pressure field $p(\mathbf{U},t)$, and perturbation of the equilibrium surface $\xi(\Theta,\varphi,t)$ will be assumed exponential, $\sim \exp(st)$, where s is a complex frequency.

The harmonic solutions of equations (2) for velocity potentials $\Psi_j(\mathbf{r},t)$ in a spherical coordinate system are naturally sought in the form of series in the normalized spherical harmonics $Y_{km}(\Theta,\varphi)$:

$$\Psi_1(\mathbf{r},t) = \sum_{k=0}^{\infty} \sum_{m=-k}^{k} A_{km} r^k Y_{km}(\Theta,\varphi)\exp(st),$$

$$\Psi_2(\mathbf{r},t) = \sum_{k=0}^{\infty} \sum_{m=-k}^{k} B_{km} r^{-k-1} Y_{km}(\Theta,\varphi)\exp(st). \tag{11}$$

Since the perturbation of the equilibrium surface of the drop $\xi(\Theta,\varphi,t)$ is related to the velocity potentials $\Psi_i(\mathbf{r},t)$ by the kinematic boundary condition (4), the function $\xi(\Theta,\varphi,t)$ should also be written as an expansion in spherical harmonics:

$$\xi(\Theta,\varphi,t) = \sum_{k=0}^{\infty} \sum_{m=-k}^{k} Z_{km} Y_{km}(\Theta,\varphi)\exp(st). \tag{12}$$

The coefficients A_{km}, B_{km}, and Z_{km} in expansions (11) and (12) are related to one another by the boundary conditions (3) and (4).

Conditions (9) and (10) impose certain restrictions on the form of the thermal perturbation $\xi(\Theta,\varphi,t)$ of the equilibrium surface of the liquid; the essence of these can be elucidated by substituting expressions (1) and (12) into (9) and (10). We note that in our statement of the problem the function $\xi(\Theta,\varphi,t)$ is independent and completely unrelated to the equilibrium spheroidal shape of the drop; in taking the integrals (9) and (10) we can therefore ignore the ''cross'' terms $\sim e^2\xi$, $e^4\xi$, etc. As a result, the integrals (9) and (10) in the linear approximation in e^2 and ξ can be reduced to the form

$$\int_0^{2\pi}\int_0^{\pi} \xi(\Theta,\varphi,t)\sin\Theta d\Theta d\varphi = 0; \tag{13}$$

$$\int_0^{2\pi}\int_0^{\pi} \xi(\Theta,\varphi,t) Y_{1m}(\Theta,\varphi)\sin\Theta d\Theta d\varphi = 0;$$

$$m = 0, \pm 1. \tag{14}$$

By virtue of the orthonormality of the spherical harmonics, it follows from (13) and (14) that $Z_{1m} = 0$ ($m = 0, \pm 1$), $Z_{00} = 0$. This means that the summation over k in (12) begins with $k = 2$. By virtue of of the boundary conditions (3) and (4), this also holds for the series (11).

To use the boundary conditions (3), one must first write out an expression for the vector normal, which in the general case is given by the relation

$$\mathbf{n} = \frac{\nabla F}{|\nabla F|}; \quad F = r - [R - e^2 h(\Theta) + \xi(\Theta,\varphi,t)].$$

In the approximation adopted here, this relation has the form

$$\mathbf{n}\approx\left[1+e^2\frac{1}{R^2}\frac{\partial h(\Theta)}{\partial\Theta}\frac{\partial\xi(\Theta,\varphi,t)}{\partial\Theta}\right]\mathbf{e}_r+\frac{1}{R}\left[-\frac{\partial\xi(\theta,\varphi,t)}{\partial\Theta}+e^2\frac{\partial h(\Theta)}{\partial\Theta}-e^2\frac{1}{R}\frac{\partial[h(\Theta)\xi(\Theta,\varphi,t)]}{\partial\Theta}\right]\mathbf{e}_\Theta-\frac{1}{R\sin\Theta}\frac{\partial\xi(\Theta,\varphi,t)}{\partial\varphi}\left[1+e^2\frac{h(\Theta)}{R}\right]\mathbf{e}_\varphi, \tag{15}$$

where $\mathbf{e}_r$, $\mathbf{e}_\Theta$, $\mathbf{e}_\varphi$ are the unit vectors of the spherical coordinate system.

Substituting expansion (11) into the boundary condition (3) and taking expression (15) into account, we obtain after straightforward but awkward transformations the following relation between the coefficients A_{km} and B_{km} in expansions (11):

$$B_{km}\approx-\frac{1}{k+1}\left\{\left[k-e^2\left(\alpha_k^1+\frac{k}{k+1}\beta_k^1\right)\varkappa_k^m\right]R^{2k+1}A_{km}-e^2\left[\alpha_k^2+\frac{k-2}{k-1}\beta_k^2\right]\gamma_{k-1}^m R^{2k-1}A_{k-2,m}-e^2\times\left[\alpha_k^3+\frac{k+2}{k+1}\beta_k^3\right]\gamma_{k+1}^m R^{2k+3}A_{k+2,m}\right\}, \tag{16}$$

where we have used the notation $\alpha_k^1=k^2-k-3$, $\alpha_k^2=0.5(k^2-3k+2)$, $\alpha_k^3=0.5(k^2+k-4)$; $\beta_k^1=k^2+3k-1$, $\beta_k^2=0.5(k^2+k-4)$, $\beta_k^3=0.5(k^2+5k+6)$,

$$\varkappa_k^m=\frac{k+k-3m}{3(2k-1)(2k+3)},$$

$$\gamma_k^m=\frac{1}{2k+1}\left[\frac{(k^2-m^2)[(k+1)^2-m^2]}{(2k-1)(2k+3)}\right]^{1/2}. \tag{17}$$

The coefficients A_{km} and B_{km} are related to the expansion coefficients Z_{km} in Eq. (12) by the kinematic boundary condition (4). Substituting (11) and (12) into (4) and using Eq. (16), we get

$$A_{km}\approx R^{1-k}k^{-1}\{[1+e^2k^{-1}\alpha_k^1\varkappa_k^m]sZ_{km}+e^2\times(k-2)^{-1}\alpha_k^2\gamma_{k-1}^m sZ_{k-2,m}+e^2\times(k+2)^{-1}\alpha_k^3\gamma_{k+1}^m sZ_{k+2,m}\};$$

$$B_{km}\approx-R^{2+k}(k+1)^{-1}\{[1-e^2(k+1)^{-1}\beta_k^1\varkappa_k^m]sZ_{km}-e^2(k-1)^{-1}\beta_k^2\gamma_{k-1}^m sZ_{k-2,m}-e^2\times(k+3)^{-1}\beta_k^3\gamma_{k+1}^m sZ_{k+2,m}\}. \tag{18}$$

3. In order to use the dynamical boundary condition (5) one must write out expressions for the electrical pressure P_E and the pressure of surface tension P_σ. The electrical potential Φ near the surface of the drop is determined by the solution of the boundary-value problem (6)–(8), which is conveniently sought in the form

$$\Phi=\Phi_0+\Phi_1+\Phi_2+\Phi_3, \tag{19}$$

where Φ_0 is the potential obtained in zeroth order in the small parameters used in the expansion, Φ_1 is the correction $\sim e^2$, Φ_2 is the correction $\sim\xi$, and Φ_3 is the correction $\sim\xi e^2$.

Substituting (19) into (6)–(8), we straightforwardly obtain the following set of boundary-value problems for finding Φ_j, where $j=0,1,2,3$:

$$\nabla^2\Phi_j=0,$$

$$r\to\infty:\quad \Phi_j\to 0;$$

$$r=R:\quad \Phi_0=Q/R;\quad \Phi_1=e^2h(\Theta)\frac{\partial\Phi_0}{\partial r};$$

$$\Phi_2=-\xi(\Theta,\varphi,t)\frac{\partial\Phi_0}{\partial r};$$

$$\Phi_3=e^2h(\Theta)\xi(\Theta,\varphi,t)\frac{\partial^2\Phi_0}{\partial r^2}-\xi(\Theta,\varphi,t)\frac{\partial\Phi_1}{\partial r}+e^2h(\Theta)\frac{\partial\Phi_2}{\partial r}. \tag{20}$$

The potentials Φ_j will be sought in the form of series in spherical harmonics:

$$\Phi_j(\mathbf{r},t)=\sum_{k=0}\sum_{m=-k}C_{km}^j(t)\cdot r^{-k-1}\cdot Y_{km}(\Theta,\varphi). \tag{21}$$

Imposing successively the boundary conditions of different orders at the surface of the drop, we obtain the following expression for the potential in the vicinity of the drop:

$$\Phi(\mathbf{r},t)\approx\frac{Q}{R}\left\{\frac{R}{r}-e^2h(\Theta)\frac{R^3}{r^3}+\frac{1}{R}\times\sum_{k=0}^{\infty}\sum_{m=-k}^{k}Z_{km}\left(\frac{R}{r}\right)^{k+1}Y_{km}(\Theta,\varphi)\exp(st)-e^2\frac{1}{R}\sum_{k=0}^{\infty}\sum_{m=-k}^{k}\left[(k+2)\varkappa_k^m Z_{km}+\frac{k}{2}\gamma_{k-1}^m Z_{k-2,m}+\frac{1}{2}(k+4)\gamma_{k+1}^m Z_{k+2,m}\right]\times\left(\frac{R}{r}\right)^{k+1}Y_{km}(\Theta,\varphi)\exp(st)\right\}. \tag{22}$$

Using the solution of equation (22), we can write the electrical pressure to the adopted accuracy as an expansion in the small parameters,

$$P_E=\frac{1}{8\pi}(\mathbf{E})^2=\frac{1}{8\pi}(\nabla\Phi(\mathbf{r},t))^2\approx P_E^0+P_E(\xi);$$

$$P_E^0\approx\frac{Q^2}{8\pi R^4}\left[1-e^2\frac{2h}{R}\right];$$

$$P_E(\xi)\approx\frac{Q^2}{4\pi R^4}\sum_{k=0}^{\infty}\sum_{m=-k}^{k}\left\{[(k-1)-e^2\times(k-4)\varkappa_k^m]\frac{Z_{km}}{R}e^2\frac{1}{2R}(k-5)\gamma_{k-1}^m Z_{k-2,m}-e^2\frac{1}{2R}(k-7)\gamma_{k+1}^m Z_{k+2,m}\right\}\times Y_{km}(\Theta,\varphi)\exp(st). \quad (23)$$

Here P_E^0 is the electrical pressure on the spheroidal surface of the drop, undistorted by the capillary motion, and $P_E(\xi)$ is the correction to it due to surface perturbation $\xi(\Theta,\varphi,t)$. The pressure of surface tension P_σ beneath the distorted surface of the liquid in the general case has the form[16]

$$P_\sigma=\sigma(\nabla\cdot\mathbf{n}),$$

where the unit vector normal to the perturbed spheroidal surface is represented by expression (15).

As a result, we can write the Laplace pressure as an expansion in small quantities:

$$P_\sigma=P_\sigma^0+P_\sigma(\xi);$$

$$P_\sigma^0=\frac{2\sigma}{R}\left[1-e^2\frac{2h}{R}\right];$$

$$P_\sigma(\xi)\approx\frac{\sigma}{R^2}\sum_{k=0}^{\infty}\sum_{m=-k}^{k}\{[(k-1)(k+2)+e^2 2(k^2+k+4)\varkappa_k^m]Z_{km}+e^2(k^2-3k+6)\gamma_{k-1}^m Z_{k-2,m}+e^2(k^2+5k+10)\gamma_{k+1}^m Z_{k+2,m}\}\times Y_{km}(\Theta,\varphi)\exp(st). \quad (24)$$

In our statement of the problem we postulated that the equilibrium shape of the drop was spheroidal; accordingly, in the dynamical boundary condition (5) the sum of the terms of zeroth order and the terms of first order in e^2 goes to zero, leaving the equilibrium shape. Then the sum of the terms linear in ξ, which are due to the deformation of the equilibrium surface of the drop by thermal capillary motion, also vanishes, i.e., Eq. (5) gives

$$\rho_1\frac{\partial\Psi_1}{\partial t}-\rho_2\frac{\partial\Psi_2}{\partial t}-P_E(\xi)+P_\sigma(\xi)=0.$$

Substituting the solutions (11) into this relation and using relations (18), (23), and (24) and the property of orthogonality of the spherical harmonics, we arrive at an infinite system of coupled equations for the coefficients Z_{km}:

$$\left\{\Omega^2\left[\frac{1}{k}\left(1-e^2\frac{k+3}{k}\varkappa_k^m\right)+\eta\frac{1}{k+1}\left(1-e^2\frac{k-2}{k+1}\varkappa_k^m\right)\right]+[(k+2)(k-1)+e^2 2(k^2+k+4)\varkappa_k^m]-4W[(k-1)-e^2(k-4)\varkappa_k^m]\right\}\frac{Z_{km}}{R}-e^2\frac{1}{2}\gamma_{k-1}^m\times\left\{\Omega^2\left[\frac{1}{k}+e^2\frac{k-3}{k^2-1}\eta\right]-2(k^2-3k+6)-4W(k-5)\right\}\times\frac{Z_{k-2,m}}{R}-e^2\frac{1}{2}\gamma_{k+1}^m\left\{\Omega^2\left[\frac{k+4}{k(k+2)}+\frac{1}{k+1}\eta\right]-2(k^2+5k+10)-4W(k-7)\right\}\frac{Z_{k+2,m}}{R}=0, \quad (25)$$

where we have introduced the dimensionless parameters

$$\Omega^2=s^2\frac{\rho_1 R^3}{\sigma},\quad \eta=\frac{\rho_2}{\rho_1},\quad W=\frac{Q^2}{16\pi\sigma R^3}.$$

The equations of system (25) have an every-other type of coupling, and the system therefore actually decomposes into two subsystems for the even and odd capillary modes, and it has nontrivial solutions when the determinant of its coefficients vanishes. This condition yields the dispersion relation of the problem, which is also of infinite order. However, in the linear approximation in the square of the eccentricity $\sim e^2$, when one can neglect the interaction of the modes (which appears only in the approximation $\sim e^4$), the system of equations (25) reduces to a system of uncoupled equations for the individual modes, and the dispersion relation can be written in the relatively simple form

$$\Omega^2\left[\frac{1}{k}\left(1-e^2\frac{k+3}{k}\varkappa_k^m\right)+\eta\frac{1}{k+1}\left(1-e^2\frac{k-2}{k+1}\varkappa_k^m\right)\right]+[(k+2)(k-1)+e^2 2(k^2+k+4)\varkappa_k^m]-4W[(k-1)-e^2(k-4)\varkappa_k^m]=0. \quad (26)$$

As the parameter W is increased and passes through a certain critical value, the square of the complex frequency passes through zero and becomes positive; this corresponds to the appearance of two solutions varying exponentially in time, one of which is experimentally decaying and the other exponentially growing, i.e., the system becomes unstable. Thus, assuming in (26) that $\Omega^2=0$, we obtain the following relation, which can be used to find the critical value of the parameter W separating the stable and unstable solutions:

$$W_*\approx\frac{k+2}{4}\left[1+e^2\frac{k^2[k(k+1)-3m^2]}{(k-1)(k+2)(2k-1)(2k+3)}\right]$$
$$(k\geqslant 2; m=0;\pm 1;\ldots\pm k). \quad (27)$$

It should be recalled that the conditions that the volume of the drop be constant (9) and that its center of gravity not move (10) forbid the excitation of modes with $k=0$ and $k=1$; the minimum value of the index k in (27) is therefore $k=2$.

It follows from (27) that for all the axisymmetric modes ($m=0$) the critical values of the parameter W increase with increasing oblateness of the drop (with increasing e^2). For the nonaxisymmetric modes the value of W can be increasing or decreasing, depending on the value of the index m. For example, for $k=2$ the coefficient of e^2 in (27) is positive for $m=0,\pm 1$ and negative for $m=\pm 2$. This means that a highly charged oblate spheroidal drop is stable against axisymmet-

ric capillary oscillations and also against nonaxisymmetric oscillations with $m=\pm 1$ but is unstable with respect to nonaxisymmetric oscillations with $m=\pm 2$.

4. Let us conclude with the observation that a similar problem can also be formulated for prolate spheroidal drops, the stability of which against axisymmetric deformations has not yet been investigated. The equation of the unperturbed surface of a prolate spheroidal drop in a spherical coordinate system is

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

Here $e=(1-b^2/a^2)^{1/2}$ is the eccentricity of the spheroid, and b and a are its semiminor and semimajor axes. Since the entire analysis was done in the approximation linear in e^2, the equation of the free surface of the drop distorted by thermal capillary wave motion is of the form

$$r=r(\Theta)+\xi(\Theta,\varphi,t)\approx R[1+e^2h(\Theta)R^{-1}+\xi(\Theta,\varphi,t)R^{-1}],$$

i.e., differs from the analogous expression for an oblate spheroidal drop only in the sign of the terms $\sim e^2$. All of the results obtained above can therefore easily be carried over to the case of a prolate spheroid — one must simply invert the signs of the terms $\sim e^2$. Ultimately, the expression determining the critical value of the parameter W is

$$W_*\approx\frac{k+2}{4}\left[1-e^2\frac{k^2[k(k+1)-3m^2]}{(k-1)(k+2)(2k-1)(2k+3)}\right]$$
$$(k\geqslant 2;m=0;\pm 1;\ldots\pm k). \qquad (28)$$

It follows from (28) that for all the axisymmetric modes $(m=0)$, increasing the eccentricity of the drop leads to a lowering of the critical conditions for instability with respect to the self-charge. For the nonaxisymmetric modes $(m\neq 0)$ the situation in regard to the stability is determined by the value of the index m. For example, for the fundamental mode $(k=2)$ the critical value of the parameter W increases with increasing e^2 in the case $m=2$, while for $m=1$ it decreases, i.e., the picture is opposite to that for an oblate spherical drop.

[1] A. G. Baily, Sci. Prog. Oxf. **61**, 555 (1974).
[2] V. I. Kozhenkov and N. A. Fuks, Usp. Khim. **45**, 2274 (1976).
[3] A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk. Mekh. Zhidk. Gazov, No. 3, pp. 3–20 (1994).
[4] O. A. Basaran and L. E. Scriven, Phys. Fluids A **1**, 795 (1989).
[5] A. I. Grigor'ev, A. A. Forstov, and S. O. Shiryaeva, in *Proceedings of the Ninth International Conference on Atmospheric Electricity*, St. Petersburg, Russia (1992), pp. 450–453.
[6] D. N. A. Jones, J. Meteorology **16**, 504 (1959).
[7] E. L. Ausman and M. Brook, J. Geophys. Res. **72**, 6131 (1967).
[8] H. R. Pruppacher and R. L. Piter, J. Atmos. Sci. **28**, 86 (1971).
[9] I. D. Grigor'eva and S. O. Shiryaeva, Zh. Tekh. Fiz. **64**(9), 202 (1994) [Tech. Phys. **39**, 973 (1994)].
[10] A. I. Grigor'ev, I. D. Grigor'eva, and S. O. Shiryaeva, J. Sci. Exploration **5**, 163 (1991).
[11] E. H. Trinh, Rev. Sci. Instrum. **56**, 2059 (1985).
[12] S. O. Shiryaeva and A. I. Grigor'ev, Zh. Tekh. Fiz. **65**(2), 1 (1995) [Tech. Phys. **40**, 122 (1995)].
[13] S. O. Shiryaeva and A. I. Grigor'ev, Zh. Tekh. Fiz. **66**(9), 12 (1996) [Tech. Phys. **41**, 865 (1996)].
[14] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* [2nd ed., Pergamon Press, Oxford (1986); 3rd ed., Nauka, Moscow (1986), 733 pp.].
[15] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* [2nd ed. with L. P. Pitaevskii, Pergamon Press, Oxford (1984); Nauka, Moscow (1982), 620 pp.].
[16] R. E. Rosentsweig, *Ferrohydrodynamics* [Cambridge Univ. Press, New York (1985); Mir, Moscow (1989), 357 pp.].

Translated by Steve Torstveit

Some observations on the appearance of instability of a plane charged liquid surface

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

Yaroslavl State University, 150000 Yaroslavl, Russia
(Submitted March 3, 1998)
Zh. Tekh. Fiz. **69**, 15–22 (July 1999)

A dispersion relation is obtained for the capillary oscillations of a hemispherical protrusion (oblate or prolate) on the plane surface of a conducting liquid in a uniform electrostatic field parallel to the symmetry axis of the protrusion. For the fundamental mode of the capillary oscillations realized on the protrusion the critical dependence of the parameter characterizing its stability in an external electrostatic field is obtained as a function of the square of the eccentricity as the protrusion is drawn out from the an oblate to a prolate hemisphere. Such a change in shape lowers the threshold electric field for instability of the protrusion.
© *1999 American Institute of Physics.* [S1063-7842(99)00407-9]

When an infinite vacuum interface of a perfectly conducting liquid of density ρ and surface tension σ is subjected to an electrostatic field $\mathbf{E}_0$ perpendicular to the free surface of the liquid, it is known[1–6] that the interface is unstable with respect to perturbations of the form

$$\zeta=\zeta_0 \exp(St-kx) \tag{1}$$

(S is the complex frequency, t is the time, k is the wave number, and x is the horizontal coordinate) if

$$W_{\mathrm{TF}}\geqslant \frac{1}{ak}+ak, \quad a=\sqrt{\frac{\sigma}{\rho g}},$$

$$W_{\mathrm{TF}}=\frac{E_0^2}{4\pi(\rho g\sigma)^{1/2}}\equiv\frac{E_0^2 a}{4\pi\sigma}. \tag{2}$$

Here a is the capillary constant, g is the acceleration of free fall, and W_{TF} is the dimensionless Tonks–Frenkel parameter, which characterizes the stability of the surface of the liquid against the charge induced on it by the field $\mathbf{E}$.

The lowest value of the electric field $\mathbf{E}_0$ for which the most easily excited mode with $k=1/a$ becomes unstable, i.e., for which the Tonks–Frenkel instability is realized, is determined from the critical value of the Tonsk–Frenkel parameter $W_{\mathrm{TF}}=2$ (Refs. 1 and 2). Perturbations of the plane surface of the liquid in the form (1) exist if only because of the thermal motion of the molecules, in which case $\zeta_0\approx\sqrt{kT/\sigma}$, where k is Boltzmann's constant and T is the absolute temperature, and if the surface charge (the field $\mathbf{E}$) reaches a high enough value that condition (2) holds, the amplitudes of some of the waves of the form (1) will begin to grow exponentially in time.

The above statements are valid in the framework of the linear theory of instability ($\zeta_0\ll 1/k$), which predicts the exponential growth of the amplitude of the perturbation to values at which the amplitude ζ of the perturbation becomes comparable to the wavelength $1/k$, i.e., $\zeta<1/k$; for $\zeta\gtrsim 1/k$ one can no longer be guided by the results of this theory. At the present time our understanding of the nonlinear stage of development of the Tonks–Frenkel instability is unclear. Nevertheless, it has been shown theoretically[1–3] and experimentally[7] that the nonlinear stage of the Tonks–Frenkel instability results in the formation of the so-called Taylor cones, which are prolate hemispherical protrusions on the liquid surface with a conical tip from which highly charged droplets are emitted, carrying away the excess charge.[8–10] The mechanism of formation of Taylor cones has been the subject of many studies.[3,7,11–14] Nevertheless, by virtue of of the nonlinearity of the problem, correct results have been obtained only for the initial and final stages of formation of these protrusions. The main regularities in the growth of emitting protrusions are known only on the qualitative level of the first paper by Tonks,[1] although the problem itself is of significant interest in connection with numerous academic, technical, and technological applications.

To reach reasoned conclusions about the intermediate stage of evolution of Taylor cones, it is useful to consider the question of stability of hemispherical protrusions (oblate or prolate) on a plane surface of a conducting liquid in an electrostatic field $\mathbf{E}$ perpendicular to the surface. By solving this problem and noting the behavior of the critical value of the electrostatic field for instability of the protrusions as a function of the value of the spheroidal deformation, one can obtain information about the evolution of emitting protrusions forming upon the onset of instability, when in the course of their growth they pass through a hemispherical shape.

Since we are interested specifically in the critical conditions for instability of hemispherical protrusions of different eccentricities, we set aside the question of the mass flow into the growing protrusion and consider a model problem of the stability of a hemispherical drop with a given eccentricity, lying on an electrically conductive solid substrate in an electrostatic field applied perpendicular to the substrate. This statement of the problem may be of interest for treating St. Elmo's fire,[9,10] for explaining the sharp enhancement of the energy loss in electrical transmission lines during rainstorms,[15] and for analysis of explosion and fire safety issues in the storage of flammable liquids.[16]

1. Consider a drop of incompressible, perfectly conducting liquid with density ρ and surface of surface tension σ,

 © 1999 American Institute of Physics

lying on a plane conducting substrate. Let the whole system be located in an external uniform electrostatic field perpendicular to the plane of the substrate. We assume that the drop has the shape of an oblate axisymmetric hemispheroid and that the equation of the free surface in a spherical coordinate system with origin at the center of the base of the drop is

$$r(\Theta)=\frac{R(1-e_1^2)^{1/3}}{(1-e_1^2\sin^2\Theta)^{1/2}},\quad e_1^2=1-\frac{d^2}{c^2},$$

where d and c are the semiminor and semimajor axes of the spheroid, with d along the symmetry axis, and R is the radius of a hemisphere of the same volume as the drop.

Let $\xi(\Theta,t)$ be a perturbation of the equilibrium hemispheroidal shape of the drop, due to the thermal motion of the molecules of the liquid. We consider the problem of capillary oscillations and stability of the surface of such a drop. We shall derive the solution in an approximation linear in the perturbation of the hemispheroidal shape of the drop $|\xi|/R$ and in the square of the eccentricity e_1^2 with accuracy to terms $\sim e_1^2|\xi|/R$. In this approximation the equation of the surface of the drop is written in the form

$$r=r(\Theta)+\xi\approx R\left[1-e_1^2\frac{h(\Theta)}{R}+\frac{\xi}{R}\right];$$

$$\frac{h(\Theta)}{R}\equiv\frac{1}{6}(3\cos^2\Theta-1). \tag{3}$$

When the drop has the shape of a prolate hemispheroid, the equation of its free surface becomes

$$r(\Theta)=\frac{R(1-e_2^2)^{1/6}}{(1-e_2^2\cos^2\Theta)^{1/2}};\quad e_2^2=1-\frac{c^2}{d^2},$$

where d and c are the semimajor and semiminor axes of the spheroid, with d along the symmetry axis.

The expansion of this expression in powers of e_2^2 differs from (3) only in the sign of the coefficients of e_1^2 and e_2^2:

$$r=r(\Theta)+\xi\approx R\left[1+e_2^2\frac{h(\Theta)}{R}+\frac{\xi}{R}\right];$$

$$\frac{h(\Theta)}{R}\equiv\frac{1}{6}(3\cos^2\Theta-1). \tag{3a}$$

We shall henceforth regard e^2 as some formal real parameter that is not required to be the nonnegative square of the eccentricity, and we shall solve the problem simultaneously for both the oblate hemispheroidal shape, with a square of the eccentricity $e_1^2\equiv-e^2$ (when $e^2<0$), and for the prolate hemispheroid, with a square of the eccentricity $e_2^2\equiv e^2$ (if $e^2>0$).

Since the liquid is assumed ideal, the wave motion in the drop will be potential motion. A complete mathematical formulation of the problem of calculating the capillary oscillations of such a drop will have the form

$$\nabla^2\psi=0; \tag{4}$$

$$r\to 0:\quad \psi\to 0; \tag{5}$$

$$r=r(\Theta)+\xi,\quad 0\leqslant\Theta<\frac{\pi}{2}:\ \frac{\partial\xi}{\partial t}\approx(\mathbf{n}\cdot\boldsymbol{\nabla})\psi; \tag{6}$$

$$\Delta\mathbf{F}-\rho\frac{\partial\psi}{\partial t}+\mathfrak{F}_E=\mathfrak{F}_\sigma; \tag{7}$$

$$\nabla^2\Phi=0; \tag{8}$$

$$\mathbf{r}\to\infty:\quad \Phi\to\Phi_\infty=-E_0z=-E_0r\cos\Theta; \tag{9}$$

$$r=r(\Theta)+\xi;\quad \Theta=\frac{\pi}{2}:\quad \Phi=\text{const}=0. \tag{10}$$

Here $\mathbf{n}$ is the unit vector normal to the surface of the drop, $\psi(\mathbf{r},t)$ is the velocity potential, Φ is the electrical potential, $\Delta\mathbf{F}$ is the pressure differential inside and outside the drop, $\mathfrak{F}_E$ is the pressure exerted by the electric field, and $\mathfrak{F}_\sigma$ is the pressure of surface tension.

2. We represent the perturbation $\xi(\Theta,t)$ of the hemispheroidal surface of the drop in the form of a series in Legendre polynomials:

$$\xi(\Theta,t)=\sum_{j=0}^{\infty}Z_j(t)P_j(\cos\Theta), \tag{11}$$

where $Z_j(t)$ are the time-dependent amplitudes of the individual capillary modes of the drop.

In the oscillations of the surface of the drop, its volume remains unchanged:

$$\int_0^{2\pi}d\varphi\int_0^{\pi/2}\sin\Theta\,d\Theta\int_0^{r(\Theta)+\xi}r^2dr-\frac{2\pi}{3}R^3$$
$$\approx\sum_{j=0}^{\infty}\frac{Z_j(t)}{R}3\int_0^1P_j(x)dx=0.$$

Here we have neglected terms $\sim e^2\xi$ on account of the fact that the spheroidal deformation of the hemispheroid is independent of the perturbation of the free surface of the liquid. Legendre polynomials on a hemisphere obey the integral relations[17]

$$\int_0^1P_m(x)dx=\begin{cases}1, & \text{if } m=0;\\ 0, & \text{if } m=2n;\\ A_m, & \text{if } m=2n+1;\end{cases}$$

$$\int_0^1P_n(x)P_m(x)dx$$
$$=\begin{cases}\dfrac{1}{2n+2}, & \text{if } m=n;\\ 0, & \text{if } n-m=2n\neq 0;\\ B_{nm}, & \text{if } m=2n+1 \text{ and } n \text{ is even};\end{cases}$$

$$A_j\equiv\frac{(-1)^{2m}(2m)!}{2^{2m+1}m!(m+1)!};$$

$$B_{nm}\equiv\frac{(-1)^{\frac{n+2j}{2}}(2m+1)!\,n!}{2^{n+2m}(n-2m-1)(n+2m+2)\left[\left(\frac{n}{2}\right)!\,m!\right]^2}. \tag{12}$$

With (12) taken into account, the condition of constant volume of the drop becomes

$$3\frac{Z_0(t)}{R}+\sum_{j=0}^{\infty}3A\frac{Z_{2j+1}(t)}{R}=0,$$

and consequently,

$$Z_0=Z_{2j+1}=0;\quad (j=0,1,2,\ldots), \tag{13}$$

i.e., the zeroth mode and all the odd modes are not excited. Therefore, expression (11) can be rewritten as follows:

$$\xi(\Theta,t)=\sum_{j=1}^{\infty}Z_{2j}(t)\,P_{2j}(\cos\Theta). \tag{14}$$

The solution of equation (4) which satisfies the boundary condition (5) has the following form in a spherical coordinate system:

$$\psi(\mathbf{r},t)=\sum_{m=0}^{\infty}C_m(t)r^mP_m(\cos\Theta). \tag{15}$$

Here $C_m(t)$ is the time-dependent amplitude of the potential motions of the liquid due to the different modes. We write the unit vector normal to the surface of the perturbed hemispheroid in the approximation linear in e^2 and $|\xi|/R$ as

$$\mathbf{n}\approx\left[1-e^2\frac{1}{R^2}\frac{\partial h(\Theta)}{\partial\Theta}\frac{\partial\xi(\Theta,t)}{\partial\Theta}\right]\mathbf{e}_r-\frac{1}{R}\left[e^2\frac{\partial h(\Theta)}{\partial\Theta}+\frac{\partial\xi(\Theta,t)}{\partial\Theta}-e^2\frac{1}{R}\frac{\partial}{\partial\Theta}(h(\Theta)\xi(\Theta,t))\right]\mathbf{e}_\Theta, \tag{16}$$

where $\mathbf{e}_r$ and $\mathbf{e}_\Theta$ are unit vectors of the spherical coordinate system.

The kinematic boundary condition (6) with allowance for expression (16) has the form

$$r\approx R\left[1+e^2\frac{h(\Theta)}{R}\right]:\quad \frac{\partial\xi}{\partial t}\approx\frac{\partial\psi}{\partial r}-e^2\frac{1}{R}-\frac{\partial h(\Theta)}{\partial\Theta}\frac{1}{r}-\frac{\partial\psi}{\partial\Theta}. \tag{17}$$

We assume that

$$Z_m(t)=Z_m\exp(St). \tag{18}$$

Substituting (14) and (15) into (17), we reduce the kinematic boundary condition to the form

$$\sum_{j=1}^{\infty}SZ_{2j}(t)P_{2j}(\cos\Theta)\approx\sum_{m=0}^{\infty}C_m(t)R^{(m-1)}\left[mP_m(\cos\Theta)+e^2m(m-1)\frac{h(\Theta)}{R}P_m(\cos\Theta)-e^2\frac{1}{R}\frac{\partial h(\Theta)}{\partial\Theta}\frac{\partial P_m(\cos\Theta)}{\partial\Theta}\right]. \tag{19}$$

In the calculations below we shall use the recurrent expansions for the Legendre polynomials:[17]

$$\cos\Theta P(\cos\Theta)=a_mP_{m+1}(\cos\Theta)+b_mP_{m-1}(\cos\Theta);$$

$$\frac{h(\Theta)}{R}P_m(\cos\Theta)=\frac{1}{2}\mu_mP_{m+2}(\cos\Theta)+\varkappa_mP_m(\cos\Theta)+\frac{1}{2}\zeta_mP_{m-2}(\cos\Theta);$$

$$\sin\Theta\frac{dP_m(\cos\Theta)}{d\theta}=ma_m[P_{m+1}(\cos\Theta)-P_{m-1}(\cos\Theta)];$$

$$\frac{1}{R}\frac{\partial h(\Theta)}{\partial\Theta}\frac{dP_m(\cos\Theta)}{d\Theta}=-m\mu_mP_{m+2}(\cos\Theta)+3\varkappa_mP_m(\cos\Theta)+(m+1)\zeta_mP_{m-2}(\cos\Theta);$$

$$a_m\equiv\frac{(m+1)}{(2m+1)};\quad b_m\equiv\frac{m}{(2m+1)};\quad \mu_m\equiv a_ma_{m+1};$$

$$\varkappa_m\equiv\frac{m(m+1)}{3(2m-1)(2m+3)};\quad \zeta_m\equiv b_{m-1}b_m. \tag{20}$$

Using (20), we transform (19) to

$$\sum_{m=0}^{\infty}\left\{SZ_m(t)\delta_{m,2j}-[m+e^2(m(m-1)-3)\varkappa_m]C_m(t)R^{(m-1)}-e^2\frac{1}{2}(m-1)(m-2)\mu_{m-2}C_{m-2}(t)R^{(m-3)}-e^2\frac{1}{2}[(m+2)(m+1)-2(m+3)]\zeta_{m+2}C_{m+2}(t)R^{(m+1)}\right\}P_m(\cos\Theta)=0.$$

By virtue of the orthogonality of the Legendre polynomials, this equation holds if the coefficients of the different P_m all vanish. As a result, we obtain an infinite system of linear coupled equations for the coefficients $C_m(t)$. For $m=2j$, where $j=1,2,3,4,\ldots$ the system will be inhomogeneous, and, solving it by the method of successive approximations, we straightforwardly find

$$C_{2j}(t)\approx\frac{1}{2jR^{(2j-1)}}\left\{SZ_{2j}(t)-e^2\left[\frac{2j(2j-1)-3}{2j}\varkappa_{2j}SZ_{2j}(t)+\frac{(j-1)(2j-1)}{2(j-1)}\mu_{2j-2}SZ_{2j-2}(t)+\frac{(j+1)(2j+1)-(2j+3)}{2(j+1)}\zeta_{2j+2}SZ_{2j+2}(t)\right]\right\}. \tag{21}$$

For $m=2j+1$, where $j=0,1,2,\ldots$, we obtain a homogeneous system of equations with a nonzero determinant. Such a system has only the trivial solution

$$C_{2j+1}(t)=0 \quad (j=0,1,2,\ldots). \tag{22}$$

Let us write out the terms in Eq. (7). The pressure exerted by the capillary forces on an arbitrary surface is determined by the well-known relation

$$\mathfrak{F}_\sigma=\sigma(\nabla\cdot\mathbf{n}), \tag{23}$$

where $\mathbf{n}$ is a unit vector normal to the surface.

For the surface (3) we find, using (16) and (14), that

$$\mathfrak{F}_\sigma\approx 2\frac{\sigma}{R}\left[1+e^2 2\frac{h(\Theta)}{R}\right]+\frac{\sigma}{R}\sum_{j=1}^{\infty}\left[(2j-1)(2j+2)-2\beta_{2j}e^2\frac{h(\Theta)}{R}\right]\frac{Z_{2j}(t)}{R}P_{2j}(\cos\Theta);$$

$$\beta_m\equiv m^2+m+4. \tag{24}$$

Using relations (20), we can reduce this expression to the form

$$\mathfrak{F}_\sigma\approx\mathfrak{F}_\sigma^{(0)}+\delta\mathfrak{F}_\sigma(\xi)=2\frac{\sigma}{R}\left[P_0(\cos\Theta)+e^2\frac{2}{3}P_2(\cos\Theta)\right]+\frac{\sigma}{R}\sum_{j=2}^{\infty}\left\{[(2j-1)(2j+2)-e^2 2\varkappa_{2j}\beta_{2j}]\frac{Z_{2j}(t)}{R}-e^2\beta_{2j-2}\mu_{2j-2}\frac{Z_{2j-2}(t)}{R}-e^2\beta_{2j+2}\zeta_{2j+2}\frac{Z_{2j+2}(t)}{R}\right\}P_{2j}(\cos\Theta), \tag{25}$$

where $\mathfrak{F}_\sigma^{(0)}$ is the pressure exerted by capillary forces on the unperturbed surface of a drop containing terms of zero order of smallness and terms $\sim e^2$; $\delta\mathfrak{F}_\sigma(\xi)$ is the correction to the capillary pressure due to the perturbation $\xi(\Theta,t)$ and contains terms $\sim\xi$ and $\sim e^2\xi$.

3. In order to write an expression for the pressure exerted by the electric field on the surface of a drop we must solve the boundary-value problem (8)–(10). We write the electrical potentials in the form of an expansion,

$$\Phi=\Phi_0+\Phi_1+\Phi_2+\Phi_3, \tag{26}$$

where Φ_0 is the zeroth order potential, Φ_1 is the correction of order e^2, Φ_2 is the correction of order ξ, and Φ_3 is the correction of order $e^2\xi$.

Substituting (26) into (8)–(10), we obtain four boundary-value problems for each of the functions Φ ($q=1,0,2,3,$)

$$\nabla^2\Phi_q=0 \quad (q=0,1,2,3), \tag{27}$$

$$\text{at} \quad r\to\infty: \quad \Phi_0\to -E_0 r\cos\Theta,$$

$$\Phi_q\to 0 \quad (q=1,2,3); \tag{28}$$

$$\text{at} \quad \Theta=\frac{\pi}{2}: \quad \Phi_q\to 0 \quad (q=0,1,2,3); \tag{29}$$

$$\text{at} \quad r=R: \quad \Phi_0=0,$$

$$\Phi_1=-e^2h(\Theta)\frac{\partial\Phi_0}{\partial r}, \quad \Phi_2=-\xi(\Theta,t)\frac{\partial\Phi_0}{\partial r},$$

$$\Phi_3=-e^2h(\Theta)\xi(\Theta,t)\frac{\partial^2\Phi_0}{\partial^2 r}-\xi(\Theta,t)\frac{\partial\Phi_1}{\partial r}e^2h(\Theta)\frac{\partial\Phi_2}{\partial r}. \tag{30}$$

The solutions of the above problems of finding potentials Φ_q which satisfy conditions (28) and (29) should be sought in the form

$$\Phi_0(r,\Theta)=\sum_{m=0}^{\infty}[A_m^{(0)}r^m+D_m^{(0)}r^{-(m+1)}]P_m(\cos\Theta);$$

$$\Phi_q(\mathbf{r})=\sum_{j=0}^{\infty}D_{2j+1}^{(q)}(t)\left(\frac{R}{r}\right)^{2j+2}P_{2j+1}(\cos\Theta)$$

$$(q=1,2,3). \tag{31}$$

Substituting these equations and relation (14) into the boundary conditions (28)–(30) and taking (20) into account, we obtain expressions for the unknown coefficients $A_m^{(j)}$ and $D_m^{(j)}$:

$$A_1^{(0)}=-E_0; \quad A_m^{(0)}=0(\forall m\neq 1);$$

$$D_1^{(0)}=E_0R^3; \quad D_m^{(0)}=0(\forall m\neq 1);$$

$$D_1^{(1)}=\frac{2}{5}e^2E_0R; \quad D_3^{(1)}=\frac{3}{5}e^2E_0R;$$

$$D_{2m+1}^{(1)}=0 \quad (m=2,3,4,\ldots);$$

$$D_{2j+1}^{(2)}=3E_0[a_{2j}Z_{2j}(t)+b_{2j+2}Z_{2j+2}(t)];$$

$$D_{2j+1}^{(3)}=e^2 3E_0\Bigg(a_{2j}\left[-\frac{4}{15}+2(j+2)\varkappa_{2j+1}+(j+1)a_{2j-1}b_{2j}\right]Z_{2j}(t)+b_{2j+2}\times\left[-\frac{4}{15}+2(j+2)\varkappa_{2j+1}+(j+3)a_{2j+2}b_{2j+3}\right]\times Z_{2j+2}(t)+(j+1)a_{2j-2}\mu_{2j-1}Z_{2j-2}(t)+(j+3)b_{2j+4}\zeta_{2j+3}Z_{2j+4}(t)\Bigg);$$

$$(j=0,1,2,\ldots). \tag{32}$$

The coefficients a_m, b_m, ζ_m, and μ_m are defined in (20).

We can now write out expressions for the electric field and the pressure exerted by it on the free surface of the drop, $r=R+e^2h(\Theta)+\xi(\Theta,t)$:

$$\mathbf{E}=-\nabla\cdot\Phi\approx\left\{E_0\cos\Theta\left[3-e^2 2\left(\frac{2}{5}-3\frac{h(\Theta)}{R}\right)\right]\right.$$
$$-6E_0\cos\Theta\left[1-e^2 2\left(\frac{7}{15}-3\frac{h(\Theta)}{R}\right)\right]\frac{\xi(\Theta,t)}{R}$$
$$+\sum_{j=0}^{\infty}(2j+2)\left[\left(1-e^2(2j+3)\frac{h(\Theta)}{R}\right)\frac{D_{2j+1}^{(2)}(t)}{R}\right.$$
$$\left.\left.+\frac{D_{2j+1}^{(3)}(t)}{R}\right]P_{2j+1}(\cos\Theta)\right\}\mathbf{e}_r$$
$$+\left\{e^2E_0\sin\Theta\left(1+6\frac{h(\Theta)}{R}\right)-3E_0\sin\right.$$
$$\left.\times\Theta\left[1+e^2\left(\frac{7}{5}+11\frac{h(\Theta)}{R}\right)\right]\frac{\xi(\Theta,t)}{R}\right.$$
$$-\sum_{j=0}^{\infty}\left[\left(1-e^2(2j+3)\frac{h(\Theta)}{R}\right)\frac{D_{2j+1}^{(2)}(t)}{R}\right.$$
$$\left.\left.+\frac{D_{2j+1}^{(3)}(t)}{R}\right]\frac{dP_{2j+1}(\cos\Theta)}{d\Theta}\right\}\mathbf{e}_\Theta.$$

$$\mathfrak{F}_E=\frac{E^2}{8\pi}\approx\frac{9E_0^2}{8\pi}\cos^2\Theta\left[1-e^2\left(\frac{8}{15}-4\frac{h(\Theta)}{R}\right)\right]$$
$$-\frac{1}{4\pi}\left\{6E_0^2\left[1+6\frac{h(\Theta)}{R}\right]\frac{\xi(\Theta,t)}{R}-e^2 6E_0^2\right.$$
$$\times\left[\frac{13}{15}-\frac{9}{15}\frac{h(\Theta)}{R}-42\left(\frac{h(\Theta)}{R}\right)^2\right]\frac{\xi(\Theta,t)}{R}$$
$$-3E_0\cos\Theta\sum_{j=0}^{\infty}(2j+2)\left[\left(1-e^2\left(\frac{4}{15}+(2j\right.\right.\right.$$
$$\left.\left.\left.+1)\frac{h(\Theta)}{R}\right)\right)\frac{D_{2j+1}^{(2)}(t)}{R}+\frac{D_{2j+1}^{(3)}(t)}{R}\right]P_{2j+1}(\cos\Theta)$$
$$+e^2E_0\sin\Theta\left[1+6\frac{h(\Theta)}{R}\right]$$
$$\left.\times\sum_{j=0}^{\infty}\frac{D_{2j+1}^{(2)}(t)}{R}\frac{dP_{2j+1}(\cos\Theta)}{d\Theta}\right\}.$$

Using (14), (20), and the expressions (32) for $D_{2j+1}^{(2)}(t)/R$ and $D_{2j+1}^{(3)}(t)/R$ and performing some straightforward but extremely awkward manipulations, we arrive at an expression for the electrical pressure on the surface of a hemispheroidal drop in the form of an expansion in Legendre polynomials:

$$\mathfrak{F}_E\approx\mathfrak{F}^{(0)}+\delta\mathfrak{F}_e(\xi)=\frac{3E_0^2}{8\pi}\left\{P_0(\cos\Theta)\right.$$
$$\left.+2\left[1+e^2\frac{18}{35}\right]P_2(\cos\Theta)+e^2\frac{48}{35}P_4(\cos\Theta)\right\}$$
$$+\frac{9E_0^2}{8\pi}\sum_{j=1}^{\infty}\left\{e^2(3j-7)\mu_{2j-4}\mu_{2j-2}\frac{Z_{2j-4}(t)}{R}\right.$$
$$+\mu_{2j-2}[(2j-2)+e^2f_j^{(1)}]\frac{Z_{2j-2}(t)}{R}$$
$$+[(2j-2)\gamma_{2j-1}+2j\gamma_{2j}+e^2f_j^{(2)}]\frac{Z_{2j}(t)}{R}$$
$$+\zeta_{2j+2}[2j+e^2f_j^{(3)}]\frac{Z_{2j+2}(t)}{R}$$
$$\left.+e^2 3(j-1)\zeta_{2j+2}\zeta_{2j+4}\frac{Z_{2j+4}(t)}{R}\right\}P_{2j}(\cos\Theta);$$

$$\gamma_m=a_m b_{m+1};$$
$$f_m^{(1)}\equiv(3m-7)\gamma_{2m-3}+(5m-7)\gamma_{2m-2}+(m-6)\gamma_{2m-1}$$
$$+(3m-4)\gamma_{2m}-\frac{1}{15}(31m-61);$$
$$f_m^{(2)}\equiv\gamma_{2m-1}\left[(m-6)\gamma_{2m-1}+(5m-7)\gamma_{2m-2}\right.$$
$$\left.-\frac{1}{15}(31m-61)\right]+\gamma_{2m}\left[(m-5)\gamma_{2m+1}\right.$$
$$\left.+(5m-2)\gamma_{2m}-\frac{1}{15}(31m-30)\right]$$
$$+2(3m-5)\gamma_{2m-1}\gamma_{2m};$$
$$f_m^{(3)}\equiv 3(m-2)\gamma_{2m-1}+(5m-2)\gamma_{2m}+(m-5)\gamma_{2m+1}$$
$$+3(m-1)\gamma_{2m+2}-\frac{1}{15}(31m-30). \qquad (33)$$

4. Let us update the hydrodynamic term in the dynamical boundary condition (7) with the use of the solution (15) and the coefficients (21) and (22), the functional form (18), and relations (20). As a result, we find that the following equations hold on the free surface of the drop:

$$\rho\frac{\partial\Psi}{\partial t}\approx\rho S\sum_{j=1}^{\infty}C_{2j}(t)R^{2j}\left[1+e^2 2j\frac{h(\Theta)}{R}\right]P_{2j}(\cos\Theta);$$
$$\rho\frac{\partial}{\partial t}\approx\rho R^2S^2\sum_{j=1}^{\infty}\frac{1}{2j}\left[\left(1+e^2\frac{(2j+3)}{2j}\varkappa_{2j}\right)\frac{Z_{2j}(t)}{R}\right.$$
$$+e^2\frac{1}{2}\mu_{2j-2}\frac{Z_{2j-2}(t)}{R}$$
$$\left.+e^2\frac{j+2}{2j+2}\zeta_{2j+2}\frac{Z_{2j+2}(t)}{R}\right]P_{2j}(\cos\Theta). \qquad (34)$$

From the dynamical boundary condition (7) on the unperturbed surface of the drop we obtain the equation

$$\Delta\mathbf{F}+\mathfrak{F}_E^{(0)}=\mathfrak{F}_\sigma^{(0)},$$

which to zeroth order gives for the pressure differential across the unperturbed surface

$$\Delta\mathbf{F}=2\frac{\sigma}{R}-\frac{3E_0^2}{8\pi},$$

and which to first order in e^2 gives the familiar relation[18,19] between the magnitude of the electric field and the square of the eccentricity of the drop, e^2:

$$e^2 \approx \frac{9}{16\pi}\frac{E_0^2 R}{\sigma}.$$

On the unperturbed surface of the drop the dynamical boundary condition (7) is written

$$-\rho\frac{\partial\psi}{\partial t}+\delta\mathfrak{F}_E(\xi)=\delta\mathfrak{F}_\sigma(\xi).$$

Substituting into this equation expressions (25), (33), and (34) and making use of the orthogonality of the Legendre polynomials, we obtain an infinite system of linear homogeneous coupled equations for the unknown amplitudes $Z_{2j}(t)/R$:

$$\begin{aligned}
&\left\{-\Omega^2\frac{1}{2j}\left[1+e^2\frac{(2j+3)}{2j}\varkappa_{2j}\right]+W[(2j-2)\gamma_{2j-1}\right.\\
&+2j\gamma_{2j}+e^2 f_j^{(2)}]-[(2j-1)(2j+2)\\
&\left.-e^2 2\varkappa_{2j}\beta_{2j}]\right\}\frac{Z_{2j}(t)}{R}+\left\{-\Omega^2 e^2\frac{(j+2)}{4j(j+1)}\right.\\
&\left.+W[2j+e^2 f_j^{(3)}]+e^2\beta_{2j+2}\right\}\zeta_{2j+2}\frac{Z_{2j+2}(t)}{R}\\
&+\left\{-\Omega^2 e^2\frac{1}{4j}+W[(2j-2)+e^2 f_j^{(1)}]+e^2\beta_{2j-1}\right\}\\
&\times\mu_{2j-2}\frac{Z_{2j-2}(t)}{R}+\{e^2 W 3(j-1)\zeta_{2j+2}\zeta_{2j+4}\}\\
&\times\frac{Z_{2j+4}(t)}{R}+\{e^2 W(3j-7)\mu_{2j-4}\mu_{2j-2}\}\frac{Z_{2j-4}(t)}{R}=0;
\end{aligned}$$

$$\Omega^2\equiv\frac{\rho R^3}{\sigma}S^2;\quad W\equiv\frac{9}{4\pi}\frac{E_0^2 R}{\sigma}. \tag{35}$$

5. System (35) has a nontrivial solution if the determinant of the coefficients of the $Z_j(t)$ equals zero. This condition yields the dispersion relation of the problem:

$$\det(\alpha_{mn})=0;$$

$$\begin{aligned}
\alpha_{mm}&=-\Omega^2\frac{1}{2m}\left[1+e^2\frac{2m+3}{2m}\varkappa_{2m}\right]\\
&+W[(2m-2)\gamma_{2m-1}+2m\gamma_{2m}+e^2 f_m^{(2)}]\\
&-[(2m-1)(2m+2)-e^2 2\varkappa_{2m}\beta_{2m}];
\end{aligned}$$

$$\begin{aligned}
\alpha_{m\,m+1}&=\left\{-\Omega^2 e^2\frac{(m+2)}{4m(m+1)}+W[2m+e^2 f_m^{(3)}]\right.\\
&\left.+e^2\beta_{2m+2}\right\}\zeta_{2m+2};
\end{aligned}$$

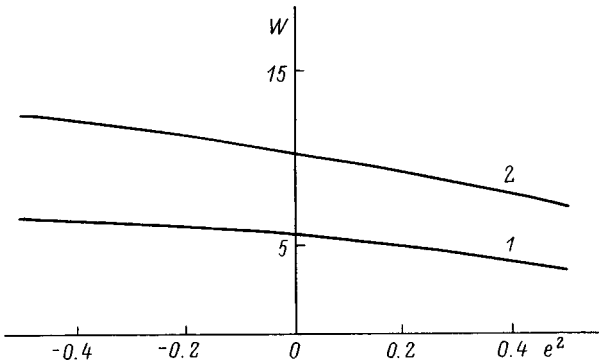

FIG. 1. Critical value of the Taylor parameter versus the square of the eccentricity of a hemispheroidal protrusion on a plane free surface of a conducting liquid.

$$\begin{aligned}
\alpha_{m\,m-1}&=\left\{-\Omega^2 e^2\frac{(1)}{4m}+W[(2m-2)+e^2 f_m^{(1)}]\right.\\
&\left.+e^2\beta_{2m-2}\right\}\mu_{2m-2};
\end{aligned}$$

$$\alpha_{m\,m+2}=3(m-1)e^2 W\zeta_{2m+2}\zeta_{2m+4};$$

$$\alpha_{m\,m-2}=(3m-7)e^2 W\mu_{2m-4}\mu_{2m-2};$$

$$\alpha_{m\,m+h}=0\quad(m=1,2,3,\ldots;\ h=\pm 3,\pm 4,\ldots). \tag{36}$$

Expressions for a_m, b_m, ζ_m, and μ_m are given in (20), and for γ_m and $f_m^{(i)}$ in (33). In a system described by the dispersion relation (36), unstable solutions, exponentially increasing in time, appear when the free coefficient of the system (36) goes to zero, which corresponds to the appearance of a zero solution for Ω^2. Therefore, setting $\Omega^2=0$ in (36), we obtain an equation of infinite order relating the critical value of the parameter W for the onset of instability of the mth mode and the deformation parameter e^2. The equation thus obtained is conveniently solved by the method of successive approximations, where the determinant of infinite order is successively approximated by determinants of increasing finite order.[20]

The equation relating W and e^2 has, according to (14), an infinite number of solutions for different even capillary modes of the hemispheroidal drop. We recall that the hemispheroidal deformation is characterized by a real parameter e^2: in the region $e^2<0$ it is equal to the square of the eccentricity of the oblate hemispheroid taken with a minus sign, $e^2=0$ corresponds to a regular hemisphere, and $e^2>0$ is the square of the eccentricity of a prolate hemispheroid.

Figure 1 shows the dependence of the critical value of W for the second (branch *1*) and fourth (branch *2*) modes calculated in the approximation of a 7th-order determinant. The calculations show that in this approximation the accuracy in the determination of the critical values of W for the fundamental mode is determined by the machine precision. From the nature of the behavior of curves *1* and *2* one can say that the critical conditions for the onset of the instability studied here tend to decrease as the hemispheroidal protrusion is drawn up from an oblate to a prolate shape. The results show, in particular, that for drops that have precipitated on

electrical transmission lines, on tall objects during storms, or on the walls of fuel tanks, the critical conditions for the onset of electrostatic instability are considerably lower than the critical conditions for the onset of the Tonks–Frenkel instability.

6. As we have said above, the critical conditions for instability of a charged plane surface obtain for $W_{\mathrm{TF}}>2$. This means that for any specific electrically conducting liquid in a uniform electrostatic field **E** normal to the plane of the free surface of the liquid, as this field is increased to a value determined from the instability criterion [see Eq. (2)]

$$E_{\mathrm{TF}}^{0}\geqslant\sqrt{8\pi\sqrt{\rho g\sigma}}, \tag{37}$$

the charged free surface of the liquid will lose stability. In other words, this happens when the attractive forces exerted by the electric field $\mathbf{E}_0$ on the charges induced on the free surface of the drop by the field $\mathbf{E}_0$ itself, which tend to increase the area of the free surface of the liquid, exceed the capillary forces that tend to diminish the area of the free surface. As a result, virtual perturbations of the free surface of the liquid due to the thermal motion of its molecules, which in the stable state have a physically infinitesimal amplitude, begin to grow exponentially in time with a growth rate that is determined by the amplitude of the virtual perturbation at the time of the onset of instability (see, e.g., Ref. 21). As the amplitude of the growing perturbation increases, so should the growth rate of the instability (or, in other words, the rate of increase of the amplitude), i.e., there is positive feedback. Some of the behavioral regularities of this phenomenon can be investigated using relations (36).

First, we see from relations (36) and Fig. 1 that as the amplitude of the virtual hemispheroidal perturbation of the free surface increases, the critical value of the parameter W_* for it decreases; this means that the critical value of the electric field E_* at which the hemispheroidal protrusion becomes unstable is lowered. Since the instability of the initial plane surface began at a value of the field E_{TF}^{0} determined by relation (37) and which remains unchanged during the entire development of the instability, we have $E_*<E_{\mathrm{TF}}^{0}$. In other words, the initial field E_{TF}^{0} will be above the critical field for the hemispheroidal protrusion, and the rate of growth of the parameter in time (its exponential growth rate) is determined by the difference $W_{\mathrm{TF}}^{0}-W_*$ at any point in time.

Using relations (36) in the approximation linear in e^2 and ignoring the interaction of the modes, we can write out an expression for the exponential growth rate of a hemispheroidal protrusion when its eccentricity is zero, $e^2=0$, i.e., when the shape of the protrusion as it changes from an oblate to a prolate hemispheroid, passes through a hemispherical shape.

If the interaction of the modes is neglected, the dispersion relation (36) decomposes into a system of mutually uncoupled dispersion relations for the individual wave numbers. The most unstable is the mode with $k=1/a$, and we shall therefore limit consideration below to only the equation of system (36) with $m=1$,

$$-\Omega^2\frac{1}{2}\left[1+e^2\frac{5}{2}\varkappa_2\right]+W[2\gamma_2+e^2f_1^{(2)}]-[4-2e^2\varkappa_2\beta_2]=0. \tag{38}$$

If it is assumed in (38) that $\Omega^2=0$, it is straightforward to find, for $e^2=\mathrm{const}$, the critical value $W=W_*$ at which a hemispheroidal protrusion having a base with a characteristic linear dimension a becomes unstable:

$$W_*=\frac{4-2e^2\varkappa_2\beta_2}{2\gamma_2+e^2f_1^{(2)}}. \tag{39}$$

We now recall that the onset of instability itself began at $W=4.5W_{\mathrm{TF}}^{0}>W_*$. Substituting this value, i.e., $4.5W_{\mathrm{TF}}^{0}$, in place of W in (38), we obtain the following expression for the instability growth rate $\delta\equiv\mathrm{Re}\Omega>0$ of a hemispheroidal protrusion on a charged plane surface of a liquid subjected to the electric field determined by (37):

$$\delta\equiv\sqrt{9W_{\mathrm{TF}}^{0}[2\gamma_2+e^2(f_1^{(2)}-5\gamma_2\varkappa_2)]-4[2-e^2\varkappa_2(\beta_2+5)]}.$$

It is easy to see that the instability growth rate δ increases with increasing e^2.

CONCLUSION

The critical conditions for the onset of instability of a hemispheroidal drop of conducting liquid on an electrically conducting solid substrate in a uniform external field **E** (a hemispheroidal protrusion on the free surface of an electrically conducting liquid undergoing the Tonks–Frenkel instability) are lowered as the eccentricity of the spheroid increases and are significantly lower than the critical conditions for the onset of instability of a plane free surface of the liquid in the field **E**. This circumstance provides a physical basis for the formation of Taylor cones and indicates that the growth rate of the Taylor cone upon the onset of the Tonks–Frenkel instability will have a faster-than-exponential growth.

[1] L. Tonks, Phys. Rev. **48**, 562 (1935).
[2] Ya. I. Frenkel', Zh. Éksp. Teor. Fiz. **6**, 348 (1936).
[3] A. I. Grigor'ev, O. A. Grigor'ev, and S. O. Shiryaeva, Zh. Tekh. Fiz. **62**(9), 12 (1992) [Sov. Phys. Tech. Phys. **37**, 904 (1992)].
[4] S. O. Shiryaeva, O. A. Grigor'ev, M. I. Munichev, and A. I. Grigor'ev, Zh. Tekh. Fiz. **65**(11), 41 (1995) [Tech. Phys. **40**, 1117 (1995)].
[5] D. F. Belonozhko, A. I. Grigor'ev, M. I. Munichev, and S. O. Shiryaeva, Pis'ma Zh. Tekh. Fiz. **22**(10), 84 (1996) [Tech. Phys. Lett. **22**(5), 425 (1996)].
[6] A. I. Grigor'ev, S. O. Shiryaeva, V. A. Koromyslov, and D. F. Belonozhko, Zh. Tekh. Fiz. **67**(8), 27 (1997) [Tech. Phys. **42**, 877 (1997)].
[7] M. D. Gabovich, Usp. Fiz. Nauk **140**, 137 (1983) [Sov. Phys. Usp. **26**, 447 (1983)].

[8] A. I. Grigor'ev and S. O. Shiryaeva, J. Phys. D **23**, 1361 (1989).
[9] A. I. Grigor'ev, I. D. Grigor'eva, and S. O. Shiryaeva, J. Sci. Exploration **5**, 163 (1991).
[10] I. D. Grigor'eva and S. O. Shiryaeva, Zh. Tekh. Fiz. **64**(9), 202 (1994) [Tech. Phys. **39**, 973 (1994)].
[11] N. M. Miskovsky, M. Chang, P. H. Cutler *et al.*, J. Vac. Sci. Technol. A **6**, 2992 (1988).
[12] M. D. Gabovich and V. Ya. Poritskiĭ, Zh. Éksp. Teor. Fiz. **85**, 146 (1983) [Sov. Phys. JETP **58**, 86 (1983)].
[13] G. I. Taylor, Proc. R. Soc. London, Ser. A **280**, 383 (1964).
[14] S. I. Shevchenko, Zh. Tekh. Fiz. **60**(2), 54 (1990) [Sov. Phys. Tech. Phys. **35**, 167 (1990)].
[15] E. M. Bazelyan, B. N. Gorin, and V. I. Levitov, *Physical and Engineering Principles of Lightning Protection* [in Russian], Gidrometeoizdat, Leningrad (1978), 224 pp.
[16] B. G. Popov, V. N. Verevkin, V. A. Bondar', and V. I. Gorshkov, *Static Electricity in the Chemical Industry* [in Russian], Khimiya, Leningrad (1977), 238 pp.
[17] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum* [in Russian], Nauka, Leningrad (1975), 439 pp.
[18] Ch. T. O'Konski and H. C. Thacher, J. Phys. Chem. **57**, 955 (1953).
[19] 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)].
[20] A. I. Grigor'ev, Zh. Tekh. Fiz. **55**, 1272 (1985) [Sov. Phys. Tech. Phys. **30**, 736 (1985)].
[21] S. O. Shiryaeva, A. I. Grigor'ev, and I. D. Grigor'eva, Zh. Tekh. Fiz. **65**(9), 39 (1995) [Tech. Phys. **40**, 885 (1995)].

Translated by Steve Torstveit

Critical equilibrium spheroidal deformation of a drop of dielectric liquid in a uniform electrostatic field

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

Yaroslavl State University, 150000 Yaroslavl, Russia
(Submitted March 3, 1998)
Zh. Tekh. Fiz. **69**, 23–28 (July 1999)

The stability of a dielectric drop, which in an external electrostatic field takes on the equilibrium shape of a prolate spheroid, is analyzed using the principle of minimum total potential energy of an isolated system. The values of the Taylor parameter and degree of spheroidal deformation at which the drop loses stability are determined for a wide range of dielectric constants of the substance of the drop. © *1999 American Institute of Physics.* [S1063-7842(99)00507-3]

The equilibrium shapes and stability of charged drops in uniform external electrostatic fields have been the subjects of numerous experimental and theoretical studies in connection with the investigation of elementary processes in storm clouds and the development of new chemical technologies and physical apparatus for analytical investigations (see, e.g., the reviews[1,2] and the literature cited therein). Nevertheless, there are still many unanswered questions on this topic. For example, the majority of the theoretical studies have been done for drops of a perfectly conducting liquid, and the question of how the critical equilibrium deformation of a drop of dielectric liquid in a uniform electrostatic field **E** is related to E and to the dielectric constant ε of the liquid has remained beyond the scope of these investigations. The present study should fill this gap.

The analysis below is carried out in an approximation in which the equilibrium shape of a drop of dielectric liquid in a field **E** is assumed to be spheroidal[3] and is based on the principle of minimum potential energy of the equilibrium shape of an isolated drop. We shall investigate the equilibrium states and the stability boundaries of a drop of dielectric liquid in the field **E**. The external medium will be assumed to be nonconducting, with a dielectric constant equal to unity, i.e., a vacuum.

1. The potential energy of an isolated drop in a field **E** consists of the energy of surface tension and the electrostatic energy of the drop in **E**. The energy of surface tension of a prolate spheroidal drop is written[4]

$$U_\sigma = 2\pi\sigma R^2\left(k^{-2/3} + k^{4/3}\,\frac{\tan^{-1}(\sqrt{k^2-1})}{\sqrt{k^2-1}}\right), \tag{1}$$

where R is the radius of a spherical drop having the same volume as the spheroidal drop under study, σ is the coefficient of surface tension, $k=a/b$, and a and b are the semimajor and semiminor axes of the prolate spheroid, respectively.

The electrostatic energy of a dielectric spheroid elongated along an external uniform electric field **E** in vacuum is given in SI units by the expression[5]

$$U_e = -\frac{1}{2}EP_x = \frac{2\pi\varepsilon_0 E^2 R^3(\varepsilon-1)}{3[1+(\varepsilon-1)n_x]}, \tag{2}$$

where P_x is the dipole moment of the spheroid, ε_0 is the dielectric permittivity of free space, and n_x is the depolarization coefficient of the spheroid, which has the form[5]

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

In the following stability analysis of the spheroidal shape of the drop, its total potential energy $U_\sigma + U_e$ will be rendered dimensionless by dividing by the potential energy of a spherical drop of the same volume,

$$U = \frac{U_\sigma + U_e}{U_{\sigma s} + U_{es}}, \tag{3}$$

where

$$U_{\sigma s} = 4\pi R^2\sigma \tag{4}$$

is the energy of surface tension of a spherical drop, and

$$U_{es} = -\frac{2\pi R^3 E^2\varepsilon_0(\varepsilon-1)}{4+\varepsilon} \tag{5}$$

is the electrostatic energy of a dielectric sphere in a field **E**.

We also introduce the dimensionless parameter W (the Taylor parameter), the value of which characterizes the stability of a drop in a field **E** (a drop of an electrically conducting liquid becomes unstable at $w=2.59$; Ref. 6):

$$W = \frac{4\pi\varepsilon_0 E^2 R}{\sigma}. \tag{6}$$

Substituting expressions (1), (2), (4), and (5) into (3), we rewrite (3) in dimensionless form as

$$U = \left\{2\pi\left[\frac{1}{k^{2/3}} + \frac{k^{4/3}\tan^{-1}(\sqrt{k^2-1})}{\sqrt{k^2-1}}\right] - \frac{W(\varepsilon-1)}{1+(\varepsilon-1)n_x}\right\}\left[4\pi - \frac{W(\varepsilon-1)}{2(2+\varepsilon)}\right]^{-1}. \tag{7}$$

 © 1999 American Institute of Physics

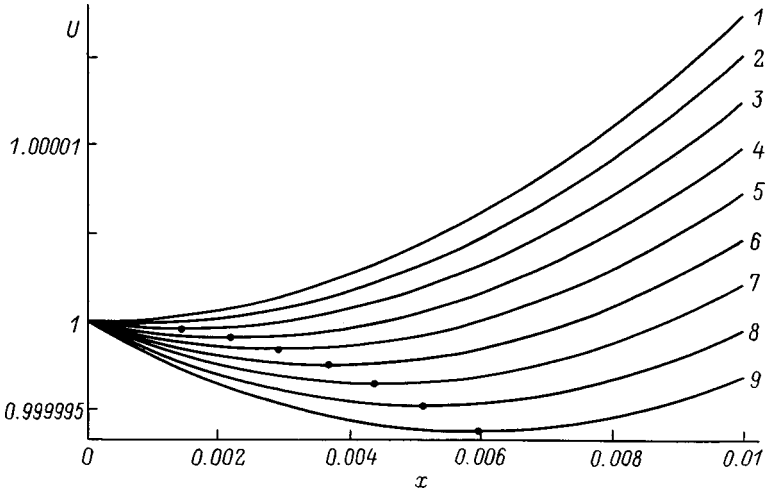

FIG. 1. Dimensionless energy of a drop of transformer oil in an electrostatic field, as a function of the deformation parameter x: $W=0.01$ (*1*), 0.1 (*2*), 0.2 (*3*), 0.3 (*4*), 0.4 (*5*), 0.5 (*6*), 0.6 (*7*), 0.7 (*8*), and 0.8 (*9*).

2. For convenience in graphical representation of the results of a numerical analysis of relation (7), we shall characterize the degree of spheroidal deformation of the drop in the field **E** by the quantity $x=\ln(a/b)$.

Figure 1 shows the curves of $U(x)$ for a drop of transformer oil ($\varepsilon=2.2$) as calculated numerically from Eq. (7) for different values of the Taylor parameter W. The presence of a single minimum on the $U(x)$ curves (the positions of the minima are indicated in Fig. 1 by dots) over a wide range of variation of the parameter W ($0\leqslant W\leqslant 40$) indicates that for this value of the dielectric constant of the substance there always exists an equilibrium shape of the drop in the form a prolate spheroid. A comparison of the positions of the minimum on the curves for different values of W (see also Fig. 1) shows that the eccentricity of the equilibrium spheroidal drop increases monotonically with increasing Taylor parameter W.

In Figs. 2a and 2b the curves $U(x)$ are shown in different scales for a drop of ethyl alcohol ($\varepsilon=46$) at different values of the Taylor parameter W. As the calculations show, and as can be seen from Fig. 2, for small values of the Taylor parameter $W<2.3$ the curves have a single minimum (curves *1–4* in Fig. 2a and 2b), corresponding to a spheroidal drop with a ratio of semiaxes $k<1.4$. As the Taylor parameter increases to values $W\geqslant 2.3$ there is a qualitative change in the shape of the $U(x)$ curves: now $U(x)$ has two minima separated by a local maximum. After the first minimum, the positions of which are denoted in Fig. 2b by dots, the $U(x)$ curve passes through a local maximum and then descends to a second minimum (the positions of which for different W are denoted in Fig. 2a by crosses), corresponding to an equilibrium spheroidal shape with a ratio of semiaxes $k>6$. Further growth of the degree of deformation x of the spheroid leads to unbounded growth of the potential energy. As the Taylor parameter W is increased, the positions of both the first and second minima are shifted to higher deformations.

As the Taylor parameter is increased to $W\geqslant 3.24$ the first minimum (and, with it, the local maximum) disappears, and $U(x)$ becomes a descending curve in the whole region $0<k<6$. Since the existence of stable equilibrium spheroidal drops with a ratio of semiaxes $k>6$, i.e., with an extremely strong elongation, is unlikely on general physical grounds that have not been incorporated in the model, the value of the Taylor parameter $W=3.24$ at which the first minimum van-

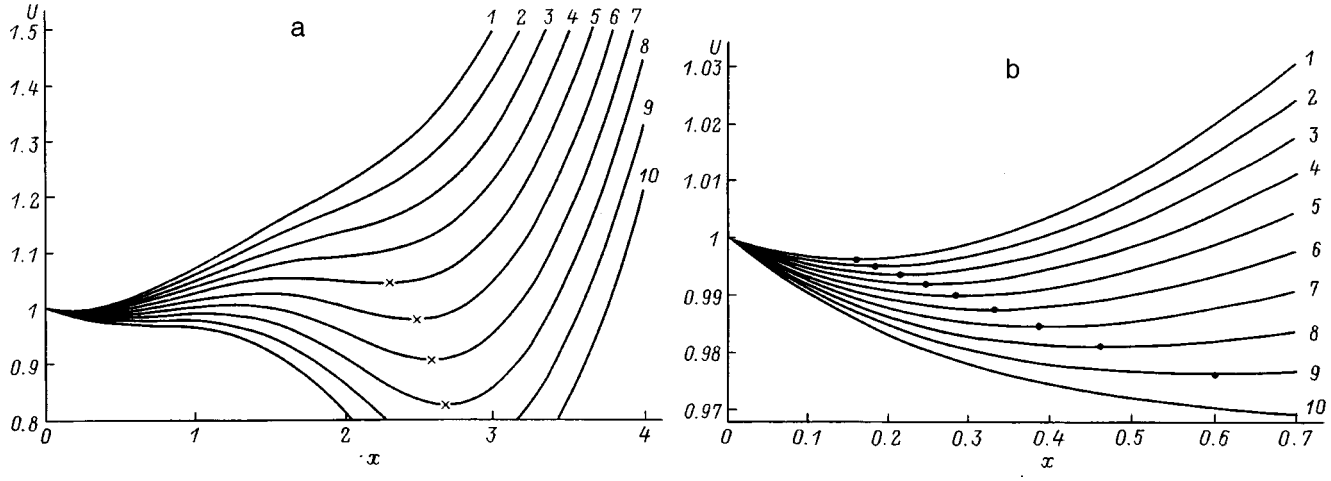

FIG. 2. Dimensionless energy of a drop of ethyl alcohol. a: The $U(x)$ curves at large deformations; the crosses indicate the minima in the region of large deformations for $W>2.3$; b: the curves at small deformations; the minimum in the region of small deformations are indicated by dots; $W=1.6$ (*1*), 1.8 (*2*), 2.0 (*3*), 2.2 (*4*), 2.4 (*5*), 2.6 (*6*), 0.8 (*7*), 3.0 (*8*), 3.2 (*9*), and 3.4 (*10*).

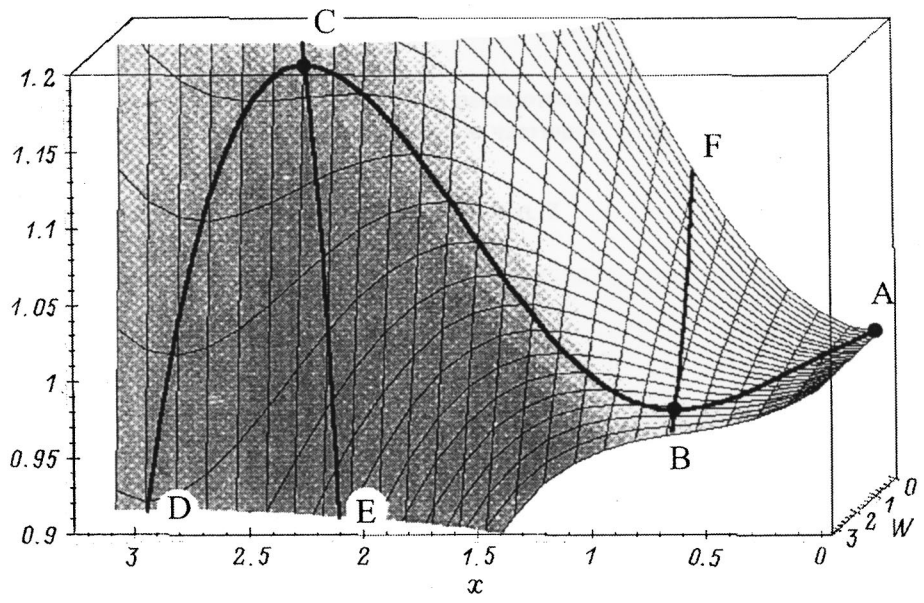

FIG. 3. Dimensionless density of a drop of water as a function of W and x.

ishes should be regarded as the upper boundary of the region of W values in which stable spheroidal drops exist for $\varepsilon=46$.

The $U(x)$ curves calculated for a water drop ($\varepsilon=81$) for different values of the Taylor parameter (Fig. 3) are qualitatively similar to those shown in Fig. 2. The significant differences are that: a) the second minimum on the $U(x)$ curve first appears at smaller values of the Taylor parameter ($W>1.43$); b) the positions of the local maximum of $U(x)$ are shifted to larger deformations; c) the vanishing of the first minimum and the local maximum of $U(x)$, i.e., its transition to a monotonically descending function in the region of deformations ($0<k<1.6$), is also observed at smaller values of the Taylor parameter ($W>2.3$).

3. The relation between the equilibrium spheroidal deformation of a drop and value of the Taylor parameter W can be found from the requirement that the total potential energy of the drop in the equilibrium state be minimum,

$$\frac{\partial}{\partial k}(U_\sigma+U_e)=0, \tag{8}$$

which leads to the equation

$$2\pi\left[-\frac{2}{3}k^{-5/3}+\frac{4}{3}\frac{k^{1/3}C}{A}+\frac{k^{1/3}}{A^2}-\frac{k^{7/3}C}{A^3}\right]+[3kA-(2k^2+1)B]\frac{WA}{D^2}=0,$$

which, with allowance for the requirement that the second derivative be positive,

$$\frac{\partial^2}{\partial k^2}(U_\sigma+U_e)\geqslant 0, \tag{9}$$

leads to the relation

$$2\pi k^{1/3}\left(\frac{10}{9}k^{-3}+\frac{4}{9}\frac{C}{A}k^{-1}+\frac{5}{3}\frac{k^{-3}}{A^2}-\frac{11}{3}\frac{kC}{A^3}-3\frac{k}{A^4}+3\frac{k^3C}{A^5}\right)+\frac{W}{D^2}\left\{-\frac{(2k^2+1)Bk}{A}+3A^2-4kAB+4k^2-1-2A[3kA-(2k^2+1)B]\times\left(\frac{3kA}{\varepsilon-1}+B\right)D^{-1}\right\}\geqslant 0,$$

$$A=\sqrt{k^2-1},\quad B=\ln(k+\sqrt{k^2-1}),$$

$$C=\tan^{-1}(\sqrt{k^2-1}),$$

$$D=\frac{(k^2-\varepsilon)\sqrt{k^2-1}}{\varepsilon-1}+k\ln(k+\sqrt{k^2-1}).$$

A graph of the curve $U=U(W,x)$ calculated from Eq. (7) for a drop of water is shown in Fig. 3. The curve AD is the projection of the solution of equation (8) on the surface $U(W,x)$, or, in other words, the curve AD passes through points at which $U(x)$ takes on extremal averages. The lines BF and CE are projections of the solution of equation (9) on the surface $U(W,x)$ and divides this surface into three parts: to the right of BF and to the left of CE are the loci of points at which

$$\frac{\partial^2}{\partial k^2}(U_\sigma+U_e)>0,$$

and between BF and CE is the locus of points at which

$$\frac{\partial^2}{\partial k^2}(U_\sigma+U_e)<0.$$

Thus the segments AB and CD correspond to minima and segment BC corresponds to maxima of the function

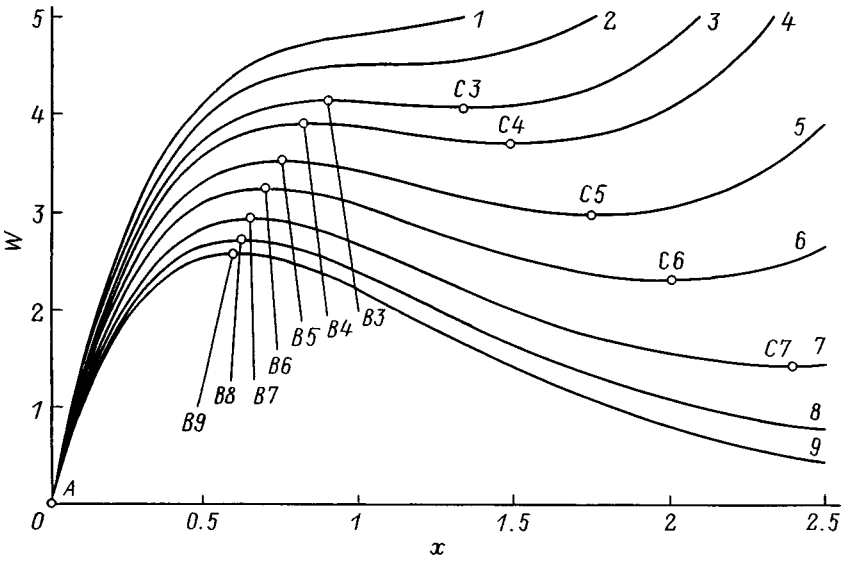

FIG. 4. Curves of $W(x)$ obtained from the solution of equation (8) for $\varepsilon = 18$ (*1*), 20 (*2*), 23 (*3*), 26 (*4*), 34 (*5*), 46 (*6*), 81 (*7*), 200 (*8*), and ∞ (*9*).

$U(x)$. It is clear from what we have said that the equilibrium shapes of the drop correspond to the curve AB.

Figure 4 shows the results of a numerical calculation using Eq. (8) for different values of the dielectric constant ε. The calculation shows that for $\varepsilon < 20.8$ an equilibrium shape of the drop in the form of a prolate spheroid exists for any value of the Taylor parameter (curves *1* and *2* in Fig. 4). For $\varepsilon > 20.8$ the equilibrium states of the spheroidal drops are observed up to a certain critical value of the Taylor parameter $W = W_*$ (denoted by points *B3–B9* in Fig. 4), and this critical value depends on the dielectric constant of the substance of the drop, in agreement with the data of Ref. 7. As the dielectric constant increases, the critical value W_* decreases, reaching at $\varepsilon \to \infty$ a minimum value $W_{*0} = 2.57$, which is in good agreement with the results of a study of the boundaries of stability of conducting drops in an electrostatic field.[6] The segments *A–Bn* in Fig. 4 correspond to equilibrium spheroidal shapes of the drop. Segments *Bn–Cn* of curves *3–9* correspond to local maxima of the function $U(x)$ for $W = \text{const}$ and $\varepsilon = \text{const}$. The value of the Taylor parameter W corresponding to the points *C3–C9* determines when the second minimum will appear on the $U(x)$ curve.

At values of the Taylor parameter less than the critical value W_*, for a drop to be brought to the region of instability it is necessary to impart to it an energy sufficient to overcome the potential barrier, the value of which is plotted in Fig. 5 (curves *Bn–Cn*) as a function of the Taylor parameter for various values of the dielectric constant ε. Segments *A–Bn* of the curves in Fig. 5 correspond to equilibrium states of the spheroidal drop, and the points *Bn* correspond to states of unstable equilibrium of the drop at the critical value of the Taylor parameter W.

The three-dimensional surface presented in Fig. 6 was determined from the relation between the dielectric constant ε of the substance of the drop, the Taylor parameter W, and the parameter x, according to the solution of equation (8). Curve AB, like the points *Bn* in Figs. 4 and 5, corresponds to states of unstable equilibrium of the drop at the critical values of the Taylor parameter W, and it separates the stable

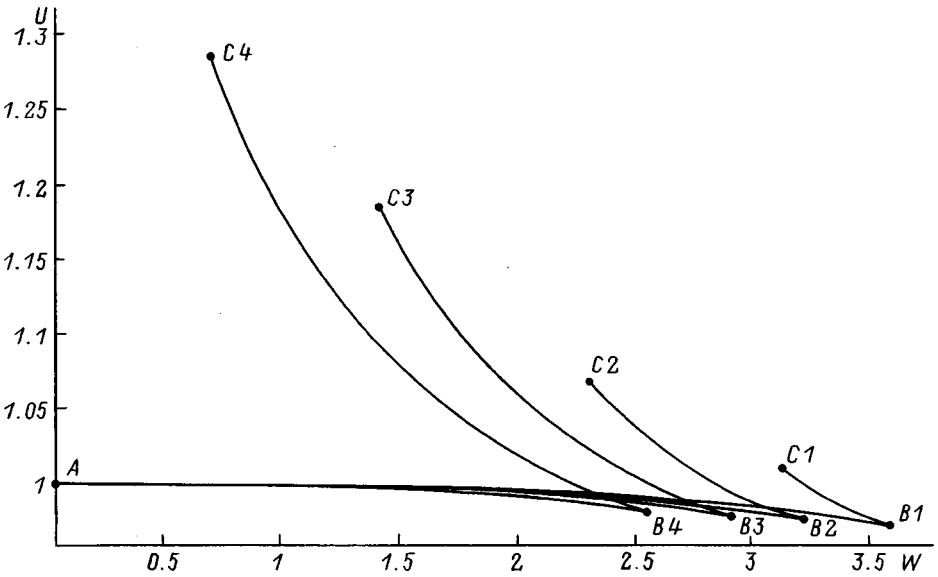

FIG. 5. Height of the potential barrier versus the Taylor parameter for $\varepsilon = 32$ (*1*), 46 (*2*), 81 (*3*), and ∞ (*4*).

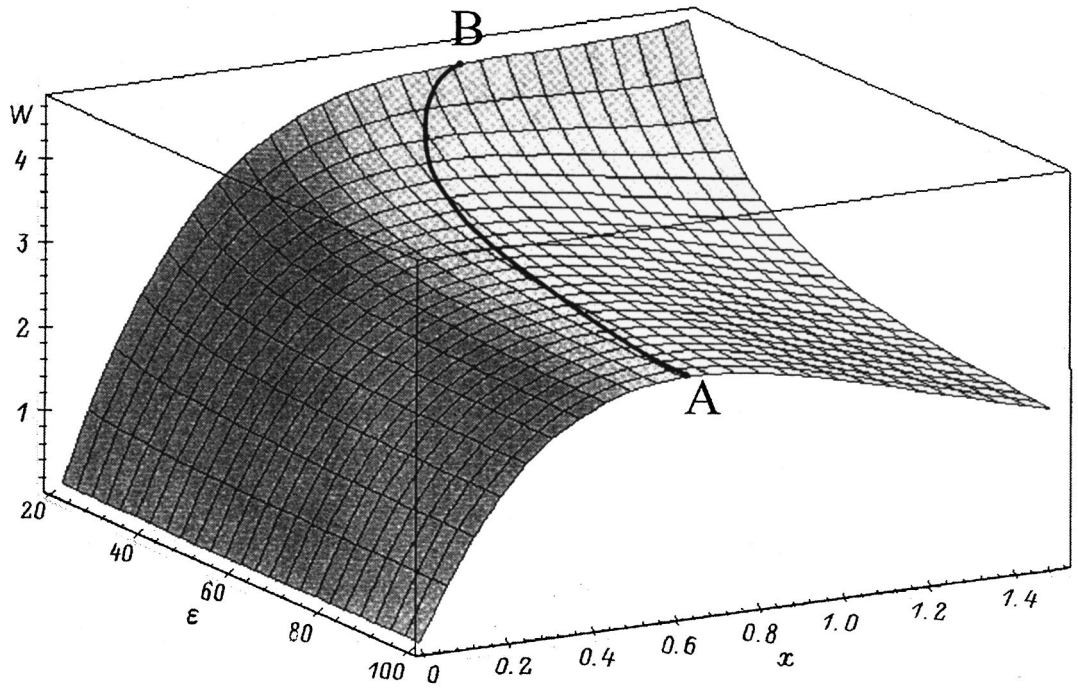

FIG. 6. Interrelationship between the Taylor parameter, the dielectric constant of the substance of the drop, and the deformation parameter.

spheroidal shapes of the drop (to the left of *AB*) from the unstable states (to the right of *AB*).

CONCLUSION

The foregoing investigation has revealed that for an isolated drop with dielectric constant ε in a uniform electrostatic field **E**, the form of the potential energy of the drop as a function of the spheroidal deformation, $U=U(x)$, depends both on ε and on the field **E** or the Taylor parameter $W\sim E^2$, the latter of which characterizes the stability of the surface of the drop against the charge induced on it. The $U(x)$ curves for different values of ε and E (or ε and W) can have either a single minimum or two minima separated by a local maximum. An analysis of $U=U(x,\varepsilon,W)$ has made it possible to map out the region of values of x, ε, and W corresponding to equilibrium spheroidal shapes of the drop and to determine the critical conditions for loss of stability.

[1] V. I. Kozhenkov and N. A. Fuks, Usp. Khim. **45**, 2274 (1976).
[2] A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk Mekh. Zhidk. Gazov, No. 3, pp. 3–22 (1994).
[3] C. T. O'Konski and H. C. Thacher, J. Phys. C **57**, 955 (1953).
[4] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* [2nd ed., Pergamon Press, Oxford (1986); 3rd ed., Nauka, Moscow (1986), 733 pp.].
[5] 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.].
[6] G. Taylor, Proc. R. Soc. London, Ser. A **280**, 383 (1964).
[7] J. D. Sherwood, J. Fluid Mech. **188**, 133 (1988).

Translated by Steve Torstveit

Point defects and ring defects in a nematic liquid crystal in a cylindrical capillary

V. K. Pershin, I. I. Klebanov, and P. B. Zalmanov

Chelyabinsk State Pedagogical University, 454080 Chelyabinsk, Russia
(Submitted February 23, 1998)
Zh. Tekh. Fiz. **69**, 29–32 (July 1999)

The bidirectional escape into the third dimension of a linear disclination of strength $m=1$ (L^p_{+1}) in a cylindrical capillary with normal boundary conditions is investigated. It is shown that in this case two types of defects arise in the capillary: point defects and ring defects, each of which can be of the radial or hyperbolic type. Exact solutions are obtained for the equation of equilibrium of the elastic field. The free energy of the point and ring defects is calculated approximately in a narrow, long capillary. New scenarios are proposed for the escape of the disclination L^p_{+1}. © *1999 American Institute of Physics.* [S1063-7842(99)00607-8]

1. In nematic liquid crystals (NLCs) the director field $\mathbf{n}(\mathbf{r})$ can have singular points, lines, and surfaces at which the direction of the director n is indeterminate. These singularities are called point, linear, and surface defects, respectively.

In an NLC in a cylindrical capillary with normal boundary conditions, the structure with a singularity on the axis (L^p_{+1}, a linear disclination with $m=1$) becomes energetically unfavorable at capillary radii ρ larger than a certain critical value ρ_c and it ''escapes'' into the third dimension.[1] The solution describing the unidirectional escape of a disclination of strength $m=1$ was obtained in Ref. 2. However, for an NLC the two directions along the axis of the capillary are equivalent. Therefore, it is of interest to consider the problem of bidirectional escape of a disclination, when the two directions are connected by a singular point. Such a possibility was indicated in Ref. 1.

The goal of the present study is to investigate the bidirectional escape into the third dimension of a linear disclination of strength $m=1$ in the one-constant approximation of the continuum theory of NLCs.

2. Let us consider an NLC in a cylindrical capillary of radius ρ_0 and length $2L$, with normal boundary conditions. The components of the director in the cylindrical coordinate system are written as

$$n_\rho=\sin\alpha(z,\rho);\quad n_\varphi=0;\quad n_z=\cos\alpha(x,\rho), \tag{1}$$

where α is the angle between the ''line of force'' and the axis of the capillary, $\alpha(\rho=\rho_0)=\pi/2$.

In the one-constant approximation the free energy of deformation is[1]

$$F=\frac{K}{2}\int_{(V)}((\nabla\cdot\mathbf{n})^2+(\nabla\times\mathbf{n})^2)dV, \tag{2}$$

where K is the Frank elastic modulus, V is the volume of the NLC, and $\mathbf{n}$ is the director.

Using (1) in expression (2), we get

$$F=\pi K\int_0^{\rho_0}d\rho\int_{-L}^{L}dz\left(\frac{\sin^2\alpha}{\rho}+\sin 2\alpha\frac{\partial\alpha}{\partial\rho}-2\sin^2\alpha\frac{\partial\alpha}{\partial z}+\rho\left(\frac{\partial\alpha}{\partial\rho}\right)^2+\rho\left(\frac{\partial\alpha}{\partial z}\right)^2\right). \tag{3}$$

Minimizing (3), we get the equation of equilibrium

$$2\rho^2\frac{\partial^2\alpha}{\partial z^2}+2\rho\frac{\partial}{\partial\rho}\left(\rho\frac{\partial\alpha}{\partial\rho}\right)=\sin 2\alpha. \tag{4}$$

Being a dimensionless quantity, the angle α can depend only on dimensionless combinations of coordinates. We assume that α depends only on the combination z/ρ. Then, after introducing the new variable $x=\sinh^{-1}(z/\rho)$, we write Eq. (4) in the form

$$2\alpha''_{xx}=\sin 2\alpha. \tag{5}$$

The intermediate result of integration is

$$(\alpha'_x)^2=c+\sin^2\alpha, \tag{6}$$

where $c\geqslant 0$ is a constant of integration.

After a second integration with the normal boundary conditions taken into account, we have

$$\int_{\pi/2}^{\alpha}\frac{d\xi}{\sqrt{c+\sin^2\xi}}=\pm\ln u, \tag{7}$$

where

$$u=\frac{z/\rho+\sqrt{1+z^2/\rho^2}}{z/\rho_0+\sqrt{1+z^2/\rho_0^2}}.$$

The substitution

$$\xi=\frac{\pi}{2}-\psi$$

reduces the left-hand side of Eq. (8) to an incomplete elliptic integral of the first kind,

$$\int_0^{\pi/2-\alpha}\frac{d\psi}{\sqrt{1-k^2\sin^2\psi}}=\pm\frac{1}{k}\ln u, \tag{8}$$

 © 1999 American Institute of Physics

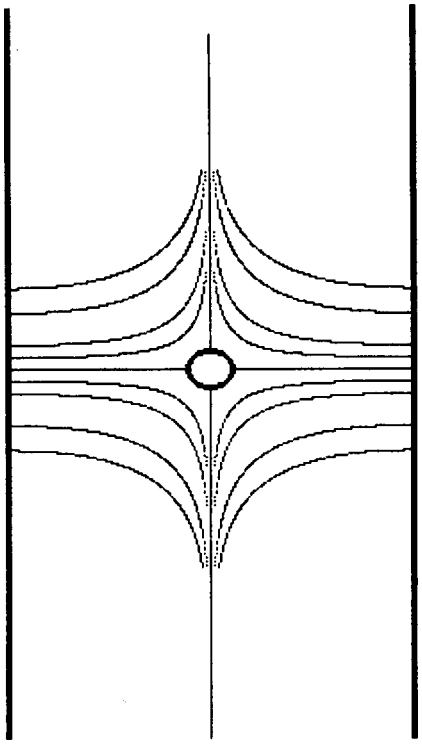

FIG. 1. Hyperbolic point defect (H).

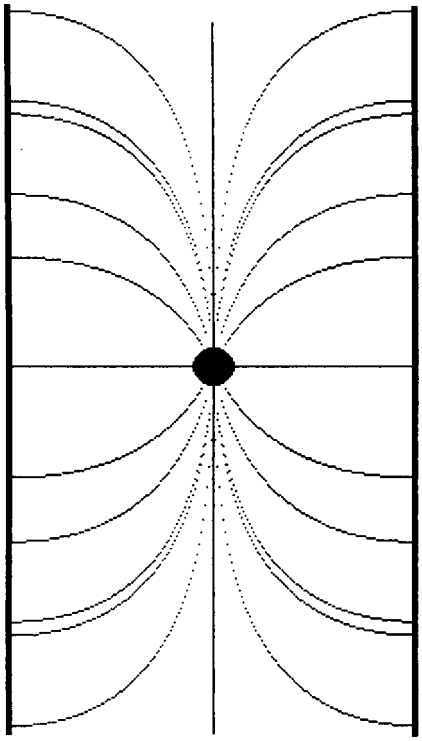

FIG. 2. Radial point defect (R).

where $k=1/\sqrt{1+c}$.

Hence

$$\alpha=\frac{\pi}{2}\pm \mathrm{am}\left(\frac{\ln u}{k}\right), \tag{9}$$

where am is the Jacobi amplitude.[3]

In order to study the elastic deformation field, we must consider a family of lines of force satisfying the differential equation

$$d\rho/n_\rho=dz/n_z, \tag{10}$$

which must be integrated with allowance for (1) and (9).

3. Let us first consider the case $k=1$ ($c=0$). Then (9) is transformed to

$$\alpha_1=2\tan^{-1}u \quad \text{or} \quad \alpha_2=\pi-\alpha_1. \tag{11}$$

The lines of force obtained as a result of the numerical integration of (10) with allowance for (1) and (11) are drawn in Figs. 1 and 2 (these and the other figures pertain to a cylindrical capillary and show a section by a plane passing through the axis). As can be seen from these figures, bidirectional escape of the disclination gives rise to point defects of the hyperbolic H and radial R types. In the limit $|z|\gg\rho_0$ we obtain the well-known Cladis–Kleman relations[2] corresponding to unidirectional escape of the disclination:

$$\alpha_1=\begin{cases}\pi-2\tan^{-1}(\rho/\rho_0), & z>0,\\ 2\tan^{-1}(\rho/\rho_0), & z<0,\end{cases}$$

$$\alpha_2=\pi-\alpha_1. \tag{12}$$

For $\rho_0\to\infty$ the angles α_1 and α_2 go over to the well-known solutions describing hyperbolic and radial point defects in an infinite space.[4]

Let us now evaluate the free energy of H and R defects. Here we limit discussion to the most interesting case of a long and narrow capillary ($\rho_0\ll L$). Substituting (11) into (1), we get

$$\frac{F}{2\pi KL}\approx 3+\frac{\pi}{2}q\frac{\rho_0}{L}+O(\rho_0^2/L^2), \tag{13}$$

where $q\approx 1.5$ for an H defect and $q\approx 3$ for an R defect.

In calculating the energy we have expanded the integrand in Eq. (3) in a series in powers of a small parameter, keeping the first two terms of the expansion. In the region $|z|>\rho_0$ the small parameter is $\rho/|z|$, and in the region $|z|<\rho_0$ it is $|z|/\rho_0$.

Expression (13) shows that the energy of a structure containing a point defect differs from the energy of a nonsingular structure by the presence of an additional term which appears because of the stronger distortion of the lines of force near the defect. Although this term does not contribute substantially to the energy, its presence suggests that the energy gap between H and R defects is smaller than in free space.

In fact, in free space Eq. (11) becomes

$$\alpha_1=2\tan^{-1}(z/\rho+\sqrt{1+z^2/\rho^2}), \quad \alpha_2=\pi-\alpha_1. \tag{14}$$

Integrating the equation of the lines of force (10) with allowance for (14), one is readily convinced that α_1 corresponds to an H defect and α_2 to an R defect. Let us now mentally select out an imaginary capillary of radius ρ_0 and

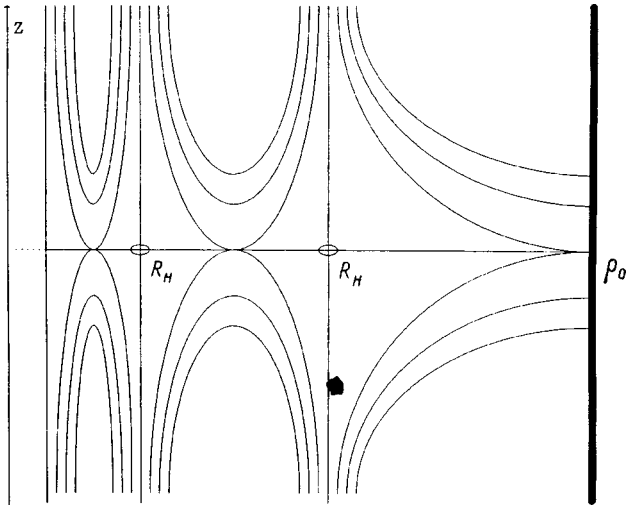

FIG. 3. System of hyperbolic ring defects (R_H).

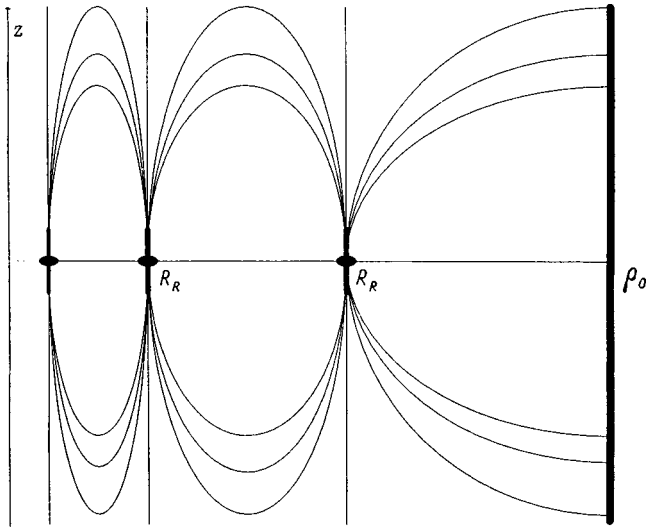

FIG. 4. System of radial ring defects (R_R).

length $2L$ and calculate the energy of the elastic field in this imaginary capillary. Substituting (14) into (13), we get

$$\frac{F^{(H)}}{4\pi K}=\rho_0 \tan^{-1}(L/\rho_0),$$

$$\frac{F^{(R)}}{4\pi K}=2\rho_0 \tan^{-1}(L/\rho_0)+L\ln(1+\rho_0^2/L^2). \tag{15}$$

Assuming that $\rho_0 \ll L$ and keeping only the first term of the series expansion in powers of ρ_0/L, we obtain

$$\frac{F}{2\pi KL}\approx \frac{\pi}{2}q\frac{\rho_0}{L}+\mathrm{O}(\rho_0^2/L^2), \tag{16}$$

where $q=2$ for an H defect and $q=4$ for an R defect.

Thus

$$\frac{F^{(R)}-F^{(H)}}{2\pi KL}\approx \frac{\pi}{2}\frac{\rho_0}{L}2, \tag{17}$$

while for a real capillary

$$\frac{F^{(R)}-F^{(H)}}{2\pi KL}\approx \frac{\pi}{2}\frac{\rho_0}{L}\frac{3}{2}, \tag{18}$$

i.e., the energy gap decreases.

4. For $k=1$ ($c=0$) the singular structures forming in the capillary are qualitatively different from those considered above: they are now ring hyperbolic R_H and radial R_R defects. Their elastic field is illustrated schematically in Figs. 3 and 4.

For an analytical description of the rings we again limit discussion to the case of a long and narrow capillary and consider the problem of unidirectional escape of the disclination in the case $c=0$. This problem can be solved exactly, and all the results obtained will clearly be valid also for bidirectional escape in the zeroth approximation in an expansion in powers of ρ_0/L. Solving Eq. (4) without the first term on the left-hand side, we get

$$\alpha=\frac{\pi}{2}\pm \mathrm{am}\left(\frac{\ln(\rho/\rho_0)}{k}\right). \tag{19}$$

Substituting (19) into (10), we get

$$\frac{dz}{d\rho}=\pm\frac{\mathrm{sn}(\ln(\rho/\rho_0)/k)}{\mathrm{cn}(\ln(\rho/\rho_0)/k)}, \tag{20}$$

where sn and cn are the sine and cosine of the Jacobi amplitude.[3]

The right-hand side of (2) has a singularity determined by the equation

$$\mathrm{cn}(\ln(\rho/\rho_0)/k)=0 \tag{21}$$

or

$$\rho=\rho_0\exp[-k(2m+1)K(k)], \quad m=0,1,2,\ldots, \tag{22}$$

where

$$K(k)=\int_0^{\pi/2}\frac{d\psi}{\sqrt{1-k^2\sin^2\psi}}$$

is the complete elliptic integral of the first kind.

In other words, we have a family of infinitely many cylindrical surfaces whose radii fall off in a geometric progression as the axis of the capillary is approached. On these surfaces the angle α tends asymptotically to 0 or π, depending on the direction of escape. On the axis of the capillary a singularity arises due to the indeterminacy of the value of the angle α for $\rho=0$ [see Eq. (19)]. Taking the relation $k=1/\sqrt{1+c}$ into account, we easily see that for $c\gg 1$ the cylinders almost uniformly fill the space inside the capillary, and for $c\ll 1$ they are concentrated mainly near the axis. For $c=0$ we have the Cladis–Kleman escape.[2]

In view of what was said above, we can conclude that ''in the zeroth approximation'' this picture holds for ring defects as well, the only difference being that for $c=0$ the system of radial rings R_R degenerates into a point r defect, while the system of hyperbolic rings R_H degenerates into a point H defect.

The free energy of the rings is found by substituting (19) into (3) and taking into account that the result thus obtained is the zeroth approximation in an expansion in powers of ρ_0/L:

$$\frac{F}{2\pi kKL}\approx\frac{2}{k}E(\mathrm{am}(\ln(\rho_0/b)/k))+\mathrm{sn}^2(\ln(\rho_0/b)/k)$$
$$-(k'/k)^2\ln(\rho_0/b)+q(k)\frac{\rho_0}{L}+\mathrm{O}(\rho_0^2/L^2)+f_0, \tag{23}$$

where

$$E(\varphi)=\int_0^{\sin\varphi}\sqrt{\frac{1-k^2t^2}{1-t^2}}dt, \quad k'=\sqrt{1-k^2},$$

$q(k)$ is a bounded function of k ($q(k)\to(\pi/2)q$ as $k\to1$) and b is a so-called cutoff parameter of the order of molecular scales; f_0 is the free energy (in units of $2\pi KL$) of the core of the singular structure (f_0 cannot be calculated in the framework of continuum theory[1]).

For $c\gg1$

$$\frac{F}{2\pi KL}\approx c\ln(\rho_0/b)+(1/c)\frac{\rho_0}{L}+\mathrm{O}(\rho_0^2/L^2)+f_0. \tag{24}$$

For $c\ll1$

$$\frac{F}{2\pi KL}\approx3+\frac{\pi}{2}q\frac{\rho_0}{L}-c\ln\left(\frac{c\rho_0}{2b}\right)+\mathrm{O}(\rho_0^2/L^2)+\mathrm{O}(c)+f_0. \tag{25}$$

Expressions (24) and (25) show that a system of rings that almost uniformly fills the entire capillary region is energetically unfavorable, since its energy is of the same order as the energy of a linear disclination of integer strength.[1] Rings bunched together near the axis, however, are almost as energetically favorable as the nonsingular structure L^{NSL}.

5. In summary, a linear disclination L^p_{+1} can escape into the third dimension in at least three ways. For a unidirectional escape a structure L^{NSC} forms which contains an infinite number of cylindrical surfaces, on each of which the direction of the director is asymptotically parallel to the axis of the capillary, and a singularity on the axis of the capillary. A degenerate case of such a structure is a disclination escaping in the Cladis–Kleman manner (L^{NSL}). For bidirectional escape a system of ring defects of the radial R_R or hyperbolic R_H type form, which in the degenerate case go over to point R or H defects, respectively.

In principle more complex scenarios for the escape of L^p_{+1} are possible, viz.,

$$L^p_{+1}\mapsto L^{NSC}\overset{\leftarrow}{\mapsto}L^{NSL},$$
$$L^p_{+1}\mapsto R_R\overset{\leftarrow}{\mapsto}R,$$
$$L^p_{+1}\mapsto R_H\overset{\leftarrow}{\mapsto}H.$$

Here the symbols $\mapsto$ and $\to$ denote the more probable and less probable scenarios, respectively. To determine which of these is actually realized it will be necessary to do a more accurate calculation of the free energy of the structures described above and to consider specific mechanisms for the onset of instability.

[1] P. G. de Gennes, *The Physics of Liquid Crystals* [Pergamon Press, Oxford (1974); Mir, Moscow (1977), 400 pp.].
[2] P. E. Cladis and M. Kleman, J. Phys. (Paris) **33**, 591 (1972).
[3] E. Jahnke, F. Emde, and F. Lösch, *Tables of Higher Functions*, 6th ed. [McGraw–Hill, New York (1960); Nauka, Moscow (1977), 344 pp.].
[4] S. Chandrasekhar and G. S. Ranganath, Adv. Phys. **35**, 507 (1986).

Translated by Steve Torstveit

Bipolar atmospheric-pressure corona discharge in a He/H_2 mixture

A. K. Shuaibov

Uzhgorod State University, 294 Uzhgorod, Ukraine
(Submitted April 2, 1998)
Zh. Tekh. Fiz. **69**, 33–37 (July 1999)

Results are presented from a study of the characteristics of a steady-state plasma in a He/H_2 mixture at high pressures. The plasma is formed in the outer region of a multielectrode corona discharge. It is shown that molecules of helium hydride form in such a medium, and their decomposition is accompanied by continuum emission in the 350–650 nm region. The corona discharge is distributed over the length and can be used in systems for the transverse electric circulation of the working medium of atmospheric-pressure plasma radiation sources. The reference spectra of the plasma emission and the dependence of the relative intensity of the lines and the brightness of the bands of He_2^* on the value of the discharge current, pressure, and composition of the working mixture are investigated. © *1999 American Institute of Physics.* [S1063-7842(99)00707-2]

INTRODUCTION

Atmospheric-pressure He/H_2 mixtures under excitation in an electron-beam-controlled discharge (duration $\geqslant 0.1$ ms) are of considerable interest for application in IR lasers working at vibrational–rotational transitions of the hydrogen halide molecules (HF, HCl, etc.), and both the case of room temperatures of the gaseous medium[1] and nitrogen temperatures ($T=100$ K)[2] have been investigated in detail.

The possibility of building a periodic-pulse laser source working at the $A–X$ transition of the excimer molecule HeH* with $\lambda_{max}=460$ nm and $\Delta\lambda=310$ nm for excitation of a He/H_2 mixture by modulated high-energy electron beams or by optical pumping was demonstrated in Refs. 3–5. In order to maintain the optimum gas temperature, such sources require circulation and cooling of the working medium. In these sources it may be appropriate to employ electric circulation of the gaseous mixtures by means of a multielectrode corona discharge[6,7] distributed uniformly over the length of the active medium of the source. Corona discharges are also used in electrofilters for removing solid particulates that have been sputtered in the electrode system of a plasma radiation source.. The conditions for achieving stable corona discharges in He/H_2 mixtures at high pressures has not been studied before. At a lower pressures ($\leqslant 0.2$ kPa) He/H_2 mixtures have been used in laser plasma radiation sources using HeI atoms ($\lambda=706.5$ nm),[8–10] but in that case the use of electric circulation is hard to implement because of instabilities of the corona discharge at low pressures of the gaseous medium. In the majority of cases the He/H_2 mixtures have been investigated under pulsed excitation by electron beams or a transverse discharge; it is therefore of physical interest to study the parameters of a dense plasma based on this mixture in the case of steady electrical-discharge pumping. The plasma under such conditions of excitation should be supercooled, a circumstance which is important for the formation of the excimer molecules HeH* and He_2H* and also for studying the kinetics of the processes occurring in such media.

In this paper we present the results of a study of the electrical and optical characteristics of a bipolar corona discharge of negative polarity, ignited in the working medium of an atmospheric-pressure plasma radiation source using a He/H_2 mixture.

EXPERIMENTAL PROCEDURES AND CONDITIONS

The multielectrode corona discharge was investigated on the apparatus described in Refs. 11 and 12. The electrode system used consisted of one row of point electrodes and a semitransparent metallic grid. The rounding radius of the tips was 0.1–0.2 mm, and the radius of the wires in the grid was 0.15 mm. The density of tips in the row was 1 per centimeter of length, and the total length of the electrode system was 11 or 17 cm. A dc voltage of negative polarity ($U\leqslant 10$ kV, $I\leqslant 15$ mA) was applied to the tips across a limiting resistor with $R=1–3$ MΩ. The emission from the corona discharge plasma was investigated with a panoramic view of the entire interelectrode gap. Radiation from the hot zones formed near the tips was shielded. The radiation from the plasma was analyzed with a DFS-12 monochromator, an FÉU-106 photomultiplier, a U5-9 dc amplifier, and a chart recorder. The resolution of the spectrometer was 0.1 nm or better. The relative calibration of the radiation detection system was done with the use of DVS-25 and SI 8-200 standard lamps.

A steady discharge regime was used. The possible presence of a pulsed component of the radiation from the plasma was monitored by means of a 14-FS linear electron multiplier and an S1-79 fast pulsed oscilloscope.

SPATIAL, ELECTRICAL, AND OPTICAL CHARACTERISTICS

A corona discharge in a He/H_2 mixture was stably ignited at a pressure of 50–300 kPa and was observed as a spatially uniform discharge similar to those in mixtures of He/Ar, Kr, and Xe.[13] The transverse aperture of the corona

 © 1999 American Institute of Physics

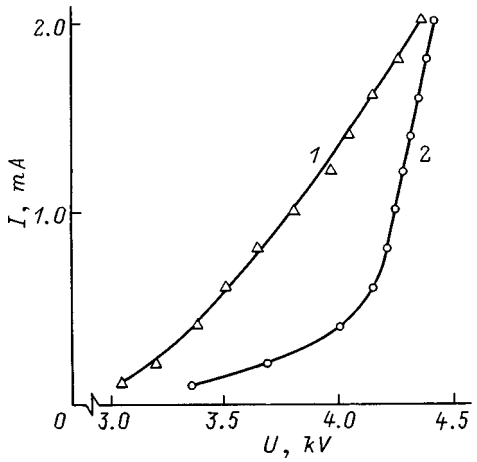

FIG. 1. Current–voltage characteristics of a negative corona discharge in helium at $P=200$ (*1*) and in a $He/H_2 = 200/1.2$ kPa mixture (*2*).

discharge plasma was in the form of an isosceles triangle with the base on the grounded electrode. Increasing the voltage across the point electrodes gave rise to streamer breakdown around one of the tips at a pressure exceeding 150 kPa or to the ignition of an arc at lower pressure. Increasing the concentration of H_2 molecules ($P \geqslant 5-10$ kPa) caused the spatially uniform discharge in the outer region of the corona discharge to vanish and led to the formation of bright hot zones near the tips, similar to what is found for a corona discharge in He/N_2 or He/CO mixtures.[11]

Typical current–voltage (I–V) characteristics of the corona discharge in He and in a He/H_2 mixture are shown in Fig. 1. In helium the I–V characteristic of the corona discharge has a close-to-linear form, while for a He/H_2 mixture it indicates the presence of two discharge stages, each of which can be described by a linear I–V characteristic. The growth rate of the current in the high-current stage of the corona discharge in a mixture with H_2 is 3 mA/kV, while in the low-current stage it is 0.25 mA/kV. The main difference between the I–V characteristic in these media may be due to the formation of H^- negative ions in the He/H_2 mixture at higher charging voltages U on the tips, and at lower voltages, to the efficient excitation of the high vibrational states $H_2(v)$ (Ref. 14), which enter into a dissociative attachment reaction with electrons.[1] The dependence of the ignition potential and the region of stable existence of the discharge in terms of the value of U as the pressure of the He/H_2 mixture is changed (at $[H_2] \leqslant 5$ kPa) was similar to the corresponding data for mixtures of He/Xe/Ar, etc.[13] The main carriers of positive change in the investigated medium are H_3^+ and He_2^+ ions[4,5] while the negative charge is carried mainly by H^- ions and electrons.

Figure 2 shows the emission spectrum of a corona discharge plasma in a He/H_2 mixture at $I=1.5$ mA. The spectrum is shown without allowance for the relative calibration of the spectrometer (k_λ). The spectra were interpreted with the aid of tables of Refs. 15 and 16. The emission spectra of a plasma based on He or a He/H_2 mixture typically exhibited

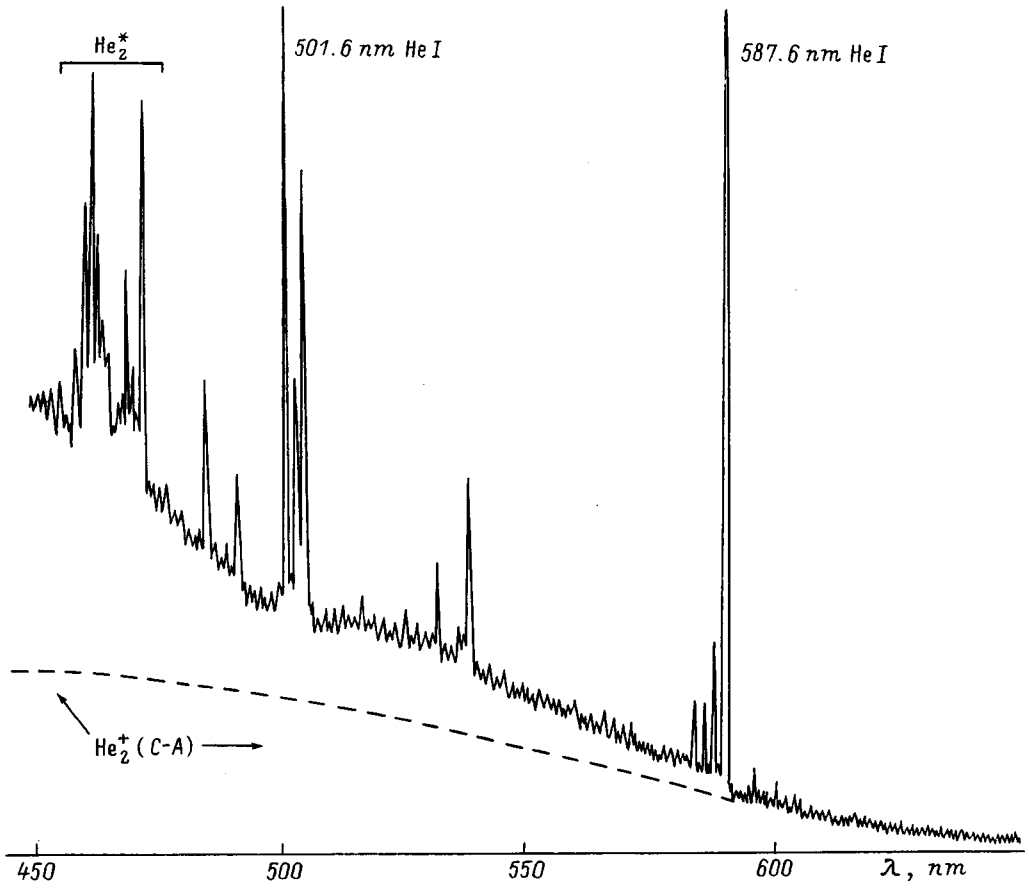

FIG. 2. Panoramic spectrum of the emission from a corona discharge plasma in a He/H_2 mixture.

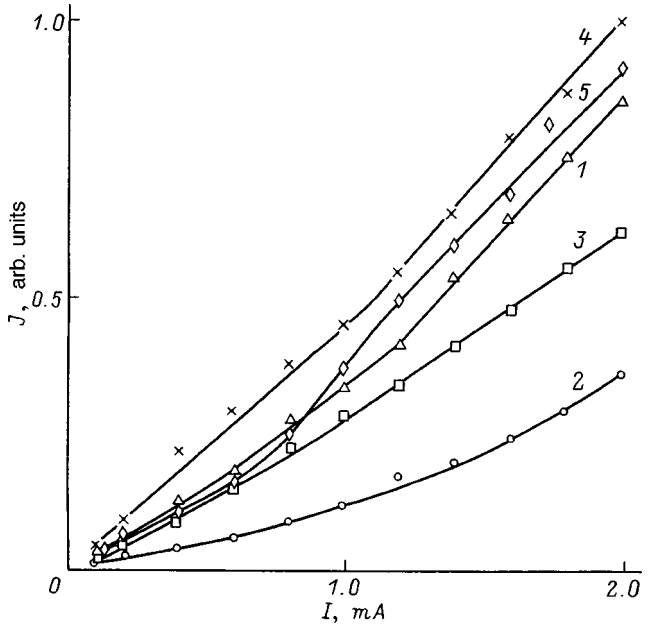

FIG. 3. Relative intensity of the emission in the lines of the He atom in a corona discharge plasma in a He/H_2 = 200/1.2 kPa mixture (*1–3*) and in He (*4,5*): 587.6 (*1,5*), 501.6 (*2*), 388.9 nm (*3,4*).

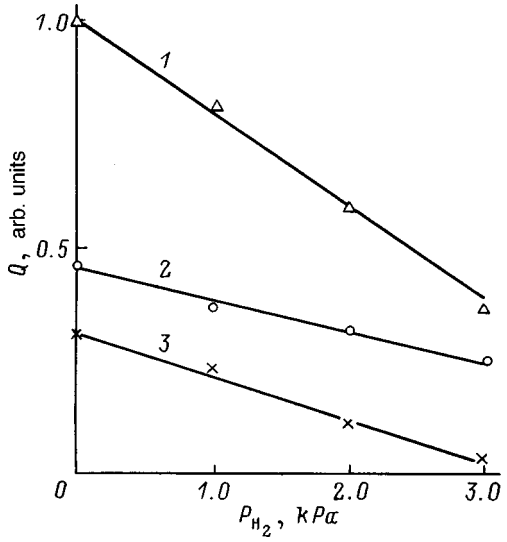

FIG. 4. Brightness of the emission in the He_2^* bands versus the H_2 partial pressure in He/H_2 mixtures with [H_2]=200 kPa: *1* — 505.7 nm line $He_2(e^3\Pi_g - a^3\Sigma_u^+)$ (0,1); *2* — 473.3 nm line of $He_2(E^1\Pi_g - A^1\Sigma_u^+)$ (1,0); *3* — 464.9 nm line of $He_2(e^3\Pi_g - a^3\Sigma_u^+)$ (0,0).

a continuum in the region 360–650 nm. Observed against the background of this continuous radiation are the most intense lines of the He atom: 388.9, 501.6, and 587.6 nm, and a series of molecular bands of He_2^* in the blue-green region of the spectrum. The dashed curve in Fig. 2 shows the continuum emission in a pure helium medium (at [He]=200 kPa) (this continuum can be identified[13] with *C–A* transitions of He_2^+) under the same excitation conditions as for the plasma in the He/H_2 mixture. The intensity of the continuum emission of the plasma in the He/H_2 mixture is significantly higher than in the helium medium.

Figure 3 shows the dependence of the relative intensity of the emission (with allowance for k_λ) on the value of the corona discharge current, and Fig. 4 shows the brightness of the He_2^* bands as a function of the partial pressure of H_2 molecules in the He/H_2 mixture. The 587.6 nm line was the most intense line in the corona discharge plasma in the He/H_2 mixture, while the 388.9 nm line was predominant in pure He. Admixtures of H_2 molecules to He had the greatest influence on the intensity of the 388.9 nm line and only a weak influence on the 587.6 nm line (Fig. 3). At discharge currents $\leqslant$0.8 mA their intensity ratio is inverted. Increasing the number of H_2 molecules in the He/H_2 mixture leads to a strong quenching of the emission in the He_2^* bands.

Thus small admixtures of H_2 molecules to He have a considerable effect on the shape of the I–V characteristics of a corona discharge; they lead to significant growth of the brightness of the continuum emission in the 360–650 nm range and to strong quenching of the emission at the 388.9 nm line of HeI and in the bands of He_2^*.

DISCUSSION OF THE RESULTS

Excited He atoms (n=2,3) are formed in a dense, weakly ionized plasma in a dissociative recombination[17] or stepwise excitation process. The dissociative recombination reaction can occur only with He_2^+ ions ($v \geqslant 3$).[18] A comparison of the distribution of the relative intensities of the HeI lines with the distribution of the effective cross sections for stepwise excitation of the corresponding lines of He (Ref. 19) showed that they do not correlate with each other. Therefore, a better candidate for the population of HeI (n=3,4) is the dissociative recombination of He_2^+ ions ($v \geqslant 3$) with electrons.

The continuum emission from a plasma in a He/H_2 mixture appears as a result of the formation of HeH* excimer molecules.[5] In the ground state the HeH molecule is unbound, while the lower excited states are stable, with a binding energy $\geqslant$2.3 eV.[20,21] The HeH continuum observed in Ref. 20, with a maximum at 250 nm, belongs to the transition $B^2\Pi - X^2\Sigma^+$ of the HeH molecule, while the longer-wavelength emission from HeH can be attributed to the system $A^2\Sigma^+ - X_2\Sigma^+$ (Ref. 21). The state $A^2\Sigma^+$ has a short lifetime with respect to predissociation, 0.65 ps, much shorter than the corresponding lifetime of HeH($B,v=1-5$), which is 56–5 ns.[22]

The formation of HeH* molecules in the corona discharge plasma in a He/H_2 mixture occurs as a result of the reactions[3–5,21]

$$\mathrm{He} + \mathrm{H_2}(C^1\Pi_u, B^1\Sigma_u^+) \rightarrow \mathrm{HeH^*} + \mathrm{H}, \quad (1)$$

$$\mathrm{H}(2s,2p) + 2\mathrm{He} \rightarrow \mathrm{HeH^*} + \mathrm{He}, \quad (2)$$

$$\mathrm{HeH^+} + e + (\mathrm{He}) \rightarrow \mathrm{HeH^*} + (\mathrm{He}), \quad (3)$$

$$\mathrm{He_2H^+} + e + (\mathrm{He}) \rightarrow \mathrm{HeH^*} + (\mathrm{He}). \quad (4)$$

The predominant population of the 3^3D state of HeI, which is the upper level for the 587.6 nm line, is due to the selective character of the dissociative recombination reaction

for He_2^+ ions ($v \geqslant 3$)[17] and the mixing of such states by electron impact. Since the rate constant for the mixing of the 3^3D states of HeI is negative,[23] they play the role of an energy acceptor, and the 587.6 nm line stands out considerably in intensity in the emission spectrum of the plasma. In a corona discharge plasma H_2 molecules lead to quenching of the emission in the lines of HeI except for the 587.6 nm line, at discharge currents $\leqslant 0.8$ mA (Fig. 3). The increase in intensity of this emission line of HeI may be due to depopulation of its lower level through processes of Penning ionization of the H_2 molecules.[24,25] At atmospheric pressure of the working mixture one of the main channels for formation of $He_2^+(C)$ ions is the reaction[26]

$$He_2^* + He^+ \rightarrow He_2^+(C) + He + e, \quad (5)$$

and therefore the admixture of H_2 molecules to He decreases the density of $He_2^+(C)$ ions. As a result, the $He_2^+(C-A)$ continuum is transformed to a more intense HeH($A-X$) continuum (Fig. 2). Interaction processes between He_2^+ ions and H_2 molecules are characterized by high rates and give a substantial contribution to the formation of HeH^+ and He_2H^+ ions:[25]

$$He_2^+ + H_2 \rightarrow \begin{cases} HeH^+ + H + He, \\ He_2H^+ + H. \end{cases} \quad (6)$$

Subsequently the HeH^+ and He_2H^+ ions recombine with slow electrons in reactions (3) and (4), which result in the formation of HeH* molecules.

A processing of the dependence of the brightness of the He_2^* band at 505.7 nm (the most sensitive to the H_2 content in the He/H_2 mixture) on the H_2 density using the Stern–Volmer formula for the quenching of luminescence[27] showed that the product $k_t\tau = (1.4 \pm 0.2) \times 10^{-16}$ cm^3 (τ is the lifetime of He_2^*). For $\tau = 10$ ns the effective rate constant for the quenching of He_2^+ by H_2 molecules is $(1.4 \pm 0.2) \times 10^{-8}$ cm^3/s.

CONCLUSIONS

A study of the characteristics of a multielectrode corona discharge in a high-pressure He/H_2 mixture has shown the following.

At $P = 50$–300 kPa and hydrogen contents $[H_2] \leqslant 5$ kPa the discharge covers the entire interelectrode gap and exists in a steady-state regime, while at higher H_2 contents it is characterized by a set of pointlike hot zones forming near the electrode tips.

The current–voltage characteristic of the discharge is described by two straight lines with different rates of current growth, indicating the presence of two different stages of the discharge.

In the 360–650 nm region the emission of HeH($A-X$) and He_2^* molecules is detected, and also the 388.9, 501.6, and 587.6 nm lines of HeI.

The main mechanism populating the excited states of HeI is the dissociative recombination of $He_2^*(v)$ ions with electrons.

H_2 molecules efficiently quench the emission at transitions of HeI and He_2.

The discharge studied can be used for transverse electric circulation of the working medium in radiation sources utilizing HeH* molecules in high-pressure plasmas.

[1] A. V. Dem'yanov, N. A. Dyatko, I. V. Kochetov *et al.*, Zh. Tekh. Fiz. **58**(1), 75 (1988) [Tech. Phys. **33**, 43 (1988)].
[2] A. V. Dem'yanov, I. V. Kochetov, A. F. Pal' *et al.*, Zh. Tekh. Fiz. **60**(1), 204 (1990) [Sov. Phys. Tech. Phys. **35**, 124 (1990)].
[3] V. S. Zuev, A. V. Kanaev, and L. D. Mikheev, Kvantovaya Élektron. (Moscow) **11**, 197 (1984) [Sov. J. Quantum Electron. **14**, 135 (1984)].
[4] F. V. Bunkin, V. I. Derzhiev, G. A. Mesyats *et al.*, Kvantovaya Élektron. (Moscow) **11**, 1277 (1984) [Sov. J. Quantum Electron. **14**, 864 (1984)].
[5] B. A. Azimdzhanov, T. U. Arslanbekov, F. V. Bunkin *et al.*, Kvantovaya Élektron. (Moscow) **12**, 1557 (1985) [Sov. J. Quantum Electron. **15**, 1029 (1985)].
[6] A. I. Baranov, K. V. Gurkov, M. I. Lomaev *et al.*, Prib. Tekh. Éksper., No. 4, pp. 108–111 (1994).
[7] A. K. Shuaibov, I. V. Shevera, A. I. Minya *et al.*, Prib. Tekh. Éksper., No. 5, pp. 98–103 (1997).
[8] D. Schmider and T. I. Salamon, Opt. Commun. **55**, 49 (1985).
[9] A. A. Berdnikov, V. I. Derzhiev, I. I. Murav'ev *et al.*, Kvantovaya Élektron. (Moscow) **14**, 2198 (1987) [Sov. J. Quantum Electron. **17**, 1400 (1987)].
[10] V. I. Derzhiev, V. F. Tarasenko, S. I. Yakovlenko *et al.*, Tr. Inst. Obshch. Fiz. Akad. Nauk SSSR (IOFAN) **21**, 5 (1989).
[11] A. K. Shuaibov, A. I. Minya, V. V. Zvenigorodskiĭ *et al.*, Pis'ma Zh. Tekh. Fiz. **22**(13), 73 (1996) [Tech. Phys. Lett. **22**, 551 (1996)].
[12] A. K. Shuaibov, L. L. Shimon, and I. V. Shevera, Kvant. Elektron. (Moscow) **24**, 20 (1997).
[13] A. K. Shuaibov and I. V. Shevera, Zh. Tekh. Fiz. **67**(11), 15 (1997) [Tech. Phys. **42**, 1254 (1997)].
[14] J. R. Hiskes, in *Negative Ions and Beams. Fourth International Symposium*, New York (1986), pp. 34–38.
[15] A. R. Striganov and G. A. Odintsova, *Tables of Spectral Lines of Atoms and Atomic Ions. A Handbook* [in Russian], Energoatomizdat, Moscow (1982).
[16] S. N. Suchard and J. E. Melzer (eds.), *Spectroscopic Data. Homonuclear Molecules*, IEI/Plemum, New York (1976), pp. 226–251.
[17] V. A. Ivanov and N. P. Penkin, Zh. Prikl. Spektrosk. **40**, 5 (1984).
[18] V. A. Ivanov and Yu. É. Skoblo, Opt. Spektrosk. **65**, 750 (1989) [Opt. Spectrosc. **65**, 445 (1988)].
[19] A. A. Mityureva and N. P. Penkin, Opt. Spektrosk. **66**, 1220 (1989) [Opt. Spectrosc. **66**, 710 (1989)].
[20] W. Ketterle, A. Dodny, and H. Wolther, Chem. Phys. Lett. **129**, 76 (1986).
[21] T. Moller, M. Beland, and G. Zimmerer, Chem. Phys. Lett. **136**, 551 (1987).
[22] A. B. Alekseev, V. S. Ivanov, A. M. Pravilov *et al.*, Opt. Spektrosk. **69**, 93 (1990) [Opt. Spectrosc. **69**, 56 (1990)].
[23] Yu. A. Piotrovskiĭ and Yu. A. Tolmachev, Opt. Spektrosk. **52**, 33 (1982) [Opt. Spectrosc. **52**, 17 (1982)].
[24] V. V. Evstigneev and S. S. Filippov, Preprint IPM AN SSSR No. 5 [in Russian], Institute of Applied Mathematics, Academy of Sciences of the USSR, Moscow (1974), 20 pp.
[25] Yu. I. Syts'ko and S. I. Yakovlenko, IAÉ Preprint No. 3138 [in Russian], I. V. Kurchatov Institute of Atomic Energy, Moscow (1979), 29 pp.
[26] P. G. Hill and P. R. Herman, Phys. Rev. A **47**, 4837 (1993).
[27] V. N. Kondrat'ev and E. E. Nikitin, *Kinetics and Mechanisms of Gas-Phase Reactions* [in Russian], Nauka, Moscow (1974).

Translated by Steve Torstveit

Modeling of diffusion and absorption processes of tritium in beryllium membranes

V. V. Kirsanov† and V. V. Grigor'eva

Tver State Technical University, 170026 Tver, Russia
(Submitted March 10, 1998)
Zh. Tekh. Fiz. **69**, 38–41 (July 1999)

The permeation of tritium produced in thermonuclear fusion through metallic membranes, which is a topical problem in radiation physics, is addressed. A physical model for the permeation of hydrogen through a beryllium membrane is proposed which takes into account the oxide layer on the surface of the membrane. The model is implemented in the form of a system of differential equations, which is solved by numerical methods. As a result of the computer modeling it is shown that as the temperature is raised, the oxide layer in a certain interval blocks the permeation of tritium in the membrane, and it is found that the tritium is distributed in a rather limited region of the membrane over its thickness. This circumstance will permit the use of special processing by etching or mechanical treatment. © *1999 American Institute of Physics.* [S1063-7842(99)00807-7]

1. INTRODUCTION

Currently the most promising power plants are controlled fusion reactors. The parts of these reactors which are in direct contact with the working plasma are exposed to intense radiation over long periods of time. The gases released in the operation of a fusion reactor (isotopes of hydrogen, helium, etc.) actively interact with the materials of the first wall and divertor. Along with the gases a large number of radiation defects are formed in these structures; modification of the crystal structure occurs, and radiation embrittlement, pitting, and blistering are observed. All of these strongly complicate the mechanisms of migration and occlusion of the gases, making it necessary in some cases to reconsider the established physicochemical ideas.

Beryllium membranes are most often proposed for use in the first wall and divertor components of fusion reactors as devices for controlling the flux of tritium and ensuring the safe operation of the reactor. Over the course of their irradiation and the introduction of structural defects, the properties of these membranes change considerably, particularly their permeability to hydrogen isotopes. In this study we have attempted to devise a computer model that would simulate the permeation of tritium through beryllium membranes for specified initial distributions of the various radiation defects in the membrane material.

2. PHYSICAL MODEL

Typical particle fluxes near the irradiated surface of a membrane in a fusion reactor are of the order of 10^{21} DT/m^{2}s. The saturation concentration in an oxide film several microns thick is reached in a few minutes. Consequently, the boundary conditions for the initial flux from the irradiated side should be the condition of constant concentration. The expected concentration of mobile hydrogen atoms at the entrance is proportional to the solubility of hydrogen in beryllium and, according to Sievert's law, to the square root of the saturation pressure. In test calculations described below we took the pressure in the subsurface layer of the membrane in the interval 1.33–13.3 kPa.

In the majority of cases it is necessary to take into account that the beryllium membrane is coated by a thin oxide layer. Data on the solubility of hydrogen isotopes in BeO are limited. In this paper we use the formula

$$S = 1.25\times 10^{-11}\exp\{0.8\ \text{eV}/k\Theta\}[\text{atomic fraction}/\sqrt{\text{Pa}}], \tag{1}$$

which was obtained by fitting the experimental data of Ref. 1. We note that the activation energy -0.8 eV was taken from the paper of Macaulay-Newcomb and Thomson[2] and was obtained at temperatures below 873 K. For the coefficient of diffusion of tritium in BeO the best approximation to the experimental data, according to Ref. 3, is the expression

$$D = 1.31\times 10^{-9}\exp\{-1.335/k\Theta\}[\text{m}^2/\text{s}]. \tag{2}$$

In view of the small interatomic distances in the crystal lattice of beryllium the question of the solubility of hydrogen isotopes in pure polycrystalline beryllium is still open to discussion. In Ref. 3 the formula

$$S = 1.76\times 10^{-7}\exp\{-0.17/k\Theta\}(\text{atomic fraction}/\sqrt{\text{Pa}}) \tag{3}$$

was proposed for temperatures in the range 500–600 °C. For the diffusion coefficient under these same conditions it is customary to use the values given by Jones and Gibson,[4]

$$D = 1.31\times 10^{-11}\exp\{-1.335/k\Theta\}[\text{m}^2/\text{s}] \tag{4}$$

or (for a membrane of ultrapure Be) by Abramov *et al.*,[5]

$$D = 6.7\times 10^{-9}\exp\{-0.294/k\Theta\}[\text{m}^2/\text{s}]. \tag{5}$$

3. MATHEMATICAL MODEL

As the initial mathematical model of the process of diffusion and absorption of tritium atoms T in a plane metallic membrane M of thickness L, in the presence of lattice defects

 © 1999 American Institute of Physics

(vacancies V and bivacancies $2V$) and the possible formation of clusters TV, $T2V$, $2T2V$, we take a system of differential equations of the Johnson–Lam type,[6]

$$\partial C_\nu/\partial t - D_\nu(\partial^2 C_\nu/\partial x^2) = F_\nu, \quad \nu = 1,2,3,5,7,10. \tag{6}$$

Here $C_\nu = C_\nu(x,t)$ are the unknown concentrations of migrating defects in atomic fractions, which by extension will also be referred to as particles or elements, at the point x of a one-dimensional sample $\Delta = \{x \in R: 0 \leqslant x \leqslant L\}$ at time t; D_ν are the corresponding diffusion coefficients, and F_ν are polynomial functions of $C_1, C_2, \ldots, C_n$ having a form determined by the character of the interactions and the interaction constants in the following seven elementary processes (''reactions'') involving the formation or annihilation of six types of defects:

$$V + V \rightarrow^{K_{11}} 2V, \quad V + T \rightarrow^{K_{13}} TV,$$

$$2V + T \rightarrow^{K_{23}} 2VT, \quad 2V \rightarrow^{K_2} V + V, \tag{7}$$

$$2VT \rightarrow^{K_7} V + VT, \quad V + VT \rightarrow^{K_{15}} 2VT,$$

$$T + 2VT \rightarrow^{K_{37}} 2V2T. \tag{8}$$

The coefficients written above the arrows in the reactions are their rate constants, which depend exponentially on the absolute temperature Θ of the sample M by the Arrhenius laws:

$$K_{\mu\nu} = K^0_{\mu\nu} \exp(-E_{\mu\nu}/k\Theta),$$

$$K_\nu = K^0_\nu \exp(-E_\nu/k\Theta).$$

Here $E_{\mu\nu}$ and E_ν are the activation energies of the reactions, k is Boltzmann's constant, and the coefficients $K^0_{\mu\nu}$ and K^0_ν are independent of Θ and are expressed in terms of the parameters of the crystal lattice.

For the sake of specificity, we stipulate that the values of the subscripts 1, 2, 3, 5, 7, and 10 in Eqs. (6) correspond to defects V, $2V$, T, TV, $2VT$, and $2V2T$ in the order written. This numbering is used because the model under discussion is a specialization of the more general model of Ref. 9, in which a system of 11 differential equations of the form (6) was used. The defects VT and $2V2T$ are assumed to be stabilized, nondecaying, and the diffusion coefficients D_5 and D_{10} are assumed to be zero. All of the remaining diffusion coefficients are taken to be functions with an exponential dependence on the absolute temperature Θ: $D_\nu = D^0_\nu \exp(-B_\nu/k\Theta)$, where B_ν is the energy of migration of the corresponding particle.

Interstitial tritium atoms T are implanted to a depth l, $0 < l < L$, through the entrance (irradiated) surface S_0 of the membrane M, which in the one-dimensional case corresponds to the point $x = 0$, at a specified pressure P of the gaseous tritium near the surface S_0. In the model a boundary layer of depth l adjacent to the irradiated surface of the membrane is assumed to be filled by a BeO oxide film, and the interval $l < x < L$ corresponds to a membrane layer filled with high-purity (99%) Be. The initial concentrations of vacancies and bivacancies over the sample are assumed given. Subsequently the above-described ensemble of point defects migrates into the interior of M and enters into reactions (7), (8). The tritium T can escape from the sample through the exit surface $S_L(x = L)$.

In accordance with the stated assumptions, let us specify the mathematical model as an initial/boundary-value problem for a system of $n = 6$ equations (6) with right-hand sides of the form

$$\begin{aligned}
F_1 &= -K_{11}C_1^2 + K_2C_2 - K_{13}C_1C_3 - K_{14}C_1C_4 \\
&\quad - K_{15}C_1C_5 - K_{16}C_1C_6 + K_7C_7 + K_8C_8, \\
F_2 &= K_{11}C_1^2 - K_2C_2 - K_{23}C_2C_3 - K_{24}C_2C_4, \\
F_3 &= -K_{13}C_1C_3 - K_{23}C_2C_3 - K_{37}C_3C_7 - K_{38}C_3C_8 \\
&\quad + K_{45}C_4C_5 + K_{49}C_4C_9 + K_9C_9, \\
F_5 &= K_{13}C_1C_3 - K_{15}C_1C_5 + K_7C_7 - K_{45}C_4C_5, \\
F_7 &= K_{23}C_2C_3 + K_{15}C_1C_5 - K_{37}C_3C_7 \\
&\quad - K_{47}C_4C_7 + K_9C_9 - K_7C_7, \\
F_{10} &= K_{37}C_3C_7.
\end{aligned} \tag{9}$$

The nonlinear terms in the right-hand sides of system (6) take into account the binary interactions of particles and clusters via ''reactions'' (7) and (8).

The penetration of defects into the membrane and the escape of gases through the surfaces S_0 and S_L are also regulated by the proper choice of initial and boundary conditions that adequately model the conditions of external pumping and removal of particles while maintaining the overall mass and flux balances. The mechanisms for the penetration and escape of particles through the surface of the membrane can be of different natures and, strictly speaking, should be described by individual models of seepage through barriers. These topics and the mechanisms of reactions of type (7) and (8) will not be discussed in detail in this paper.

In the spatially one-dimensional model under consideration the simplest forms of boundary conditions are those which characterize either the values of the concentrations C_ν of defects or the fluxes of these concentrations at the entrance and exit. We choose the boundary conditions in the form

$$\partial C_\nu/\partial x|_{x=0} = 0, \quad \partial C_\nu/\partial x|_{x=L} = 0, \quad \text{if} \quad \nu \neq 3;$$

$$C_3(0,t) = S \cdot \sqrt{P}; \quad C_3(L,t) = 0, \tag{10}$$

and the initial conditions

$$C_\nu(x,0) = C^0_\nu(x), \quad \nu = 1, 2, 3, 5, 7, 10 \tag{11}$$

From a mathematical standpoint, nonlinear systems of the type (1) often turn out to be stiff, and the corresponding explicit finite-difference schemes are unstable even for small time steps. A general approach for dealing with problems of stiffness is to use implicit iterative methods.[7,8] The physical cause of these effects is the onset of rapidly damped transient processes that must be dealt with by the numerical scheme and which, in the end, have no effect on the solution. For numerical modeling of diffusion processes at macroscopic times one must go over to a reduced model. The averaging

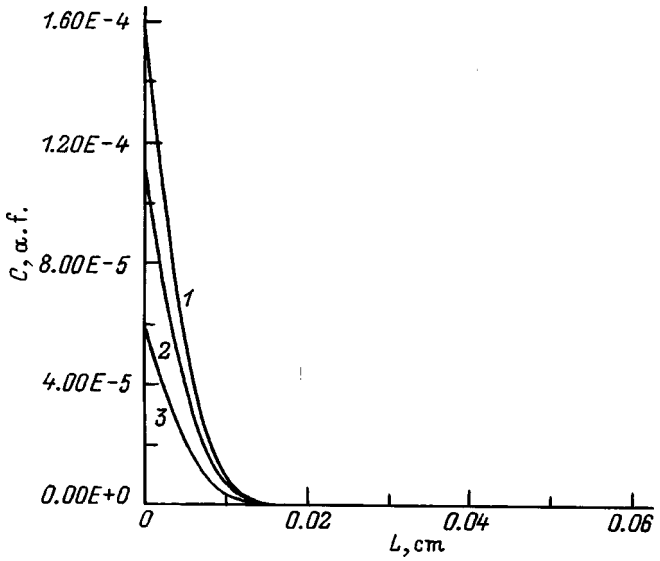

FIG. 1. Distribution of the tritium concentrations over the sample at the time $t=1800$ s for a temperature $\Theta=573$ K and pressures of $P_1=13.3$ kPa (*1*), $P_2=7.0$ kPa (*2*), and $P_3=2.0$ kPa (*3*).

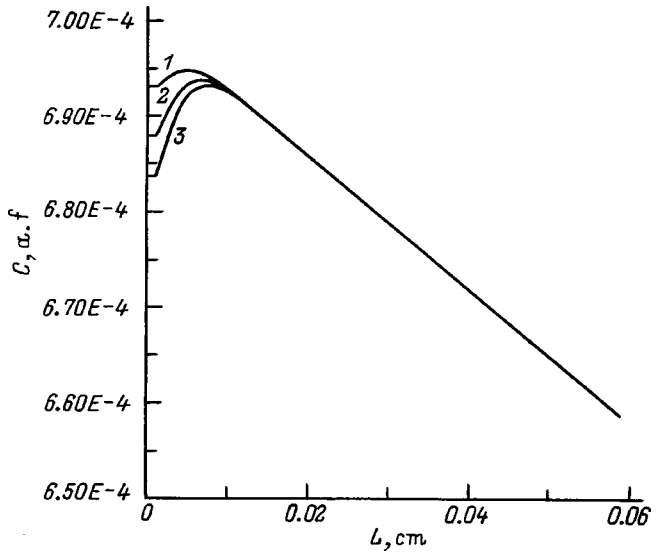

FIG. 2. Distribution of the vacancy concentrations over the sample at time $t=1800$ s under the same conditions as in Fig. 1, curves *1–3*.

property of the reduction operator, which is described in Ref. 9, leads to a computationally stiff system of differential equations which can be integrated numerically even at macroscopic (of the order of 1 s) time steps with a simultaneous improvement of the computational stability.

Because the reduced model coincides exactly with the initial system, up to the notation for the coefficients, we will not stop to write it out here but will go on to a description of a discrete analog of the model.

Consider on the interval $[0,L]$ a uniform mesh $\omega_H=\{x_i=iH;i=0\ldots N;NH=L\}$ with step $H>0$. At the internal mesh points of the grid we approximate each equation of system (6) with right-hand sides (7) and (8) by the familiar difference scheme of second-order accuracy in time and a predictor–corrector scheme in space.[8] The algebraic system of difference equations thus obtained is supplemented by the difference forms of the boundary-value and initial-value problems (10) and (11). The implantation depth l is assumed to be comparable to the mean free path of the T atoms in the membrane; in the discrete problem we assume that $l=2H$.

The algorithm for solving the stated problem inclusions the successive application of tridiagonal inversion[8] at each time step, and the time step τ is chosen from the conditions for stability of the calculation.

4. RESULTS OF THE COMPUTER MODELING

The main results of the calculation were obtained under the following general conditions. A beryllium membrane of thickness $L=0.6$ mm, the irradiated side of which is coated by a thin oxide film 0.0024 mm thick, is irradiated by a tritium plasma at temperatures in the interval 573–773 K. The pressure P in different simulations had values of 2, 7, and 13.3 kPa. The solubility and diffusion coefficients in BeO and Be and the initial concentrations of vacancies and bivacancies over the sample were taken according to the recommendations of Ref. 3, as described in Sec. 1 above. The segment $[0,L]$ was divided into $N=200$ parts. The following values were used for the diffusion coefficients of mobile defects: $D_1=5.0\times10^{-5}\exp(-0.68/k\Theta)$, $D_2=8.0\times10^{-5}\exp(-0.66/k\Theta)$, $D_3=1.31\times10^{-7}\exp(-1.335/k\Theta)$, and $D_7=1.0\times10^{-6}\exp(-0.9/k\Theta)$. The remaining D_ν were assumed to be zero. Figure 1 shows the results of test calculations of the distribution of tritium over the sample in a membrane of thickness $L=0.6$ mm at $T=573$ K for various tritium pressures in the subsurface layer. The calculation time was 1800 s. The diffusion coefficient was calculated according to Gibson's formula.[4] The tritium concentrations at the entrance were proportional to the solubility S and to the square root of the pressure. The initial concentrations of vacancies over the sample was linear, with a maximum at the entrance of $(C_0-(x/L))\cdot7\times10^{-4}$ atomic fraction. The concentration of bivacanies was constant and equal to 1×10^{-4} atomic fraction.

As we see from Fig. 2. The final distribution of vacancies over the volume after the calculation time could change appreciably in an initial segment of length 0.01 cm; here the bottom curve corresponds to the highest initial pressure, 13.3 kPa. The trend of the curves showing the distribution of the complex TV over the sample under the conditions described above have the same general features as the curves of Fig. 1, but the maximum penetration depth of tritium is approximately 1.3 times larger than the maximum depth at which an appreciable number of TV complexes is observed.

The calculations showed that the final distribution of bivacancies in the initial segment, to a depth of 0.01 cm, qualitatively reiterates the vacancy distribution.

To study the temperature dependence of the concentrations of T, V, $2V$, and TV in the model, with the entrance surface of the membrane coated with a thin oxide film, the vacancy and bivacancy concentrations were taken to be constant and equal to 5×10^{-4} and 1×10^{-4} atomic fractions, respectively, and the temperatures were $\Theta_1=573$ K and $\Theta_2=773$ K. The tritium pressure was 13.3 kPa, and the time

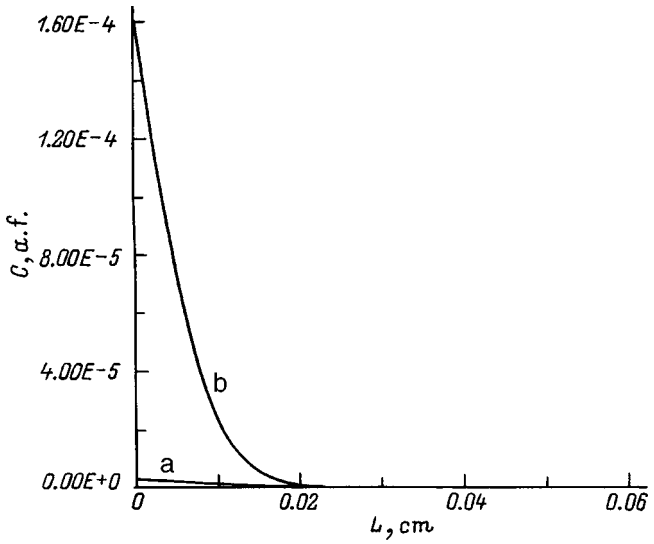

FIG. 3. Distribution of the tritium concentration over the sample in the presence of an oxide layer, for $\Theta = 773$ K (a), 573 K (b) and $t = 3600$ s.

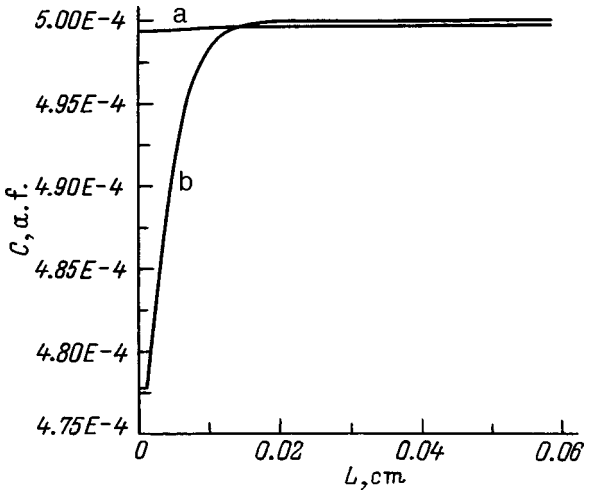

FIG. 4. Curves of the vacancy distribution over the sample for $C_1(0,x) = 5\times10^{-4}$ atomic fraction, $\Theta = 773$ K (a), 573 K (b) and $t = 3600$ s.

$t = 3600$ s. We see from Fig. 3 that as the temperature is raised there is a sharp decrease in the initial and final distributions of tritium over the sample, even though the penetration depth of tritium is approximately the same in the two cases, a little more than 0.02 cm. The final distribution of TV complexes over the sample has a similar trend (Fig. 4). The qualitative behavior of the curves of the final distributions of V and $2V$ remains unchanged, as expected. Noticeable deviations to lower values from the initial distribution are observed down to depths of 0.015 cm. Thus as the temperature is increased from $\Theta_1 = 573$ K to $\Theta_2 = 773$ K the oxide layer block the flux of tritium in the membrane.

The model calculations showed a slowing of the escape of tritium from the unirradiated side of the membrane as the initial concentrations of vacancies was increased and in comparison with the defect-free case; this can be explained by the formation of immobile mixed clusters of TV and $2T2V$ in the membrane.

The distribution of tritium trapped by the membrane over its thickness (as we see from Fig. 2) occupies a rather limited region of the membrane. From everything we know, this may make it possible to remove the accumulated tritium by etching or mechanical treatment.

[†]Deceased

[1] D. L. Balduoin, in *Woorkshop on Beryllium for Fusion Application*, Karlsruhe, Germany (1993), KfK 5271.
[2] R. G. Macaulay-Newcomb and D. A. Thomson, J. Nucl. Mater. **212–215**, 942 (1994).
[3] G. R. Longhurst *et al.*, Fusion Technol. **28**, 1217 (1995).
[4] P. Jones and R. Gibson, J. Nucl. Mater. **21**, 353 (1967).
[5] E. Abramov, M. P. Riehm, and D. A. Thomson, *Deuterium Permeation and Diffusion in High Purity Beryllium*, Canadian Fusion Fuel Technology Project (1990), N CFFTP-G-9013.
[6] V. V. Kirsanov and S. B. Kislitsyn, *Computers and the Modeling of Defects in Crystals* [in Russian], A. F. Ioffe Physicotechnical Institute, Academy of Sciences of the USSR, Leningrad (1982), pp. 106–107.
[7] J. M. Ortega and W. G. Poole Jr., *An Introduction to Numerical Methods for Differential Equations* [Pitman, Marshfield, Mass. (1981); Nauka, Moscow (1986)].
[8] A. A. Samarskiĭ and A. V. Gulin, *Numerical Methods* [in Russian], Nauka, Moscow (1989).
[9] V. V. Kirsanov and V. V. Grigor'eva, in *Collected Works of Young Scientists, Commemorating the 75th Anniversary of Tver State Technical University* [in Russian], Tver State Technical University, Tver (1997), pp. 42–50.

Translated by Steve Torstveit

Relation between the reduced and unreduced hardness in nanomicroindentation tests

S. I. Bulychev

Moscow State Industrial University, 109280 Moscow, Russia
(Submitted March 18, 1998)
Zh. Tekh. Fiz. **69**, 42–48 (July 1999)

Hooke's law is generalized to the case of arbitrary elastic or plastic indentation $\varepsilon=(2/\sqrt{\pi})\times(w_1/\sqrt{A})$, where $\varepsilon=q/E_r$ is the elastic strain, q is the average pressure over the contact area, E_r is the reduced elastic (Young's) modulus, A is the projected area of the contact, w_1 is the deformation in elastic indentation by a flat punch. On this basis a relation is obtained between the reduced hardness H and unreduced hardness H_h, which depends on the ratio $w_s/w_1=m_s$; w_s is the elastic deformation along the perimeter of the indent, and $m_s\cong 0.78$. It is shown that the correction ΔE_r to the elastic modulus E_r determined from the condition of linearity of the initial part of the unloading diagram, is $\Delta E_r=0.27(\Delta P/P_m)$, where ΔP is the value used in the calculation of E_r for the length of the linear part of the diagram, reckoned from the maximum load P_m. It is shown that for metallic construction materials of medium hardness one has $q=HM$, where HM is the Meyer hardness. With increasing HM and increasing angle φ at the tip of the indenter, the ratio HM/q grows by an exponential law. © *1999 American Institute of Physics.* [S1063-7842(99)00907-1]

INTRODUCTION

In domestic practice the terms reduced (H) and unreduced (H_h) (micro)hardness have been regulated by the the GOST 9450-76 (ST SÉV 1195-78) standard and are determined in the first case by its conventional measurement from the projection of the loaded plastic indent and in the second case (H_h) from the displacement h (depth of the indent). Here H_h is measured continuously as the load P on the indenter is increased, by recording three variables: P, h, and the time t (Ref. 1). In the rest of the world no such terminology has been developed,[2–8] even though there is an obvious need to distinguish this sort of test from the conventional measurement of the (micro)hardness. Analysis and experiments show[1] that the quantities H and H_h characterize two independent properties of a material, and their ratio is a new indicator of its physicomechanical properties. The efficacy of testing in which the $P-h-t$ diagram is recorded is enhanced many-fold and can provide comprehensive information not only about the physicomechanical properties but also about the structural characteristics of the material.[1,9–18] The limiting sensitivity in the recording of the displacement and the load are of the order of 1 nm and 1×10^{-4} N, respectively,[2–7] which permits investigation not only of thin films, coatings, and individual phases but also the structure of a single-phase material, with a resolution of the order of 10 nm. At such a resolution it becomes possible to investigate phase transformations in silicon on the basis of an analysis of the features of the hysteresis loop.[7,16–18]

HOOKE'S LAW DURING INDENTATION

A typical $P-h$ diagram is shown in Fig. 1. Parts *1–3* depend, for a fixed shape of the indenter, on three properties of the material: the resistance to plastic deformations (curve *1*), creep or stress relaxation (curve *2*), and the elastic properties (curve *3*). The hysteresis loop of width δ is recorded under repeated loading of the same indent. It determines the intensity of the accumulation of reversible plastic deformations under cyclic loading and, hence, the intensity of wear and the kinetics of crack growth in fatigue.

In the analysis of the $P-h$ diagrams a decisive role is played by the elastic deformations in the indentation, and the

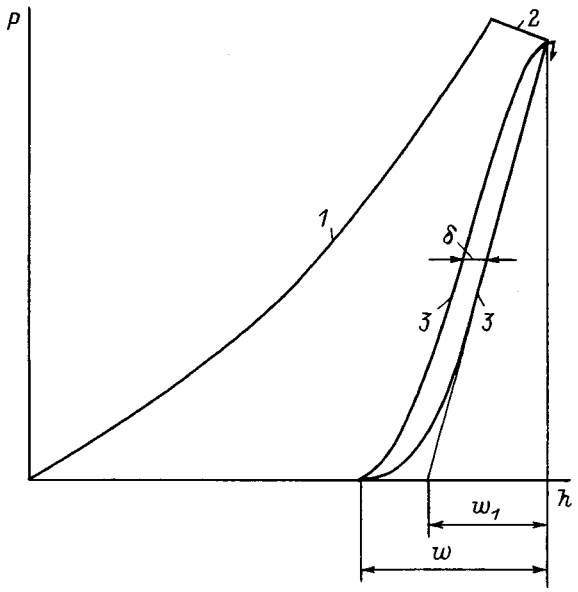

FIG. 1. Typical indentation diagram with three segments: *1* — loading and continuous measurement of the hardness H_h as a function of h; *2* — hold under load (relaxation or creep regime), *3* — unloading with repeated loading and recording of a hysteresis loop of width δ. The scales along the P and h (or w for unloading) axes depend on the hardness H_h of the material.

 © 1999 American Institute of Physics

beginnings of a theory of such testing based on their analysis are being worked out.[9–12] A number of properties have been discovered by which an analytical relation between H and H_h can be established and experimentally confirmed.[12]

Figure 2 illustrates the elastic recovery of an indent, including that in the material forming a noticeable ridge of extruded material (rim) around it, with a height Δh. Because the normal recovery along the perimeter of the indent is less than at its center, the angle of the unloaded indent increases.[19,20] It is noteworthy that the no change in the transverse dimension a of the indent is observed.[20,21]

Let us digress from the shape of the indenter and discuss its average pressure HM (Meyer hardness), which is distributed according to an *a priori* unknown law over the area of the plastic indent. An analysis[10] based on the known solutions of contact problems in the theory of elasticity[22] shows that the initial stage of the elastic unloading of the plastic indent, which commences after the creep on segment *2* has practically ceased,[10] is independent of the specific character of the pressure distribution over its area and is determined by the level of hardness and the elastic modulus E of the material.

Including allowance for the stiffness dP/dh_c of the device, the formula for E becomes[10]

$$E=\frac{1-\nu^2}{\frac{2-\sqrt{A}}{\sqrt{\pi}}\left(\frac{dh}{dP}-\frac{dh_c}{dP}\right)-\frac{1-\nu_i^2}{E_i}}, \tag{1}$$

where A is the projected area of the indent, and ν, ν_i, E, and E_i are the Poisson's ratios and elastic moduli of the materials of the sample and indenter.

To describe the process of unloading of the plastic indent, it is advisable to place the origin of the coordinate system at the point where the unloading starts, with the maximum values $P=P_m$ and $h=h_m$ and to change from h to the symbol w used in mechanics. When this is done, expression (1) assumes the more compact form

$$\frac{dP}{dw}=S=\left(\frac{2}{\sqrt{\pi}}\right)\sqrt{A}\cdot E_r, \tag{2}$$

where S is the stiffness, including the stiffness dP/dh_c of the device,[10] E_r is the contact elastic modulus: $(1/E_r)=(1-\nu^2/E)+(1-\nu_i^2)/E_i$.

A circular indent and a square indent with a side of $\sqrt{A}$ are practically equivalent:[10,22,23] in the case of uniform pressure for the square indent the stiffness s decreases by 0.6% according to the estimate of Ref. 10 and by 1.2% according to Ref. 23 (calculation by the finite-element method; for the trihedral indent made by a Berkovich pyramid this decrease reaches 3.4%).

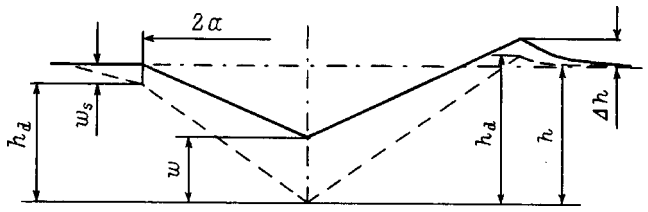

FIG. 2. Relationships among the different components of the elastic deformations in an indent, both with allowance for the rim of height Δh (to the right of the axis) and without allowance for the rim (to the left of the axis). The dashed and solid curves are contours of the indent under load and after removal of the load.

In analyzing the relationship between H and H_h it is important to pay attention to the not-so-obvious circumstance that the displacement h on the diagram does not respond to the value of E_i and is equal to the displacement recorded for an absolutely stiff indenter. This follows from the condition that at the instant the indenter touches the surface of the sample it is not yet loaded, and so the subsequent dependence $P(h)$ is determined solely by the geometry of the indenter (instrument), since the transverse dimension of the indenter, like that of the indent, does not change on loading.[20,21] Consequently, h is the instrumental displacement, set by the geometry of the indenter. Another feature of this displacement is that it is measured from the initial surface of the sample, whereas in measurements of the reduced hardness the transverse dimension of the indent is measured with allowance for the rim height Δh. Therefore, the relation between H and H_h depends solely on the normal elastic deformation w_s along the perimeter and the rim height Δh. Then, by definition, the relation between H and H_h in an indent obtained at a fixed load P_m is written in accordance with Fig. 2 in the form[1,12]

$$\left(\frac{H}{H_h}\right)^{1/2}=\frac{h}{h_d}=\frac{h_d+w_s-\Delta h}{h_d}. \tag{3}$$

An analysis of the known solutions of the contact problem in the theory of elasticity[22,24] has been systematized in Table I,[1,12] where the following additional notation has been adopted: $P_e=P_m/A$ is the mean pressure over the area of the indent, w_1 is the deformation in an elastic indentation by a flat punch having a fixed area A and product $E_r\sqrt{A}$, m is an

TABLE I. Basic relations between the elastic deformations in circular and square indents.

Means of loading	Pressure distribution law	$w/w_1=m$	w_s/w_1	$-u_s/w_s$
Punch with flat end	$p=P/(2\pi ax)$	1.0	1.0	
Uniform pressure	$p=p_e$	4π	0.81	$\pi(1-2\nu)/8(1-\nu)$
Conical punch	$\exp(2p/p_e)=(1+x)/(1-x)$	2	$2-(4/\pi)$	
Spherical indenter	$p=3p_ex/2$	1.5	0.75	$4(1-2\nu)/3\pi(1-\nu)$
Concentrated force	$p\to\infty$			$(1-2\nu)/2(1-\nu)$
Square indent	$p=p_e$	$4/\pi$	0.635*	
			0.865**	

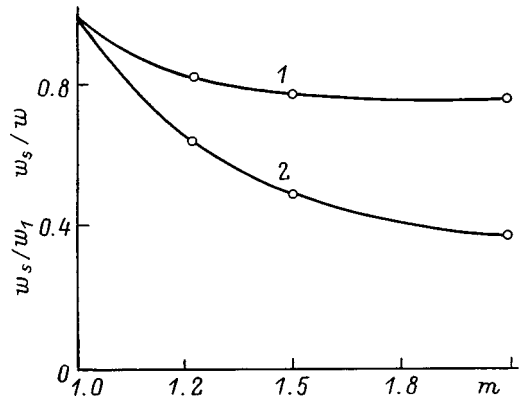

FIG. 3. Analysis values of the elastic deformations in circular and square indent of dimension $2a$, as functions of the index m of the pressure distribution over the area of the indent (1 — w_2/w_1, 2 — w_s/w).

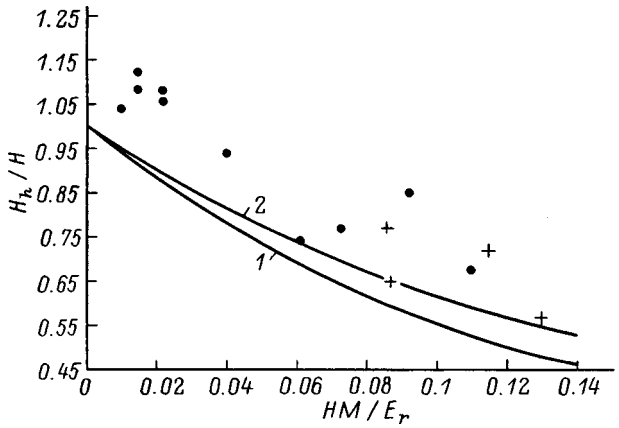

FIG. 4. Calculated (solid curves) and experimental (points) plots of H_h/H versus HM/E_r for $w_s/w_1=0.78$ (*1*), 0.61 (*2*); + — data of Ref. 26.

index characterizing the pressure distribution over the contact area, and u_s is the radial deformation along the perimeter of the indent. This last deformation has the opposite sign, so that the dimension $d=2a$ increases during unloading, but the increase is practically zero, as follows from the formula in Table I and as is confirmed by experiment:[20,21] $w_s/d \ll 1$; $w_s \cong -5u_s$; $u_s/d \cong 0$.

For a flat punch it follows from Eq. (2) that P_m and w_1 are linearly related:

$$P_m = w_1 E_r d. \tag{4}$$

The derivative dP/dw at the initial point of the elastic unloading curve of a plastic indent intercepts the h axis at a distance w_1 from the value h_m. This is the minimum possible value at a given hardness HM and elastic modulus E (Refs. 1, 12, and 25). It follows from Fig. 1 that there is a simple relation between w and w_1 if the unloading curve is approximated[4,11] by $P=Bw^m$ (see Table I):

$$m = w/w_1, \tag{5}$$

where m can vary in the interval from 2 to 1.

In elastic indentation by a cone one has $m=2$ and $w=2w_1$, for a spherical indenter $m=1.5$ and $w=1.5w_1$, for a flat punch $m1$ and $w=w_1$, and in the model of a uniform pressure distribution $m=4/\pi$ and $w \cong 1.27w_1$. When the deformation w_s is normalized by w_1 we obtain, according to Table I, its extrapolated dependence on m (Fig. 3).[12] For plastic indent made by a Vickers pyramid one has $m=1.3-1.5$ (Refs. 1 and 12). We see from Fig. 3 that for a circular indent in this interval of m values the function $(w_s/w_1)\times(m)$ is close to a constant, in contrast to the function $(w_s/w)(m)$. For a Berkovich pyramid a wider interval of m values, 1.2–1.6, was obtained in Ref. 4.

Table I also gives two values for the deformation along the perimeter of a square indent during its uniform loading — w_{sd} (denoted by *), measured at the ends of the diagonals, and w_{sm} (denoted by **), measured in the middle of the square. In a linear approximation, proceeding from the known deformation w_s, we obtain the following expressions for a circular indent in the interval $m=1.27$–1.5:

$$w_s/w_1 = 0.81 - 0.26(m-1.27), \tag{6}$$

$$w_{sd}/w_1 = 0.635 - 0.204(m-1.27), \tag{7}$$

$$w_{sm}/w_1 = 0.865 - 0.28(m-1.27). \tag{8}$$

If one should start from the averaged deformation $w_{sc}=w_s$ along a side of the square indent, the approximation formula (6) would apply. It is preferable to formula (7), since present-day hardness meters measure the real projected area of the indent with allowance for the rim-related distortions of the square shape.

In Fig. 4 the experimental values of H_h/H are compared with the values calculated according to formula (3). The deformation on the contour of the indent was found using formula (6) for a mean value $m=1.4$, which gives $w_s/w_1=0.78$ for a circular indent. The height of the rim was taken equal to zero. Also shown in Fig. 4 are the results of experimental measurements, including the data of Ref. 26. The position of the experimental points shows that if the values of m and of the blunting of the tip of a sharp indenter are measured more carefully, then the ratio H_h/H can be used to measure the height of the rim, which depends on the coefficient of strain hardening of the material and determines its basic mechanical characteristics. More complete results of experimental measurements are presented in Table II, which also gives the ratio of the differentials of the elastic work dA_e to the total work dA in the indentation; this ratio was analyzed in Ref. 25. Note the different deviation from the calculated curve for the two points corresponding to two types of glass: construction-grade soda–lime glass, and KV-grade quartz glass. The calculated rim heights for these two differ by more than 10%. Experiments show that this difference can be much larger for metallic glasses of different compositions. Careful interference measurements of the rim height and the profiles of unloaded indents were made in Ref. 27. It follows from an analysis of these data that on traditional construction materials the rim height can alter the ratio H_h/H by more than a factor of 1.2.

Taking into account that $P_m/A=HM$, we obtain from relation (4)

$$\frac{HM}{E_r} = \frac{2}{\sqrt{\pi}}\frac{w_1}{\sqrt{A}} = \frac{4}{\pi}\frac{w_1}{d}. \tag{9}$$

TABLE II. Physico-mechanical properties of materials.

No.	Material	HV, GPa	HV_h/HV	E, GPa	HM/E	HM/E_r	δ/w_1, %	dA_e/dA
1	30KhGSA	2.50	1.08	210	0.013	0.015	14.0	0.96
2	12Kh18N9T	1.60	1.04	204	0.085	0.010	7.0	0.31
3	45	2.46	1.12	195	0.013	0.015	14.0	0.97
4	L62 brass	1.33	1.08	98	0.014	0.015	9.0	0.62
5	D16 Duralumin	1.61	1.06	80	0.021	0.022	4.5	0.45
6	Ti alloy	2.47	1.08	130	0.020	0.022	7.0	0.70
7	Ni–Al–Cr	7.10	0.94	210	0.037	0.040	6.0	1.02
8	Glass	6.25	0.85	74	0.091	0.092	8.5	3.16
9	KV glass	6.95	0.68	71	0.110	0.110	5.0	1.99
10	$FeAl_3$	7.20	0.77	110	0.071	0.073	1.0	0.28
11	GaAs	5.93	0.74	118	0.055	0.061	⋯	0.46
12	Si	11.00	0.65	150	0.079	0.087	⋯	0.62
13	α-Al_2O_3	25.00	0.77	430	0.063	0.086	⋯	0.65
14	BN	45.00	0.57	550	0.088	0.130	⋯	0.87
15	B	32.00	0.72	400	0.080	0.115	⋯	0.86

Relation (9) does not change when HM is replaced by the true mean pressure q and true contact area A. We obtain the following expression for Hooke's law for any, elastic or plastic, local contact loading:

$$\frac{q}{E_r}=\frac{2}{\sqrt{\pi}}\frac{w_1}{\sqrt{A}}. \tag{10}$$

For tensile testing, Hooke's law is $s/E=\varepsilon$. Consequently

$$\varepsilon=\frac{2}{\sqrt{\pi}}\frac{w_1}{\sqrt{A}}. \tag{11}$$

ELASTIC CONTACT AREA

We have taken $HM=q$ and $d=d_c$, where d_c is the dimension including the plastic and elastic contact zone around the perimeter of the indent. Analytical estimates made in Ref. 28 show that this approximation holds in many cases. Let us use the data of Table I to obtain a relation between HM and q. In addition, we must find the conditions for a more precise experimental determination of the elastic modulus ε, as discussed in Refs. 3–6 and 23.

In elastic indentation by a pyramid or an equivalent cone, the elastic contact along the perimeter of the plastic indent is described by the equations

$$P=\frac{2}{\sqrt{\pi}}\frac{E_r}{\cot\varphi}w^2, \quad w=\frac{\pi}{4}d\cot\varphi, \tag{12),(13}$$

and the variable pressure p is distributed as a function of the polar coordinate r by the law

$$p=\frac{p_e}{2}\ln\frac{1+x}{1-x}, \tag{14}$$

where 2φ is the angle at the vertex of the cone, p_e is the mean pressure for a purely elastic contact, x is a quantity which determines the dimension r along the perimeter of which the pressure p acts,

$$x^2=1-r^2/a^2, \tag{15}$$

where $a=d_c/2$ is the radius of the perimeter on which the pressure p is equal to zero.

We are interested in the boundary of the plastic region of dimension r on which the pressure $p=HM$. Let the same load P_m be distributed in one case over the region a and in the other over the region r. Dividing the numerator and denominator of the right-hand side of (14) by P_m, we get

$$x^2=1-(q/HM), \tag{16}$$

where $q=P_m/(\pi a^2)$ is the mean pressure over the total elastic and plastic contact area, and $P_m/(\pi r^2)=HM$ is the hardness measured from the unloaded plastic indent.

The unloading along a perimeter of width $\Delta r=a-r$ is elastic. For a conical indenter we substitute (13) into (12) and divide by A to get

$$p_e=\frac{1}{2}E_r\cot\varphi. \tag{17}$$

The pressure distribution in the plastic contact zone is close to uniform.[29] Therefore, we suppose, with a certain margin of error, that the pressure p in the expression for the dimension x is equal to HM (analysis of the pressure distribution using the data of Ref. 29 gives an order-of-magnitude estimate of $HM=1.1q$). After substituting this value into (14) and doing some straightforward manipulations we obtain

$$\exp\left(\frac{4}{\cot\varphi}\frac{HM}{E_r}\right)=\frac{1+x}{1-x}. \tag{18}$$

With the values found for x, formulas (16) and (15) give values for the ratios q/HM and $\Delta r/a$, and, as can be seen from a comparison of (16) and (15),

TABLE III. Relation between the mean contact pressure q and the Meyer hardness HM for a Vickers pyramid for different values of HM/E ($HM/E=y/11.32$).

y	$\exp y$	HM/q	d_c/d
0.1111	1.1175	1.003	1.001
0.3333	1.3956	1.028	1.014
0.5555	1.7428	1.079	1.039
0.7777	2.1765	1.159	1.077
1.1110	3.0375	1.342	1.158
1.4443	4.2392	1.619	1.272
1.7777	5.9161	2.021	1.422
2.2221	9.2266	2.834	1.683

$$\left(\frac{HM}{q}\right)^{1/2}=\frac{d_c}{d}=\frac{r+\Delta r}{r}. \tag{19}$$

Solutions of equations (18) and (19) are presented in Table III. The quantity $(4/\cot\rho)(HM/E_r)$ is denoted as y. For a Vickers pyramid $2\varphi=136°$, and so $4/\cot\varphi=11.11$. Then $y=11.11HM/E_r$. For this particular case the values of HM/q and d_c/d are given for various HM/E_r. It is seen from the table that for $HM/E_r>0.03$ the ratio $HM/q>1.03$, and it increases rapidly with increasing HM/E_r. Let us also look at the calculations for another particular case, typical for steels: $E=200$ GPa, $\nu=0.25$, $E_i=1141$ GPa, $\nu_i=0.07$ (Ref. 4). For these values we have $HM/E_r=1.19HM/E$. Then values $HM/q>1.03$ are already reached at $HM/E>0.027$. Similar calculations can be done for other angles φ. For $2\varphi>136°$ it is convenient to use the approximation $\tan\psi=\psi$, where $\psi=(90°-\varphi)$. If ψ is decreased by a factor of 5, so that $\psi=(22/5)°$, then the ratio of the tangents of these angles is 5.2, i.e., it increases in comparison with the ratio of the tangents of these angles by only 4%. For such an indenter, values $HM/q>1.03$ are reached at $HM/E>(0.03)/5=0.006$.

Let us examine indenter shapes further. For an elastic process of indentation by a parabolic (spherical) indenter, according to Table I, we write in analogy with (14)

$$p=(3/2)p_e x. \tag{20}$$

On the other hand, for an elastic process

$$P=d_c^3E_r/(6R). \tag{21}$$

After dividing by A, we get

$$p_e=\frac{4}{3\pi}\frac{a}{R}E_r. \tag{22}$$

Under the condition $p=HM$ the substitution of (20) into relation (22) yields

$$\frac{q}{HM}=1-\left(\frac{\pi}{2}\frac{HM}{E_r}\frac{R}{a}\right)^2. \tag{23}$$

Let us compare the two indenters in more detail. Taking the ratio of Eq. (20) to Eq. (14), making use of the fact that for small x one can write $\ln[(1+x)/(1-x)]\cong 2x$, and substituting p_e from Eq. (22) into (17), we obtain

$$\frac{p_s}{p_c}=\frac{4}{\pi}\frac{a}{R}\frac{1}{\cot\varphi}, \tag{24}$$

where the subscript s and c denote spherical and conical indenters, respectively.

Here a/R and $\cot\varphi$ characterize the angle between the normal to the surface of the sample and the tangent to the profile of the indent at a point located along its perimeter. For equal values of φ we obtain from the geometric relations

$$(a/R)(1/\cot\varphi)=\sin\varphi. \tag{25}$$

As φ increases, the value of expression (25) approaches unity, and, hence,

$$p_s/p_c\cong(4/\pi). \tag{26}$$

From an analysis of relations (24)–(26) one can conclude that to a first approximation the width of the elastic girdle along the perimeter of the plastic indent is the same for indents made by conical and spherical indenters at the same angles of indentation $\psi(90°-\varphi)$.

The accuracy to which E and, hence, the ratio H_h/H are determined is higher if the initial segment of the unloading curve of the plastic indent is linear. The existence of this indicator was pointed out in Refs. 1, 3, 9, and 10, and the necessity of taking the nonlinearity of this segment into account was demonstrated in Refs. 4 and 16. In this connection let us analyze the flat-punch model, which is also given in Table I. In this model the elastic contact is described by the linear relation (4). The sharp edge of the punch is dulled on the segment of the transition from d to d_c. The geometry of this transition is specified by the geometry of the real conical or spherical indenter. Consequently, the unloading on the segment on which the dimension of the contact area decreases from d_c to d is described by the relations already considered, which also determine its length. Comparing formulas (19) and (12) for a conical indenter with relation (13) substituted into it, we conclude that the load P in the unloading from dimension d_c to d is given by $P/P_m=q/HM$. It follows from the foregoing discussion that this segment on the unloading curve become appreciable for $HM/E_r>0.03$ and grows rapidly as this ratio increases. From a comparison of formulas (26) and (21) it follows that this segment increases for a spherical indenter.

If the ratio E_i/E is large, then the further elastic unloading approaches the model of a rigid punch, since all points on the plastic indent area recover by the same amount Δw_s, equal to the rise of the indenter. The distribution of the counterpressure in the unloading will vary in accordance with the formula

$$p_x=\Delta P/(2\pi a^2/x)=p_e/(2x), \tag{27}$$

where ΔP is the decrease of the load during unloading, and p_x is the variable value of the counterpressure.

On the part of the unloading from d_c to d the pressure along the perimeter of the indent declines to zero and then increases to a value $p_x\cong -HM$. Under this condition Eq. (27) becomes $-HM=p_e/(2x)$, or

$$x=p_e/(2HM), \tag{28}$$

where p_e increases during unloading, from zero to HM at the completion of the unloading, while x increases from 0 to 0.5.

Substituting p_e for q in (16), we obtain at $x=0.5$ the limiting value of the counterpressure

$$p_{em}=0.75HM. \tag{29}$$

Up to this limiting value of the counterpressure the dimension of the contact area remains unchanged and equal to d. Therefore, $p_e/HM=\Delta P/P_m$. As the unloading progresses, reverse plastic deformations arise, and along the perimeter of the indent there is an increase in the width $\Delta r/r$ of the girdle over which a uniform pressure with a value q is distributed. For such a pressure distribution the pressure $w_e=(4/\pi)w_1\cong 1.27$, where $4/\pi=m_e$, and the unloading also takes place by a linear law similar to (4), but here the slopes of the unloading branches are different: $dP/dw_1(4/\pi)dP/dw_e$. In the linear relation (4) and the others like it, the quantities d and w_1 or d and w_e are interchangeable: the value of P does not change if one of these is increased and the other decreased by the same factor. This property allows one to treat the unloading of a plastic indent as a sum of two linear processes in which the parameters d_1 and d_2 vary but their sum remains equal to d. For one of theses processes $m_e=\pi/4$. Their specific contribution is determined from the boundary conditions. At the start of the unloading, condition (2) holds, and accordingly $dw/dP=w_1/P_m$. At this time the girdle over which the pressure is distributed uniformly has zero width. If upon completion of the unloading the pressure is distributed uniformly over the entire area of the plastic indent, then we obtain a deformation $w_e=(4/\pi)w_1$. At an intermediate position the contribution of this deformation is given by $\Delta w_e=(4/\pi)\Delta w_1(\Delta P/P_m)$. The specific contribution of both processes is the sum

$$\Delta w=\frac{4}{\pi}\Delta w_1\frac{\Delta P}{P_m}+\Delta w_1\frac{P}{P_m}=\Delta w_1\left(\frac{4}{\pi}\frac{\Delta P}{P_m}+\frac{P}{P_m}\right). \tag{30}$$

It follows from a comparison of relations (30) and (2) that the expression in parentheses in the right-hand side of (30) is equal to the correction to the value E_w found under the assumption that ΔP is a straight line segment. The true value of the modulus E_r for this sum of two linear processes, with each of their properties (4) taken into account, is found from the equation

$$d\cdot E_r=\frac{\Delta P}{\Delta w}\left(\frac{4}{\pi}\frac{\Delta P}{P_m}+\frac{P}{P_m}\right). \tag{31}$$

At the initial point of the unloading process $\Delta P/P_m=0$, $P=P_m$, and the correction is zero. As P decreases and ΔP increases, this correction increases linearly. We obtain for E_r the expression

$$E_r/E_w=1+[(4/\pi)-1]\Delta P/P_m, \tag{32}$$

where E_w is the value found from relation (2) along the segment of the unloading curve of length ΔP that was assumed to be a straight line segment.

In the model of the unloading of a rigid punch the limiting value was $\Delta P_m=0.75$. In a real process this is less and depends on several parameters. If E/E_i increases, then the contribution of the elastic deformation of the indenter increases, as a result of which the flat model of the punch is transformed into a convex model, and ΔP_m decreases. The real pressure distribution over the area of the indent is characterized by a rise toward the center of the indent, which also decreases ΔP_m. There is a similar influence of deviations from a circular shape of the indent. It increases in the sequence from conical or spherical indenter to Vickers, Berkovich, and Knoop pyramids. With increasing number of repeated loading cycles the reverse plastic deformation leads to a self-matching of the surfaces of the indent and indenter, the reverse plastic deformation becomes exhausted, and ΔP_m also tends toward zero, and the hysteresis loop changes accordingly. The intensity of the reverse plastic deformation processes and their exhaustion as the number of repeated loading cycles increases also depend on the strain-hardening coefficient of the material. As it increases, so does the dimension D_z/d of the plastic zone under the indent, the intensity of the reverse plastic deformations decreases owing to their retardation by the plastic zone, and a greater number of repeated unloading cycles is needed in order to exhaust them.

CONCLUSION

In summary, the branch of the first unloading of an indent consists of three segments. The initial segment involves elastic unloading of a girdle along the perimeter of the indent. Its length and index m depend on HM/E_r and the type of indenter used. For a conical or pyramidal indenter $m=2$, and for a spherical indenter $m=1.5$. On the second segment the unloading occurs in accordance with a rigid-punch model, with an accumulation of reverse plastic deformation as the unloading proceeds. The length of this segment decreases with increasing ratio E/E_i and asymmetry of the indent. This last is most pronounced for a Knoop pyramid. On the final segment the contact area of the indenter with the surface of the plastic indent begins to decrease. The concentration of the pressure at the center of the indent causes the index m and the curvature of this segment to increase. After 10–15 repeated loading cycles the reverse plastic deformation has decreased and stabilized. The index m increases with increasing ratio E/E_i and strain-hardening coefficient k. The influence of k is due to the fact that the length of the second segment decreases with increasing D_z/d, while on the third segment the index m increases and pressure is increasingly concentrated at the center of the indent. This scheme of the unloading processes and the technique of refined determination of the elastic modulus agree with the data of Ref. 4, where the following values of m were obtained: for tungsten and aluminum 1.51 and 1.38, respectively, and for sapphire, quartz, and fused silica 1.47, 1.43, and 1.25, respectively.

The experimental measurements presented in Table II and discussed above also agree with the analysis given for the relation between the reduced and unreduced hardness and the determination of the height of the rim.

[1] S. I. Bulychev and V. P. Alekhin, *Materials Testing by Continuous Indentation* [in Russian], Mashinostroenie, Moscow (1990), 224 pp.
[2] J. B. Petica, R. Hutckings, and W. C. Oliver, Philos. Mag. A **48**, 593 (1983).
[3] M. F. Doerner and W. D. Nix, J. Amer. Cerm. Soc. **1**, No. 4–5, pp. 601–606.
[4] W. C. Oliver and G. M. Pharr, J. Mater. Res. **7**, 1564 (1992).
[5] Y. Mutakami, K. Tanaka, M. Itokazu, and A. Shimamoto, Philos. Mag. A **69**, 1131 (1994).
[6] N. V. Novikov, M. A. Voronkin, S. N. Dub *et al.*, Diamond Relat. Mater. **6**, 574 (1997).
[7] S. N. Dub, Sverkhtverdye Mater., No. 6, pp. 17–25 (1997).
[8] N. G. Chechemin, J. Bottiger, and J. P. Krog, Thin Solid Films **304**, 70 (1997).
[9] A. P. Ternovskiĭ, V. P. Alekhin, M. Kh. Shorshorov *et al.*, Zavod. Lab. **39**, 1242 (1973).
[10] S. I. Bulychev, V. P. Alekhin, M. Kh. Shorshorov *et al.*, Zavod. Lab. **41**, 1137 (1975).
[11] S. I. Bulychev, V. P. Alekhin, M. Kh. Shorshorov, and A. P. Ternovskiĭ, Problemy Prochnosti, No. 9, pp. 79–83 (1976).
[12] V. P. Alekhin and S. I. Bulychev, Fiz. Khim. Obrab. Mater., No. 3, pp. 134–178 (1978).
[13] V. P. Alekhin and S. I. Bulychev, Dokl. Akad. Nauk SSSR **238**, 1328 (1978) [Sov. Phys. Dokl. **23**, 140 (1978)].
[14] S. I. Bulychev and V. N. Malyshev, Fiz. Khim. Obrab. Mater., No. 3, pp. 98–102 (1986).
[15] S. I. Bulychev and V. P. Alekhin, Zavod. Lab. **53**, 76 (1987).
[16] G. M. Pharr, W. C. Oliver, and D. R. Clarke, Scr. Metall. **19**, 881 (1989).
[17] G. M. Pharr, W. C. Oliver, and D. S. Harding, J. Mater. Res. **6**, 1129 (1991).
[18] G. M. Pharr, Mater. Res. Soc. Symp. Proc. **239**, 301 (1992).
[19] M. P. Markovets, Zh. Tekh. Fiz. **19**, 371 (1949).
[20] N. A. Stillwell and D. Tabor, Proc. Phys. Soc. London **78**, 169 (1961).
[21] H. Bückle, Z. Metallkd. **45**, 623 (1954).
[22] A. I. Lur'e, *Three-Dimensional Problems in the Theroy of Elasticity* [in Russian], GTTI, Moscow (1955), 408 pp.
[23] R. B. King, Int. J. Eng. Sci. **3**, 1657 (1987).
[24] I. N. Sneddon, Int. J. Eng. Sci. **3**, 47 (1965).
[25] M. Kh. Shorshorov, S. I. Bulychev, and V. P. Alekhin, Dokl. Akad. Nauk SSSR **259**, 839 (1981) [Sov. Phys. Dokl. **26**, 769 (1981)].
[26] B. A. Galanov, O. N. Grigor'ev, Yu. V. Mil'man *et al.*, Dokl. Akad. Nauk SSSR **274**, 815 (1984) [Sov. Phys. Dokl. **29**, 146 (1984)].
[27] M. M. Khrushchov and E. S. Berkovich, *Determination of Wear of Machine Parts by the Method of Artificial Baselines* [in Russian], Izd. Akad. Nauk SSSR, Moscow (1959), 219 pp.
[28] O. N. Grigor'ev, Yu. V. Mil'man, and V. N. Trefilov, Dokl. Akad. Nauk **274**, 815 (1984).
[29] A. Yu. Ishlinskiĭ, Prikl. Mat. Mekh. **3**, 201 (1944).

Translated by Steve Torstveit

Electron emission during pulsed polarization switching of ferroelectric ceramics

A. N. Pavlov, I. P. Raevskiĭ, and V. P. Sakhnenko

Scientific-Research Institute of Physics at Rostov State University, 344104 Rostov-on-Don, Russia
(Submitted March 18, 1998; resubmitted June 19, 1998)
Zh. Tekh. Fiz. **69**, 49–52 (July 1999)

It is shown that in the pulsed polarization switching of polycrystalline ferroelectrics in high fields the external field is damped by the electronic subsystem associated with the crystallite boundaries. Therefore, the reorganization of a ferroelectric system with a high density of surface states at the crystallite boundaries is not irreversible, and the system returns to its former state after the polarization-switching voltage is removed. Because of this, electron emission from the surface of a ferroelectric ceramic can be brought about by unipolar pulses. © *1999 American Institute of Physics.* [S1063-7842(99)01007-7]

INTRODUCTION

Electron emission from ferroelectrics is observed under various kinds of external influences (light, heat, mechanical, etc.). There has recently been considerable interest in the phenomenon of electron emission from ferroelectrics under the influence of an external polarization-switching electric field.[1–5] It has been established that this emission can be obtained both from crystals and from ceramics, that it has a unipolar character, that it last a short time (of the order of tens of nanoseconds), that the emission capability can recover after the polarization-switching field is removed, and that the amount of charge emitted correlates with the value of the spontaneous polarization P_s. In Refs. 1–5 the questions of the formation of the emission potential and different microscopic mechanisms of emission were discussed for ferroelectric crystals. However, there have been no reports of models that take into account the specifics of the emission processes during the polarization switching of a ferroelectric ceramic, for which promising experimental results have been obtained.[3] At the same time, our studies[6–8] of charge transport processes have shown that the formation of the internal electric field in ferroelectric ceramics has a number of fundamental differences from that in crystals on account of the presence of intercrystallite boundaries, and that these differences affect the formation of the emission potential.

DESCRIPTION OF THE MODEL

Let us consider the behavior of a polarized polycrystalline ferroelectric in an external polarization-switching field (Fig. 1). In polycrystalline ferroelectrics there are a number of interacting subsystems that respond to the external electric field, among them: 1) ferroelectric (ionic), 2) electronic in the bulk of the crystallites, 3) electronic, related to the surface of the crystallites. The interaction with this last subsystem is due to the fact that under certain conditions of synthesis and annealing, localized surface states of the acceptor type arise at the interfaces between crystallites[6,9] the screening of these filled states owing to the redistribution of electrons in the bulk of the crystallites gives rise to the formation of an electron-depleted surface layer near the boundaries of the crystallites; this layer is of the Schottky type and is positively charged[9] (Fig. 2). The appreciable electric field existing in the Schottky regions has a large influence on the distribution of the polarization P.[7] In the equilibrium state of the ferroelectric the spontaneous polarization P_s is screened by the electronic subsystem. If the ferroelectric subsystem is taken out of its former state (e.g., by a sudden polarization reversal), the electronic subsystem will adjust to the new state of the ferroelectric subsystem, i.e., a rescreening will occur, which can come about through motion of the charges in both the internal and external circuits. The rescreening due to the motion of charges in the external circuit can be accompanied by the escape of electrons from the open surface of the ferroelectric, i.e., emission. The response times of the electronic (surface and bulk) and ionic subsystems to an external influence are different. If a sudden polarization reversal is brought about by a field $E \geqslant E_c$ (where E is the external electric field and E_c is the coercive field, e.g., for $BaTiO_3$ the theoretical value is $E_c \sim 10^7$ V/m), then the ferroelectric subsystem readjusts over $\tau_{c0} \sim 10^{-12}$ s. For $E < E_c$ the switching time of the ferroelectric system increases as $\tau_c \sim \tau_{c0} \exp(E_c/E)$ (Ref. 10). For example, for $E \sim 10^6$ V/m (Ref. 3) we obtain $\tau_c \sim 10^{-8}$ s. In order to observe any kind of switching effect it is necessary to satisfy the condition $\tau_c < \tau_i$, where τ_i is the time of the external polarization-switching pulse. The response time of the electronic bulk subsystem to an external influence can be estimated by proceeding from the Maxwell relaxation time $\tau_r = \varepsilon_0 \varepsilon \rho \sim 10^{-1}$ s (ε_0 is the permittivity of free space, $\varepsilon \sim 10^4$ is the dielectric constant, $\rho \sim 10^6\ \Omega \cdot$m (Ref. 11) is the electrical resistivity).

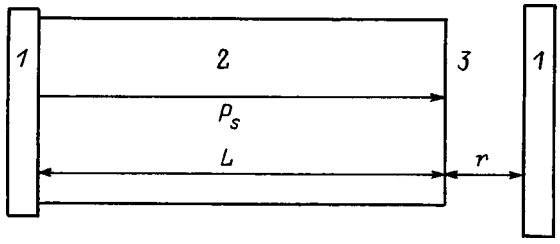

FIG. 1. Arrangement of the objects: *1* — electrodes, *2* — sample, *3* — gap.

 © 1999 American Institute of Physics

Owing to the large dielectric constant of a ferroelectric, the external voltage applied to the electrodes creates a significant electric field in the gap.[2] In order for an electron to leave the surface of a ferroelectric, it must undergo a transition from some surface state to a free state. At an average external electric field $E_{av} \sim 10^6$ V/m, $\varepsilon \sim 10^4$, a sample thickness $L \sim 10^{-3}$ m, and a gap thickness $r \sim 10^{-6}$ m (Fig. 2), the electric field in the gap will be $\sim 10^9$ V/m,[5] which causes electron emission[4] from the surface energy states with a significant activation energy (~ 0.5 eV) over the time $\sim 10^{-12}$ s necessary for the displacement of an electron to a distance of the order of one lattice constant under the influence of the external electric field. The integrated value of the emitted charge is limited by the density of surface states, which can reach $\sim 10^{18}$ m^2.

In the crystal the screening processes occurring in the external circuit due to emission of the surface electrons can lead to pinning of the new state of the ferroelectric system. However, in a polycrystalline material during a brief ($\sim 10^{-7}$ s) polarization-switching pulse the electrons in the bulk of the crystallites do not have sufficient time to react to the pulse, and as a result the electric fields due to this electronic subsystem in the Schottky regions remain the same as they were before the polarization-switching pulse was applied. Thus, after the polarization-switching pulse has ended, the ferroelectric system in a ceramic is subjected two two competing factors: the field of the emission-modified electronic system associated with the open surface of the sample tends to stabilize the new polarization direction, while the conserved fields of the Schottky regions tend to restore the previous state. The restoration of the initial state of the ferroelectric system plays an extremely important role in organizing an efficient emission process. It can happen that, after suppression of the new polarization-switching pulse of the previous polarity, the whole cycle is repeated anew. Therefore, the time of the polarization-switching pulse should be less than the time τ_r required for reorganization of the Schottky regions at the grain boundaries. Thus, in order for the emission current to be large and for the switching back to occur after removal of the polarization-switching field it is necessary to satisfy the condition $\tau_r > \tau_i > \tau_c > \tau_s$ and to have the restoring effects due to the Schottky regions be predominant over the emission-related effects that pin the new state. It is most important to consider the role played in these processes by N_s, the density of surface states at the crystallite boundaries, since that is primarily what determines the fields in the Schottky regions.

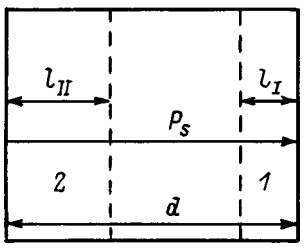

FIG. 2. Diagram of the electron-depleted surface layers: *1* and *2* are Schottky regions I and II.

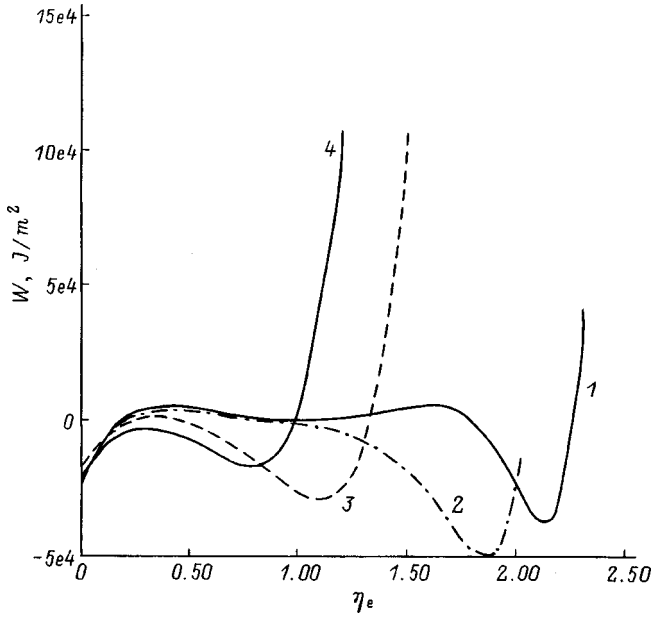

FIG. 3. Total energy W versus the degree of emission η_e in the absence of external voltage ($u=0$) for a ferroelectric crystal (*1*) and for ferroelectric ceramics with $N_s = 3\times10^{18}$ (*2*), 4×10^{18} (*3*), and 5×10^{18} m^{-2} (*4*).

RESULTS OF THE CALCULATIONS

Let us give some illustrative numerical estimates for barium titanate $BaTiO_3$, since electron emission during polarization switching has been observed for this compound[1] and the parameters of the thermodynamic potential have been well determined. To trace how the emission processes are influenced by the density N_s of deep surface states of the acceptor type at the crystallite boundaries, we have made some model calculations based on the thermodynamic parameters of $BaTiO_3$ inferred from the data of Ref. 12. In the calculations we used the following values: temperature $\theta = 300$ K, density of impurity states of the donor type in the bulk of the crystallite $N_d = 2\times10^{25}$ m^{-3}, activation energy of these donor states $E_d = 0.5$ eV, activation energy of the surface states $E_s = 1.2$ eV, and crystallite thickness $d = 10^{-6}$ m. Then the bulk conductivity in the crystallites is $\rho \sim 10^3 \Omega \cdot$m, which ensures satisfaction of the condition $\tau_r > \tau_i > \tau_c > \tau_s$. Therefore, in the polarization switching and the accompanying electron emission from the open surface of the ferroelectric, the distribution of the electron density and the Schottky regions in those crystallites which are located in the interior of the ferroelectric remain uncharged. Figures 3 and 4 show the results of calculations of a number of parameters as functions of the degree of emission, which is characterized by the quantity η_e, which is related to the charge ΔQ escaping from the surface of the sample by the relation $\eta_e = \Delta Q / P_s$. We assume that $qN_{s,fr} > P_s$ (q is the elementary electron charge, and $N_{s,fr}$ is the density of surface states on the free surface of the ferroelectric sample). Therefore, the surface charge $Q = qN_{s,fr}$ and the adjacent Schottky region together provide complete screening of P_s in the ground state. Since emission alters the surface charge on the free surface, the field in the bulk of the crystallite changes, and a new polarization $P = P_s(1-\eta_e)$ is estab-

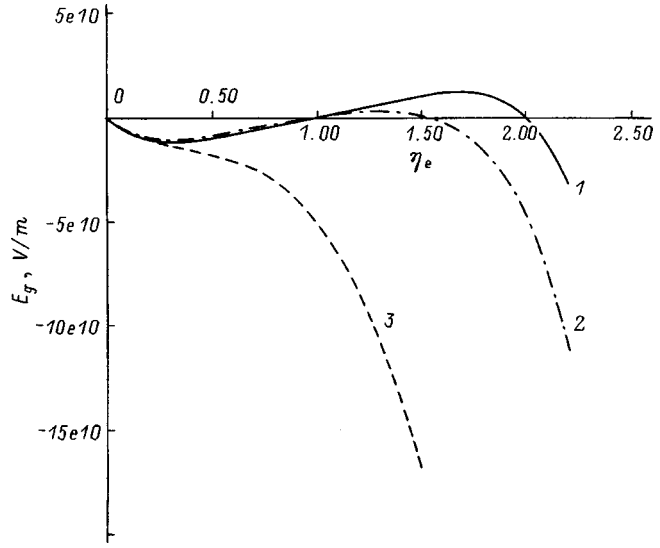

FIG. 4. Field E_g in the gap versus the degree of emission η_e at $u=0$ for a crystal (*1*) and for ceramics with $N_s=3\times10^{18}$ (*2*) and 5×10^{18} m^{-2} (*3*).

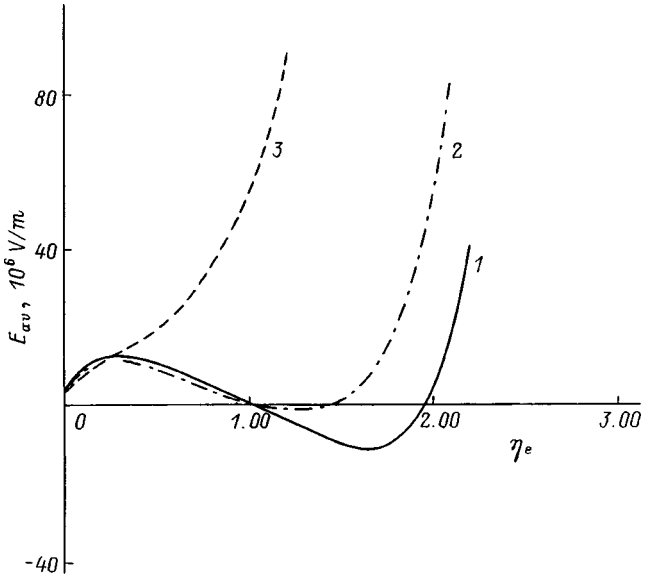

FIG. 5. Average field $E_{\rm av}$ versus η_e at a field in the gap $E_g=10^9$ V/m for a crystal (*1*) and for ceramics with $N_s=3\times10^{18}$ (*2*) and 5×10^{18} m^{-2} (*3*).

lished. Figure 3 shows the total energy per unit area of the sample together with the gap as calculated for different degrees of emission in the absence of an external voltage; this energy includes the energy of the ferroelectric and electronic subsystems and the energy of their interaction. It follows from Fig. 3 that for a crystal or ceramic with small N_s, polarization switching accompanied by emission leads to a new stable state of the system, i.e., is irreversible, since the total energy of the new state is lower than that of the initial state. At the same time, Fig. 3 shows that for a ceramic with large N_s, polarization switching accompanied by emission leads to an unstable new state, and the initial state is energetically favorable. The calculated dependence of the electric field E_g in the gap (Fig. 4) shows that a deviation of the external field (at zero potential difference between the electrodes) for a ceramic with a large density of surface states (curve *3*) gives rise to an electric field in the gap in the direction corresponding to restoration of the original surface charge ($E_g<0$); this promotes reversibility of the polarization switching process. Thus the calculations whose results are presented in Figs. 3 and 4 suggest that in the case of a large density of surface states the polarization switching is reversible at values of the degree of emission ($\eta_e>1$) for which the region of irreversible processes has been reached in a crystal or ceramic with a low density of surface states. At lower degrees of emission ($\eta_e<1$), where, in accordance with Fig. 4, a crystal or ceramic with a low density of surface states exhibits a tendency toward restoration of the initial state when the external polarization-switching voltage is turned off, this tendency is weaker for a ceramic with large N_s.

Let us now examine Fig. 5, which corresponds to an external field of 10^9 V/m, i.e., a field such that emission conditions will be attained in the gap. Plotted along the vertical axis is the average of the external electric field $E_{\rm av}$, and along the horizontal axis is η_e, a measure of the degree of emission. The figure shows the value of $E_{\rm av}$ that is required in order for emission to continue at a given η_e. The negative values of $E_{\rm av}$ for a crystal and for a ceramic with small N_s indicate that during polarization switching at that stage, which corresponds to the metastable part of the hysteresis loop, the sample itself become a voltage source. This tendency is less pronounced for a ceramic with small N_s and is altogether absent for a ceramic with large N_s. This is because a significant portion of the potential in a ceramic falls across the nonreorganizing Schottky regions, where there is a high-field region (Fig. 6) because of the small ε (Fig. 7). The local value of ε is given by the expression $\varepsilon_0(\varepsilon-1)=(\alpha+3\beta P^2+5\gamma P^4)^{-1}$. Here α, β, and γ are the parameters of the thermodynamic potential of the ferroelectric, and ε_0 is the permittivity of free space. Negative values of ε correspond to metastable parts of the hysteresis loop.[7] The larger the value of N_s, the larger is that part of the Schottky region which corresponds to the metastable segments of the hysteresis loop, where ε is small. Since the linear dimensions $\lambda_{\rm I}$

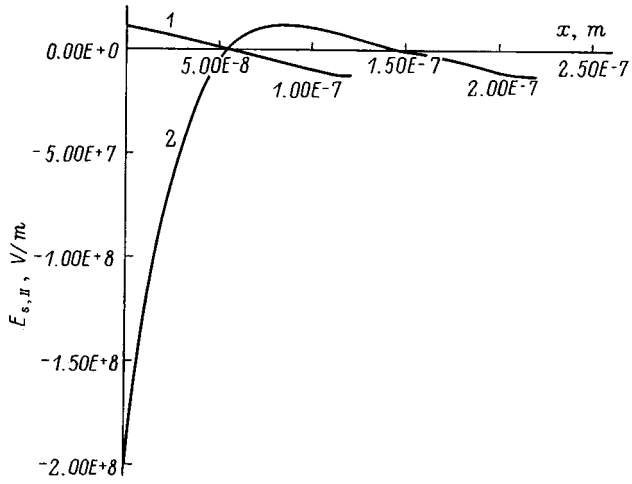

FIG. 6. Distribution of the field $E_{s,\rm II}$ in Schottky region II for an average field $E_{\rm av}=14\times10^6$ V/m and $N_s=3\times10^{18}$ (*1*) and 5×10^{18} m^{-2} (*2*).

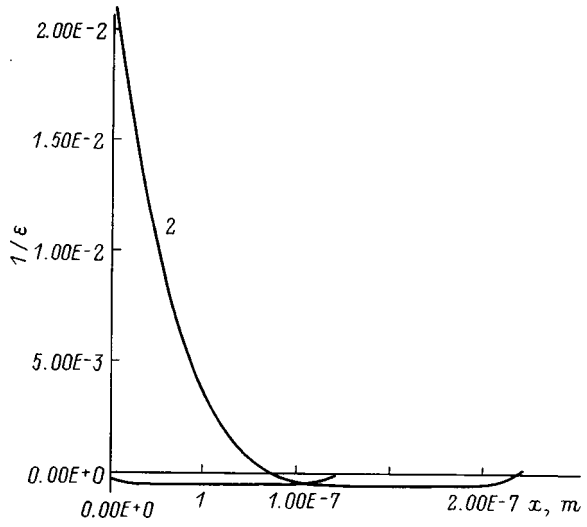

FIG. 7. Distribution of the inverse dielectric constant ε^{-1} in Schottky region II at an average field $E_{av}=14\times10^6$ V/m and $N_s=3\times10^{18}$ (*1*) and 5×10^{18} m^{-2} (*2*).

and λ_{II} of the Schottky regions I and II on the parts of the crystallite surface onto which the positive and negative bound charges due to P_s emerge are quite different ($l_I<l_{II}$) because of the difference in the dielectric constants of these regions,[8] the data in Figs. 6 and 7 are given for Schottky region II, which has the larger dimensions. Figure 5 shows that for a crystal or for a ceramic with small N_s the dependence of E_{av} on η_e is multivalued, and therefore even when the average field is lower than the coercive field ($E_{av}<E_c$) it is possible, if the polarization-switching field is applied for a long enough time, to have a transition to states for which the polarization switching becomes irreversible (the region of large η_e). For a ceramic with large N_s, irreversible polarization-switching processes will not occur even for high polarization-switching fields ($E_{av}>E_c$).

CONCLUSION

During the polarization switching of ferroelectric materials, which have large dielectric constants, electric fields are produced which can cause electron emission from the surface of the ferroelectric. In polycrystalline ferroelectrics there are regions of low dielectric constant in the regions of the crystallites near the grain boundaries; these affect the formation of the internal fields and promote restoration of the initial state after the polarization-switching voltage is turned off. For this reason, at high densities of surface states one can achieve higher degrees of emission without leaving the region of reversible polarization-switching phenomena than is possible in a crystalline ferroelectric or a polycrystalline ferroelectric with a low density of surface states.

[1] K. Biedrzycki and R. Le Bihan, Ferroelectrics **126**, 253 (1992).
[2] I. I. Ivanchik, Ferroelectrics **111**, 147 (1990).
[3] H. W. Gundel, *Electron Emission from Ferroelectrics: a New Generation of Pulsed Electron Beam Sources*, Shaker, Aachen (1996), 120 pp.
[4] G. I. Rosenman, Ferroelectrics **133**, 235 (1992).
[5] D. Shur and G. Rosenman, J. Appl. Phys. **80**, 3446 (1996).
[6] I. P. Raevskii, E. I. Bondarenko, A. N. Pavlov, O. I. Prokopalo, Ferroelectrics **76**, 55 (1987).
[7] A. N. Pavlov, Fiz. Tverd. Tela (St. Petersburg) **36**, 579 (1994) [Phys. Solid State **36**, 319 (1994)].
[8] A. N. Pavlov and I. P. Raevskiĭ, Zh. Tekh. Fiz. **67**(12), 21 (1997) [Tech. Phys. **42**, 1390 (1997)].
[9] W. Heywang, J. Am. Ceram. Soc. **47**, 484 (1964).
[10] G. A. Smolenskiĭ and I. A. Bokov *et al.*, Ferroelectrics and Antiferroelectrics [in Russian], Nauka, Leningrad (1971), 476 pp.
[11] H. Gundel, J. Handerek, and H. Riege, J. Appl. Phys. **69**, 975 (1991).
[12] D. Y. Wang and K. Umeya, J. Am. Ceram. Soc. **73**, 1574 (1990).

Translated by Steve Torstveit

Inclusion of a nonzero volume fraction of the new phase in the kinetics of crystallization of melts

A. V. Koropov

Sumy State Agricultural University, 244030 Sumy, Ukraine

S. A. Kukushkin and D. A. Grigor'ev

Institute of Problems of Machine Building, 199178 St. Petersburg, Russia

(Submitted March 27, 1998)
Zh. Tekh. Fiz. **69**, 53–58 (July 1999)

The growth rate of nuclei is calculated in self-consistent mean field approximation with allowance for the screening of a selected nucleus from heat fluxes, and an expression is obtained for the screening length of the heat flux. It is shown that the growth rate of a nucleus depends on the degree of crystallization of the melt. The influence on the crystallization process of collisions of the particles as they coalesce into a polycrystalline solid is investigated. A refinement of the size distribution function in the thermal Ostwald ripening stage is obtained.
© *1999 American Institute of Physics.* [S1063-7842(99)01107-1]

At present there is much interest in the theoretical description of the crystallization of melts.[1–6] The construction of a complete theory describing all stages of this process (nucleation, independent growth of the nuclei, Ostwald ripening) is of enormous significance both for the physics of phase transitions and for practical metallurgy. The stage of nucleation of the new phase in melts was investigated in Refs. 2 and 3, and a theory of the Ostwald ripening of the ensemble of nuclei in a crystallizing multicomponent melt was constructed in Refs. 4 and 5. However, all of the proposed theories of Ostwald ripening[4–6] are based on the approximation of zero volume fraction, i.e., in the derivation of all the basic relations it was assumed that the volume occupied by the particles of the new phase is strictly zero. This simplification affects such parameters of the nucleation theory as the screening length of the heat fluxes and the growth rate of an island of given radius.[7] It must also be taken into account that in studying the thermal Ostwald ripening the heat sinking is taken into account only in the heat balance equation, i.e., the theory treats the evolution of the system as a whole, on the macroscopic level. In this paper we shall consider the influence of heat sinking on a microscopic level, i.e., at the level of the growth of an individual nucleus.

The influence of a nonzero volume fraction of the new phase in the diffusional growth of a nucleus and in the decomposition of supersaturated solid solutions was first investigated in Refs. 8–10. One should also mention Ref. 11, in which a method based on a classical diagram technique was used.

In this paper we shall consider the crystallization of melts with allowance for the finite distance between nuclei and derive expressions for the growth rate and screening length in the self-consistent effective-medium approximation with allowance for the screening of heat fluxes at a selected nucleus.

Another goal of this study is to investigate the influence of the nonzero volume fraction in the late stage of the first-order phase transition in melts — the stage of thermal Ostwald ripening.[4–6]

Let us turn to a derivation of expressions for the growth rate of a nucleus of radius R with allowance for the screening of the heat fluxes.

SELF-CONSISTENT EFFECTIVE-MEDIUM APPROXIMATION WITH ALLOWANCE FOR THE SCREENING OF THE HEAT FLUXES TO A SELECTED NUCLEUS

Let us select a nucleus of the new phase in a supercooled one-component melt. For the sake of definiteness, let us assume that it has spherical symmetry. Heat is removed from the melt with a sink strength $K<0$ that is constant in time. The temperature $T(\mathbf{r},t)$, where $\mathbf{r}$ is the coordinate of a point in space and t is the time, will be determined by the heat conduction equation in a nonmoving fluid:

$$\rho C_P \frac{\partial T}{\partial t} = -\nabla \cdot \mathbf{q} + K, \quad \mathbf{q} = -\chi \nabla T \tag{1}$$

with boundary conditions at the surface S_i of some ith nucleus of radius R:

$$(\mathbf{q}\cdot\mathbf{n})|_{S_i} = \beta(T - T_R)|_{S_i}, \tag{2}$$

where ρ is the density of the melt, χ is the thermal diffusivity, C_P is the specific heat, and β is the specific boundary flux of atoms to the nucleus.

To solve the problem of the many-particle dynamics, it is necessary to solve Eq. (1) with the boundary conditions (2) imposed at all the nuclei. This problem obviously cannot be solved exactly. To solve it we shall use the self-consistent mean field approximation.[9,10] The essence of this approach is to replace the spatially nonuniform thermal field $T(\mathbf{r},t)$ by a uniform field that has been averaged over all the islands of the ensemble and which is modulated by the field of the

 © 1999 American Institute of Physics

selected nucleus. The evolution of the temperature field in time is governed by the heat balance equation in the system, and that makes this approach self-consistent. In the present study we take into account that the volume fraction of the new phase is nonzero, and so the distance between islands is finite. It is clear that the thermal field in the region close to the selected nucleus is formed mainly by that nucleus, and the influence of the entire ensemble begins to play a substantial role only at a certain distance from the nucleus. Thus for a selected nucleus one can define a zone of influence in which there are no other nuclei, and approximate this zone by a sphere of radius $R_0(R)$. Physically this means that all the heat removed by the external sink K is released by the given nucleus. Thus the entire space occupied by the crystallizing melt will be divided into two zones: the sphere of influence $R<r<R_0$, and the outer region $r \geqslant R_0$, which we shall first describe separately and then match the solutions obtained. It can be shown that inside the sphere of influence, $R<r<R_0$, Eq. (1) holds in the quasisteady approximation:

$$-\nabla \cdot \mathbf{q}+K=0. \tag{3}$$

In the outer region $r \geqslant R_0$ we should go over from the set of localized heat sources at each nucleus to a continuous medium with continuously distributed sinks. For this we generalize the approach developed in Ref. 10 for the case of island growth from an adsorbed vapor on the surface of a substrate. We shall perform an averaging of equation (1) in a ''physically'' infinitesimal volume ΔV containing a large number of nuclei:

$$\langle \cdots \rangle = \lim_{\Delta V \to 0} \frac{1}{\Delta V} \overline{\int_{\Delta V'} \cdots dV'}. \tag{4}$$

After performing operation (4) for Eq. (1), we write the latter in the form

$$\rho C_P \frac{\partial \langle T \rangle}{\partial t} = -(\nabla \cdot \langle \mathbf{q} \rangle)_{\text{macro}} - \tilde{I} + K(1-\eta), \tag{5}$$

where

$$(\nabla \cdot \langle \mathbf{q} \rangle)_{\text{macro}} = \frac{1}{\Delta V} \oint_S \mathbf{q} \cdot d\mathbf{S};$$

S is a surface enclosing the individual nuclei in the volume ΔV,

$$\tilde{I} = \frac{1}{\Delta V} \sum_i \oint_{S_i} \mathbf{q} \cdot d\mathbf{S}$$

is the specific flux of heat released by the entire ensemble of nuclei at the physical point $\mathbf{r}$, and η is the volume fraction constituted by the nuclei.

We shall show below that the temperature changes substantially within the sphere of influence of a given nucleus, while outside it $\langle T \rangle$ differs from the mean temperature $\overline{T}$ of the melt only over the screening length for the screening of the heat flux by the surrounding centers, and we can replace the $\partial \langle T \rangle / \partial t$ in Eq. (5) by $d\overline{T}/dt$.

Combining Eqs. (3) and (5), we obtain the desired averaged heat conduction equation for the selected nucleus in a self-consistent mean field $\overline{T}$ with allowance for the screening of the heat fluxes:

$$[-\nabla \cdot \mathbf{q}+K]\Theta(R_0-r) + \left[-(\nabla \cdot \langle \mathbf{q} \rangle)_{\text{macro}} - \tilde{I} + K(1-\eta) - \rho C_P \frac{d\overline{T}}{dt}\right] \times \Theta(r-R_0)=0, \tag{6}$$

where $\Theta(x)$ is the Heaviside step function,

$$\Theta(x)=\begin{cases} 1, & x \geqslant 0, \\ 0, & x<0. \end{cases}$$

The boundary conditions for (6) are

$$\chi \left.\frac{dT}{dr}\right|_{r=R} = \beta(T(R)-T_R); \quad \langle T \rangle|_{r \to \infty} = \overline{T}(t). \tag{7}$$

The mean temperature $\overline{T}(t)$ is determined self-consistently from the heat balance equation. In addition, for the solution of equation (6) we must supplement (7) with the condition for the ''matching'' of the temperatures T and $\langle T \rangle$ and the heat fluxes $\mathbf{q}$ and $\langle \overline{q} \rangle$ at the boundary of the sphere of influence ($r=R$).

We use the solution of equation (6) with boundary conditions (7) in the following way. We write the temperature T of the melt and the heat flux density $\mathbf{q}$ in the form

$$T=T_0+T_1, \quad \mathbf{q}=\mathbf{q}_0+\mathbf{q}_1=-\chi(\nabla T_0+\nabla T_1). \tag{8}$$

Here, of course, for the average temperature of the melt, $\overline{T}=\overline{T_0}+\overline{T_1}$, the flux $\mathbf{q}_1$ and temperature T_1 are due solely to the presence of the sink K (for $K=0$ we have $T_1=0$ and $\mathbf{q}_1=0$) and are nonzero only inside the sphere of influence. They are determined from the natural condition at the boundary of the sphere of influence: $\mathbf{q}_1|_{r=R_0}=0$. Inside the sphere of influence we assume that $\langle T_1 \rangle = \overline{T_1}$.

The flux $\mathbf{q}_0$ in (8) is due to the exchange of heat between nuclei, and we will call it the exchange flux. For this flux the boundary of the sphere of influence does not have any significance. Making a change of various in Eq. (6) and boundary conditions (7), we get the heat conduction equation for T_1 (9) with the corresponding boundary conditions (10):

$$-\nabla \cdot \mathbf{q}_1 + K = 0, \tag{9}$$

$$\chi \left.\frac{dT_1}{dr}\right|_{r=R} = \beta T_1(R); \quad T_1|_{r=R_0}=T^*; \quad T_1|_{r>R_0}=\overline{T_1}. \tag{10}$$

The quantity T^* is determined from the self-consistency condition

$$\frac{4\pi}{3} \int_0^\infty R_0^3(R) f(R,t) dR = 1 \tag{11}$$

and the requirement

$$\chi \frac{dT_1}{dr}\bigg|_{r=R_0} = 0.$$

Here $f(R,t)$ is the size distribution of the nuclei. Assuming that in the effective-medium region $(\nabla \cdot \langle \mathbf{q}_1 \rangle)_{\text{macro}} = 0$, we obtain Eq. (12) with the boundary conditions (13) for calculating the value of T_0:

$$-\nabla \cdot \mathbf{q}_0 \cdot \Theta(R_0 - r) + \left[-(\nabla \cdot \langle \mathbf{q}_0 \rangle)_{\text{macro}} - \tilde{I} + K(1-\eta) - \rho C_p \frac{d\overline{T}}{dt} \right] \times \Theta(r - R_0) = 0, \tag{12}$$

$$\chi \frac{dT_0}{dr}\bigg|_{r=R} = \beta(T_0(R) - T_R); \quad T_0|_{r\to\infty} = \overline{T}_0. \tag{13}$$

Thus we have reduced Eq. (6) with boundary conditions (7) to the system of equations (9), (12) with boundary and auxiliary conditions (10), (11), and (13).

First we must find the flux q_1 from Eq. (9). Since its solution does not present any difficulties, we shall immediately write the value of the heat flux q_1 from a growing nucleus of radius R:

$$-q_{1R} = \chi \frac{dT_1}{dr}\bigg|_{r=R} = -\frac{KR}{3} + \frac{\chi R_0^{\frac{1}{6}}}{R} \frac{\left(\frac{K}{\chi}\right)(R_0^2 - R^2) + T^* + \frac{KR}{3\beta}}{(R_0 - R) + \frac{\chi R_0}{\beta R}}, \tag{14}$$

where

$$T^* = \frac{1}{3}\left(\frac{K}{\chi}\right)\frac{R_0^3}{R}\left(1 - \frac{R}{R_0}\right)\left(1 + \frac{1}{2}\frac{R}{R_0}\right) + \frac{K}{3\beta R^2}(R_0^3 - R^3). \tag{15}$$

Then, substituting (15) into (14), we obtain

$$-q_{1R} = \frac{K}{3R^2}(R_0^3 - R^3). \tag{16}$$

Let us find the radius of the sphere of influence of the selected nucleus $R_0(R)$ from expressions (11) and (15). We introduce the notation $x \equiv R_0/R$, $\alpha = 3\chi T^*/KR^2$, $\gamma \equiv \chi/\beta R$, $\gamma_0 = \chi/R\overline{R}$. Then system (11) and (15) becomes

$$\frac{4\pi}{3}\int_0^\infty x^3 \cdot f(R,t) \cdot R^3 dR = 1, \tag{17}$$

$$(x-1)^2\left(x + \frac{1}{2}\right) + \gamma(x^3 - 1) = \alpha. \tag{18}$$

The solution of system (17), (18) does not present any particular mathematical difficulties, and we therefore turn immediately to the results. Let $x \gg 1$ $(R \ll R_0)$, which corresponds to a rather sharp distribution function $f(R,t)$ and a low degree of crystallization $(\eta^{1/3} \ll 1)$. Then $R_0(R)$ will assume the form

$$R_0(R) = \frac{R}{2(1+\gamma_0)}\left[\lambda + \left(\frac{\overline{R}}{R}\right)^{2/3} \cdot \lambda^{1/3}\frac{2(1+\gamma_0) - \eta^{1/3}}{\eta^{1/3}}\right], \tag{19}$$

where $\lambda \equiv (1+\gamma_0)/(1+\gamma)$.

Let us examine two limiting cases.

1. If $\beta \to 0$, then the growth or melting of the nuclei is limited by the boundary kinetics. In this case $R_0(R)$ can be written in the form

$$R_0(R) = \frac{R^{2/3}\overline{R}^{1/3}}{\eta^{1/3}} \sim R^{2/3}. \tag{20}$$

If $\beta \to \infty$ (the growth or melting of the nuclei is limited by the heat removal, and the boundary kinetics does not play a role), we will have

$$R_0(R) = \frac{R}{2}\left[1 + \left(\frac{\overline{R}}{R}\right)^{2/3}\frac{2 - \eta^{1/3}}{\eta^{1/3}}\right] \simeq \frac{R^{1/3}\overline{R}^{2/3}}{\eta^{1/3}} \sim R^{1/3}. \tag{21}$$

Thus from the solution of system (9), (10) we have obtained an expression for the flux (16) due to the sinking of heat from the system and expressions (19)–(21) for the radius of the sphere of influence. Now we must find the exchange flux q_0. We obtain a solution of equation (12) with the boundary conditions (13). In the effective-medium region $(r \geqslant R_0)$, for a random distribution of the nuclei with respect to their size and location, we can write the following expression for the averaged heat flux $\tilde{I}(r)$:

$$\tilde{I}(r) = \int_0^\infty 4\pi R^2[\langle q_{0R}(r) \rangle + q_{1R}] f(R,t) dR. \tag{22}$$

Here $\langle q_{0R} \rangle$ is the averaged exchange flux to a nucleus of radius R, and q_{1R} is the r-independent flux q_1 to the nucleus. The total heat flux I from all the nuclei in the system can be written as

$$I = \frac{1}{V}\sum_i \oint_{S_i} \mathbf{q} \cdot d\mathbf{S}_i = -\int_0^\infty 4\pi R^2(\overline{q_{0R}} + q_{1R}) f(R,t) dR. \tag{23}$$

Writing $\langle q_{0R} \rangle$ in the form

$$-\langle q_{0R} \rangle = \chi\varphi(R)(\overline{T}_0 - T_R) + \chi\varphi(R)(\langle T_0 \rangle - \overline{T}_0), \tag{24}$$

where $\varphi(R)$ is a function to be determined, and combining (22)–(24), we obtain an expression for $\tilde{I}$:

$$\tilde{I} = I + \frac{\chi}{l^2}[\langle T_0 \rangle - \overline{T}_0],$$

where $l^{-2} = 4\pi\int_0^\infty R^2\varphi(R) f(R,t) dR$.

Then Eq. (12) can be reduced to

$$-\nabla \cdot \mathbf{q}_0 - \frac{\chi}{l^2}(T_0 - \overline{T}_0)\Theta(r - R_0) + \left[K(1-\eta) - I - \rho C_P \frac{d\overline{T}}{dt}\right]\Theta(r - R_0) = 0. \tag{25}$$

The expression in the square brackets in Eq. (25) is independent of **r** and must go to zero in order for a physical (bounded as $\mathbf{r}\to\infty$) solution to exist; this corresponds to the heat balance equation

$$K(1-\eta)=C_P\frac{d\overline{T}}{dt}\rho+I. \tag{26}$$

Thus Eq. (12) goes over to the following equation:

$$-\nabla\cdot\mathbf{q}_0-\frac{\chi}{l^2}(T_0-\overline{T}_0)\Theta(r-R_0)=0, \tag{27}$$

$$\chi\frac{dT_0}{dr}\bigg|_{r=R}=\beta[T_0(R)-T_R],\quad T_0|_{r\to\infty}=\overline{T}_0. \tag{28}$$

We note that the parameter l in (27), which has the dimensions of length and depends on the distribution function, has the physical meaning of the screening length for the heat flux to the given nucleus. An expression for this index will be obtained below.

Since the solution of equation (27) with boundary conditions (28) does not present any fundamental difficulties, we will immediately write out the expression for the heat flux $\overline{q}_{0R}$:

$$-\overline{q}_{0R}=\frac{\chi}{R\left(1-\dfrac{R}{R_0+l}\right)+\dfrac{\chi}{\beta}}(\overline{T}_0-T_R), \tag{29}$$

where the screening length l for the heat flux is determined from the equation

$$l^2=\frac{1}{4\pi N\overline{R}}\left[1+\gamma_0-\frac{\overline{R}}{R_0(\overline{R})+l}\right].$$

Let us now give a formula for the growth rate of the nucleus with allowance for the screening of the thermal fields,

$$\frac{dR}{dt}=\frac{1}{\rho L}(\overline{q}_{0R}+q_{1R}),$$

and, substituting into this equation the expressions for $\overline{q}_{0R}$ and q_{1R} and considering, for example, the case $\beta\to 0$, i.e., when the growth rate of the particle is limited by the rate of incorporation of the atoms, we obtain an expression for dR/dt in the form

$$\frac{dR}{dt}=\frac{\beta\alpha}{RL\rho_s}\left[\frac{R}{R_{\rm cr}}-\frac{K}{3}\left(\frac{1}{R^{1/3}}\sqrt{\frac{(1-\eta)^{1/3}}{\eta}}\times\left(\frac{1}{3/4\pi N}\right)^{1/9}-R^2\right)\right]. \tag{30}$$

Comparing expression (30) with the formula for the growth rate of a nucleus obtained[5] without taking the screening of the thermal fields into account, we see that as the degree of crystallization η increases, the growth rate becomes comparatively smaller. We note that the results obtained above were obtained in the approximation of a rather low degree of crystallization. Let us now turn to an analysis of the late stage of the first-order phase transition, when the particles grow together into a continuous polycrystalline solid and the degree of crystallization approaches unity.

ALLOWING FOR THE ''COLLISIONS'' OF NEW-PHASE NUCLEI IN THE OSTWALD RIPENING STAGE

In Refs. 5 and 6 the theory of Ostwald ripening during crystallization of a melt was treated in an approximation in which the volume fraction of the new-phase nuclei in the melt is assumed to be zero. As the volume of the new phase increases, the influence of collisions[1)] of nuclei on their size distribution should become more noticeable. For example, the collision of grains of the new phase occurs during crystallization of melts in the stage of formation of a continuous polycrystalline solid.[6]

Let us write a system of equations describing the heat of Ostwald ripening of an ensemble of nuclei in a one-component melt with allowance for the nonzero volume fraction of the new phase. This system consists of the continuity equation (31), which includes a collision integral $I_{\rm col}$, the heat balance equation (32), and an expression in general form (33) for the growth rate of the nuclei:

$$\frac{\partial f(R,t)}{\partial t}+\frac{\partial}{\partial R}[f(R,t)V_R]=I_{\rm col}, \tag{31}$$

$$Q_0=Q(t)+\frac{4}{3}\pi L\rho_s\int_0^\infty f(R,t)R^3dR, \tag{32}$$

$$\frac{dR}{dt}=\frac{\rm const}{R^{p-1}}\left(\frac{R}{R_{\rm cr}}-1\right). \tag{33}$$

Here L is the latent heat of fusion, ρ_s is the density of the new-phase fraction of the particles, $f(R,t)$ is the size distribution function of the islands, $I_{\rm col}$ is the collision integral, $Q(t)$ is the quantity of heat in the melt at a given time, const is the constant in the equation for the growth rate of a nucleus,[5] and Q_0 is the quantity of heat in the melt at the start of the Ostwald ripening.

To take the collisions of the nuclei into account, we use the method developed by Lifshits and Slezov in Ref. 11, which treated the decomposition of supersaturated solid solutions. In fact, it has been shown[5] that in dimensionless form, the equations describing the thermal and diffusional Ostwald ripening of an ensemble of new-phase nuclei have the same mathematical structure. We use this formalism and introduce the relative variables $u=R/R_{\rm cr}(t)$ and ''time'' $\tau=\ln x^2$, where $x=\Delta T/\Delta T_0$, where ΔT and ΔT_0 are the instantaneous and initial values of the supercooling of the melt.

Choosing the origin of coordinates at some nucleus, we note that all the other nuclei will move toward the chosen center (since the coordinates of their centers r_i do not change, and $R_{\rm cr}(x)\to\infty$). Here the majority of the nuclei will melt without having reached the chosen center, and collisions will not occur. We write the total system of equations describing the thermal Ostwald ripening with allowance for collisions of the grains in the relative variables u and τ as

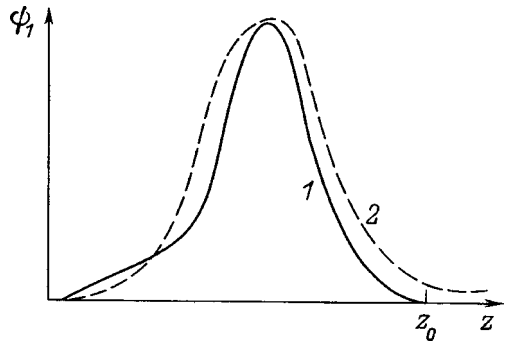

FIG. 1. Form of the size distribution function of the particles: *1* — for zero volume fraction of the new phase;[5] *2* — with a nonzero volume fraction of the new phase.

$$\frac{\partial\varphi}{\partial\tau}+\frac{\partial}{\partial u}\left(\varphi(u,\tau)\frac{du}{d\tau}\right)=I_{\rm col},$$
$$\varphi|_{\tau=0}=\varphi_0, \tag{34}$$

$$1-\frac{\Delta_0 T}{Q_0}l^{-\tau/3}=l^{\tau}\int_0^{\infty}\varphi(u,\tau)u^3du, \tag{35}$$

$$\frac{du^3}{d\tau}=(u-1)\gamma-u^3. \tag{36}$$

Here Eq. (34) is the continuity equation in the space of variables u,τ; Eq. (35) is the heat balance equation in the new variables, and Eq. (36) is the growth rate of an island. Equation (35) describes the heat balance in the system, expression (36) is the dimensionless growth rate of the nucleus, and φ is the dimensionless distribution function. We note that the system of equations (34)–(36) is mathematically identical to the system described in Ref. 11 and can be solved by the same method. Here we shall write down immediately the expression for the dimensionless size distribution function of the particle with allowance for ''collisions'' in the first approximation, $\varphi_1(\tau,z)$, where $z=R^3/R_{\rm cr}^3$,

$$\varphi_1=\begin{cases}\dfrac{1}{e^{\tau}}\dfrac{1}{\tau_{\rm col}}e^{\Delta}A\dfrac{e^{-\psi}}{\beta} & z\leqslant z^0\quad A=\displaystyle\int_{z'}^{2z^0}e^{\psi^*}I_{\rm col}^0(z')dz',\\ \dfrac{1}{e^{\tau}}\dfrac{1}{\tau_{\rm col}}\dfrac{e^{-\psi^*}}{\beta}\displaystyle\int_z^{2z^0}e^{\psi}I_{\rm col}(z')dz' & z^0\leqslant z<2z^0,\\ 0 & z\geqslant 2z^0.\end{cases} \tag{37}$$

Here

$$\psi=\begin{cases}\psi^*+\Delta\quad z>z^0 & \psi^*=\dfrac{4}{3}\ln(\sqrt{z}+3)+\dfrac{5}{3}\ln|\sqrt{z}-\sqrt{z^0}|\\ & \quad+\dfrac{1}{1-\sqrt{\dfrac{z}{z^0}}}-\ln\dfrac{3^3e}{2^{5/3}},\\ \psi^*\quad z<z^0 & \Delta=\dfrac{3\pi}{2\sqrt{\dfrac{3}{4}\dfrac{\Delta\gamma}{\gamma_0}}}\ \Delta\gamma=\gamma_0-\gamma>0.\end{cases}$$

The collision integral has the form

$$I_{\rm col}=\frac{1}{\chi}\frac{1}{\tau_{\rm col}}\frac{3\mu}{4\pi}\left\{\frac{1}{2}\int_0^{\infty}\frac{1}{e^{2\tau}}\omega(z-z',z')\varphi(z-z',\tau)\varphi(z',\tau)dz'-\frac{1}{e^{\tau}}\varphi(z,\tau)\int_0^{\infty}\frac{1}{e^{\tau}}\omega(z,z')\right\}.$$

Here $\omega(z,z')$ is the effective relative volume in which nuclei z interact with nuclei z'. The relative collision time is of the order of unity, $t_{\rm col}\simeq 1$, since all the parameters in the expression describing the melting of a smaller grain in a collision, $du^3/d\tau\simeq-\gamma-u^3$, and in the initial conditions $u|_{\tau=0}=u_0\simeq 1$.

Figure 1 shows the form of this function. We see that the distribution function (37) has a ''tail'' corresponding to the formation of particles of large sizes.

This study was supported in part by the Russian Fund for Fundamental Research (Grants Nos. 96-03-32396 and 98-03-32791) and the Russian Federal Center ''Integration'' (Project No. 589).

[1)] By collision we mean the direct coalescence or melting of smaller-radius nuclei upon entering the sphere of influence (see above) of a larger particle. The heat absorbed in the melting promotes growth of the larger particle. Here the possible slight amount of absorption by the other particles of the heat released on coalescence will not be taken into account.

[1] V. P. Skripov and V. P. Koverda, *Spontaneous Crystallization of Supercooled Liquids* [in Russian], Nauka, Moscow (1984), 231 pp.
[2] V. V. Slezov and S. F. Kukushkin, Fiz. Tverd. Tela (St. Petersburg) **38**, 433 (1996) [Phys. Solid State **38**, 239 (1996)].
[3] A. V. Osipov, Fiz. Tverd. Tela (St. Petersburg) **36**, 1213 (1996) [*sic*].
[4] S. A. Kukushkin and V. V. Slyosov, J. Phys. Chem. Solids **57**, 195 (1996).
[5] S. A. Kukushkin, Fiz. Tverd. Tela (Leningrad) **27**, 2987 (1985) [Sov. Phys. Solid State **27**, 1794 (1985)].

[6] S. A. Kukushkin and D. A. Grigor'ev, Zh. Tekh. Fiz. **65**(10), 154 (1995) [Tech. Phys. **40**, 698 (1995)].
[7] V. N. Detsik, S. A. Kukushkin, A. V. Osipov *et al.*, Fiz. Tverd. Tela (St. Petersburg) **39**, 118 (1997) [Phys. Solid State **39**, 104 (1997)].
[8] V. V. Slezov, Fiz. Tverd. Tela (Leningrad) **31**, 20 (1989) [Sov. Phys. Solid State **31**(8), 1289 (1989)].
[9] A. V. Koropov, P. N. Ostapchuk, and V. V. Slezov, Fiz. Tverd. Tela (Leningrad) **33**, 2835 (1991) [Sov. Phys. Solid State **33**, 1602 (1991)].
[10] I. M. Lifshits and V. V. Slezov, Fiz. Tverd. Tela (Leningrad) **1**, 1401 (1958).
[11] M. Marder, Phys. Rev. A **36**, 858 (1987).

Translated by Steve Torstveit

Theory of the generation of mechanical vibrations by laser radiation in solids containing internal stresses on the basis of the thermoelastic effect

K. L. Muratikov

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, St. Petersburg, 194021 St. Petersburg, Russia
(Submitted April 15, 1998)
Zh. Tekh. Fiz. **69**, 59–63 (July 1999)

The behavior of the nonsteady deformations in solids containing internal stresses under irradiation by temporally modulated laser radiation is analyzed. In the framework of the nonlinear theory of thermoelasticity a model is proposed for the excitation of mechanical vibrations with allowance for the dependence of the thermoelastic coupling parameter on the initial deformation. For the case of a piezoelectric method of detecting the mechanical vibrations in a uniformly deformed sample, an analytical expression is obtained for the electrical signal taken off from the piezoelement. The behavior of the piezoelectric signal under various conditions is investigated, and the results are compared with the available experimental data and found to be in qualitative agreement. © *1999 American Institute of Physics.* [S1063-7842(99)01207-6]

The detection of mechanical stresses in solids is an important research area in the development of modern mechanics, nondestructive monitoring, and diagnostic tools. Various methods are used for this purpose. Among the most important of these are ultrasonics,[1] Raman spectroscopy,[2,3] x-ray[4] and neutron[5,6] diffraction, magnetics,[7] and also methods based on the use of holographic interferometry.[8–10] In addition, recently there has been serious attention devoted to investigating the possibility of using the thermoelastic effect for diagnostics of the mechanical stresses in solids.[11–16] In that approach the mechanical stresses are ordinarily detected from the thermoelastic strains generated in the object by temporally modulated laser radiation. An important advantage of the thermoelastic method is that it can be applied to objects of various natures, since the thermoelastic effect is quite universal. The feasibility of this approach has already been confirmed by a number of experimental studies for metals[11,14,15] and ceramics.[12,13,16] Nevertheless, the mechanism by which the mechanical stresses influence the results of the laser thermoelastic measurements remains insufficiently clear. In Ref. 14 a model of the formation of the thermoelastic signal was proposed which attributed its dependence on the mechanical stresses mainly to the stress dependence of the thermophysical parameters of the material. On the other hand, in Ref. 16 it was shown experimentally that a dependence of the thermoelastic signal on the residual stresses can be observed in ceramics even in the absence of any noticeable changes in their thermophysical properties.

To explain the aforementioned features, in this paper I propose a different model for the formation of the thermoelastic signal generated in solid objects by laser radiation. The important distinction of this model is that it takes into account the dependence of the thermoelastic coupling coefficient on the mechanical stresses. Such a dependence has been noted previously both for the coefficient of thermal expansion[17] and for the elastic modulus.[18] Since the thermoelastic coupling coefficient in the case of an isotropic solid is a product of these two quantities, it is important to include such a dependence in the thermoelastic coupling coefficient as well.

Let us discuss the generation of mechanical vibrations by laser radiation in a solid containing mechanical stresses in the framework of nonlinear mechanics with initial strains.[19] Here we shall assume that the initial strains are not small. Therefore, the displacement vector of points of the solid will be assumed to be specified in the form

$$\mathbf{u}(\mathbf{r})=\mathbf{r}+\mathbf{U}(\mathbf{r})+\Delta\mathbf{u}(\mathbf{r}), \tag{1}$$

where $\mathbf{U}(\mathbf{r})$ describes the initial strain, and the vector $\Delta\mathbf{u}(\mathbf{r})$ describes the displacement of particles of the solid due to thermoelastic strains under the influence of the laser radiation.

The equation of motion of the elements of the solid in nonlinear mechanics[20] can be written in the form

$$\frac{\partial P_{ik}}{\partial x_k}=\rho_0\Delta\ddot{u}_i, \tag{2}$$

where $P_{ik}=(\partial u_i/\partial x_m)t_{km}$ is the Piola–Kirchhoff tensor, t_{km} is the stress tensor, which is related to the internal energy density W of the solid by the relation $t_{km}=\partial W/\partial u_{km}$, with the strain tensor

$$u_{km}=\frac{1}{2}\left(\frac{\partial u_k}{\partial x_m}+\frac{\partial u_m}{\partial x_k}+\frac{\partial u_l}{\partial x_k}\frac{\partial u_l}{\partial x_m}\right),$$

and ρ_0 is the density of the solid in the initial state.

The energy density of a strained solid can be written in the form a sum, $W=W_E+W_T$ (W_E is the mechanical energy density, and W_T is the energy density due to the thermoelastic strains). In this paper it is assumed that the solid is isotropic in the initial state, and the mechanical part of its energy is determined in the Murnaghan model,[21] according to which the mechanical energy density can be written

 © 1999 American Institute of Physics

$$W_E=(\lambda+2\mu)\frac{I_1^2}{2}-2\mu I_2+(l+2m)\frac{I_1^3}{3}-2mI_1I_2+nI_3, \tag{3}$$

where λ and μ are the Lamé coefficients, l,m,n are the Murnaghan constants, $I_1=u_{kk}$,

$$I_2=\frac{1}{2}[(u_{kk})^2-u_{lm}u_{lm}],$$

and

$$I_3=\frac{1}{3}\left[u_{ik}u_{il}u_{kl}+\frac{3}{2}u_{ik}u_{ik}u_{ll}+\frac{1}{2}(u_{ll})^3\right].$$

The stress tensor $t_{ij}^{(E)}$ associated with the mechanical energy can be determined from the relation

$$t_{ij}^{(E)}=\frac{\partial W_E}{\partial I_1}\frac{\partial I_1}{\partial U_{ij}}+\frac{\partial W_E}{\partial I_2}\frac{\partial I_2}{\partial U_{ij}}+\frac{\partial W_E}{\partial I_3}\frac{\partial I_3}{\partial U_{ij}}. \tag{4}$$

Taking expression (3) into account, we write the stress tensor $t_{ij}^{(E)}$ in explicit form as

$$t_{ij}^{(E)}=\left[\lambda I_1+\left(l+\frac{n}{2}\right)I_1^2-2mI_2\right]\times\delta_{ij}+[2(\mu+mI_1)-nI_1]U_{ij}+nU_{pi}U_{pj}. \tag{5}$$

The thermoelastic energy density is due solely to the strains caused by the action of the laser radiation on the solid. In determining it we shall assume, in analogy with Ref. 17, that the thermoelastic coupling coefficient depends linearly on the strain tensor. Then the thermoelastic energy density of the solid can be written as

$$W_T=-\gamma_{ik}(u_{ik}-U_{ik})\Delta T, \tag{6}$$

where $\gamma_{ik}=\gamma_0(\delta_{ik}+\beta U_{ik})$, γ_0 is the thermoelastic coupling coefficient for the unstrained solid, β is a coefficient that determines the dependence of the thermoelastic coupling on the initial strain, U_{ik} is the initial strain of the solid, and $\Delta T=T-T_0$, with T_0 the temperature of the surrounding medium.

In what follows we shall assume that the strains Δu_{ik} arising as a result of the action of laser radiation on the solid are small. Then the thermoelastic energy (6) of the solid will be determined up to terms linear in Δu_{ik} by the equation

$$W_T=-\gamma_0(\delta_{ik}+\beta U_{ik})(T-T_0)\Delta u_{ik}. \tag{7}$$

For $\beta=0$ Eq. (7) reduces to the usual expression for the thermoelastic energy density of an isotropic solid. We note that the value of the coefficient β in accordance with Ref. 17 is given approximately by $\beta\cong KE\gamma_g/3$ (K is the compressibility, E is the elastic modulus, and γ_g is the Grüneisen coefficient), and for metals at small strains it typically has a value between 1 and 2.

The stress tensor $t_{ij}^{(T)}$, corresponding to the thermoelastic energy (6), can be found from the expression

$$t_{ij}^{(T)}=\frac{\partial W_T}{\partial U_{ij}}=-\gamma_0(\delta_{ij}+\beta U_{ij})(T-T_0). \tag{8}$$

Using expressions (2), (5), and (7), we can obtain the equation of motion for the components of the displacement vector Δu_i in explicit form. Assuming the oscillations of the temperature in the sample are small and keeping only the term to the first power of ΔT, we obtain the equation of motion in the form

$$a_{iklp}\frac{\partial^2\Delta u_k}{\partial x_l\partial x_p}+b_{ikp}\frac{\partial\Delta u_k}{\partial x_p}=\gamma_0\left[\frac{\partial\Delta T}{\partial x_i}+\frac{\partial}{\partial x_k}\left(\frac{\partial U_i}{\partial x_k}\Delta T\right)\right]+\gamma_0\beta\frac{\partial}{\partial x_k}\left[\left(U_{ki}+\frac{\partial U_i}{\partial x_p}U_{kp}\right)\Delta T\right]+\rho_0\Delta\ddot{u}_i, \tag{9}$$

where

$$a_{iklp}=\delta_{ik}t_{lp}^{(0)}+N_{likp}+\frac{\partial U_i}{\partial x_m}N_{lmkp},$$

$$b_{ikp}=\delta_{ki}\frac{\partial t_{lp}^{(0)}}{\partial x_l}+\frac{\partial N_{likp}}{\partial x_l}+\frac{\partial^2U_i}{\partial x_l\partial x_m}N_{lmkp}+\frac{\partial U_i}{\partial x_m}\frac{\partial N_{lmkp}}{\partial x_l},$$

$$N_{ijkp}=2c_{ij}a_{kp}+4m\delta_{ij}a_{km}U_{mp}+b(a_{ki}\delta_{pj}+a_{kj}\delta_{pi})+n(U_{mj}a_{km}\delta_{pi}+a_{ki}U_{pj})+n(U_{mi}a_{km}\delta_{pj}+a_{kj}U_{pi}),$$

$$a_{ki}=\frac{1}{2}\left(\delta_{ik}+\frac{\partial U_k}{\partial x_i}\right),$$

$$c_{ij}=\left[\lambda+2\left(l-m+\frac{n}{2}\right)I_1\right]\delta_{ij}+(2m-n)U_{ij},$$

and $t_{ij}^{(0)}$ are the components of the initial strain tensor.

Equation (9) can be used to determine the nonsteady strain in the solid provided that one knows the distribution of the initial strain in it and the temperature distribution produced by the laser radiation. In addition, for solving the equation obtained one needs to specify the boundary conditions. In accordance with the nature of the problem to be solved, the initial strain of the solid should be regarded as given. As to the thermophysical properties of the solid, we shall assume that the appearance of the internal stresses in it does not lead to substantial changes. Such a situation has been observed experimentally in a number of ceramics, for example.[16]

The boundary conditions on Eq. (9) in the general case are determined by the means employed for detection of the variable strains in the object. In this paper we consider the case when these strains are detected by means of a piezoelement attached to the sample (see Fig. 1). For this case it is necessary to write the boundary conditions at the upper and lower surfaces of the solid. Under the influence of external stresses the surface of the sample is slightly deformed. Therefore, e.g., on the upper surface of the sample, the boundary condition can be written in the form

$$n_kP_{ik}|_{z=Z(x,y)}=0, \tag{10}$$

where $z=Z(x,y)$ is the equation of the upper surface of the object deformed by the internal stresses.

When applying boundary condition (10) to the problem under study one should keep in mind that we are interested only in the nonsteady component of the strains. Below, in

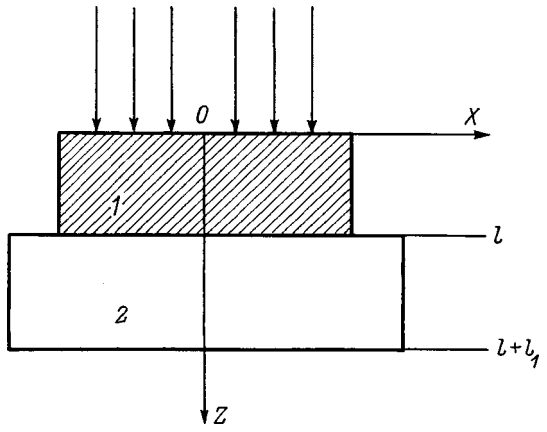

FIG. 1. Geometry of the sample and piezoelement: *1* — sample, *2* — piezoelement.

accordance with the usual assumptions of nonlinear mechanics, we shall replace the boundary condition at the deformed surface by a boundary condition at the undeformed surface.[22] Here we take into account the fact that the initial deformation of the surface occurs under the influence of static internal stresses. Then, after linearization with respect to Δu_i and ΔT, we obtain on the basis of the foregoing comments the following form for the boundary condition (10):

$$n_k\left[t_{km}^{(0)}\frac{\partial u_i}{\partial x_m}+\left(\delta_{ik}+\frac{\partial U_i}{\partial x_m}\right)(\Delta t_{km}^{(E)}+t_{km}^{(T)})\right]\Bigg|_{z=0}=0, \qquad (11)$$

where $\Delta t_{km}^{(E)}$ is the variable component of the mechanical stress generated in the object by the laser radiation.

At the boundary of the object with the piezoelement, $z=l$, we use the condition of continuity of the normal component of the stress vector. Using Eq. (9) and the stated boundary conditions, one can find the components of the displacement vector of the particles of the solid when mechanical vibrations are excited in it by laser radiation. In this paper we limit consideration to the case of a uniformly strained solid with vector components of the initial strain specified in the form $U_i=A^{(i)}x_i$ (the $A^{(i)}$ are constants characterizing the uniform strain along different directions). In this case Eq. (9) simplifies and can be put in the form

$$f_k^{(i)}\frac{\partial^2\Delta u_i}{\partial x_k\partial x_k}+h_k^{(i)}\frac{\partial^2\Delta u_k}{\partial x_k\partial x_i}=g^{(i)}\frac{\partial\Delta T}{\partial x_i}+\rho_0\Delta\ddot{u}_i, \qquad (12)$$

where

$$f_k^{(i)}=\left[t_{kk}^{(0)}+\frac{(1+A^{(i)})^2}{2}(b+nU_{ii}+nU_{kk})\right],$$

$$b=2\mu+(2m-n)U_{pp},$$

$$h_k^{(i)}=(1+A^{(i)})(1+A^{(k)})\times\left[c_{ii}+2mU_{kk}+\frac{1}{2}(b+nU_{ii}+nU_{kk})\right],$$

$$g^{(i)}=\gamma_0(1+A^{(i)})(1+\beta U_{ii}),$$

$$c_{ii}=K-\frac{2}{3}\mu+2\left(l-m+\frac{n}{2}\right)U_{pp}+(2m-n)U_{ii}.$$

We note that the expressions for $f_k^{(i)}$, $h_k^{(i)}$, c_{ii}, and $g^{(i)}$ on the right-hand side are not to be summed over repeated indices. Equation (12) can be used to determine the strain in the solid, provided that the temperature distribution produced in it by the exciting laser radiation is known. As we have said above, we are assuming that the appearance of the internal stresses in the solid will not appreciably alter its thermophysical parameters. In addition, we shall assume that the surface of the sample is illuminated uniformly by the laser radiation and that the laser radiation is modulated in time according to a harmonic law. Then if the exciting laser radiation is completely absorbed at the surface of the sample and is modulated in time by a harmonic law, the nonsteady component of the temperature inside the sample will be given by the expression

$$\Delta T(z,t)=\Delta T_s e^{-\sigma z+i\omega t}, \qquad (13)$$

where $\sigma(1+i)\sqrt{\omega/2\kappa}$, where κ is the thermal diffusivity of the sample, ΔT_s is the amplitude of the oscillations of the surface temperature of the sample, and ω is the angular frequency of modulation of the exciting radiation.

The requirement of continuity of the normal component at the boundary between the sample and piezoelement can be used to find the signal taken off from the piezoelement. For this one must use the familiar equations relating the mechanical and electrical characteristics of a piezoelement.[23] For the problem under consideration these equations can be written in the form

$$t_{33}^{(P)}=C^{(ET)}\frac{\partial u_3^{(p)}}{\partial z}-e^{(T)}E_3, \qquad (14a)$$

$$D_3=e^{(T)}\frac{\partial u_3^{(p)}}{\partial z}+\varepsilon^{(ST)}E_3, \qquad (14b)$$

where $t_{33}^{(P)}$ is the mechanical stress inside the piezoelement, $u_3^{(p)}$ is the displacement vector of the points of the piezoelement, D_3 is the electric displacement, E_3 is the electric field, and $C^{(ET)}$, $\varepsilon^{(ST)}$, and $e^{(T)}$ are characteristics of the piezoelectric which are defined as in Ref. 23.

For the quasistatic case, when the acoustic wavelength is much greater than the characteristic dimensions of the piezoelement, its deformation along the thickness can be assumed uniform. Then with the use of Eq. (14b), we obtain the following expression for the voltage signal $V(t)$ detected at the output of an open-circuited piezoelement:

$$V(t)=-\frac{e^{(T)}}{\varepsilon^{(ST)}}l_1\frac{\partial u_3^{(p)}(t)}{\partial z}\Bigg|_{z=l}, \qquad (15)$$

where l_1 is the thickness of the piezoelement.

In expression (15) one can go over from the strains in the piezoelement to the strains in the sample by making use of the continuity equation for the normal component of the stress at the sample–piezoelement boundary,

$$t_{33}^{(p)}=P_{33}|_{z=l}. \qquad (16)$$

In accordance with the character of the quantities appearing in expression (16), for finding the variable components of the signal from the piezoelement it is sufficient to know the $\Delta u_3(z,t)$ component of the displacement vector. For harmonic modulation of the exciting laser radiation, we write this component in the form $\Delta u(z,t)=\Delta u(z,\omega)e^{i\omega t}$, and using Eq. (12) and the indicated boundary conditions, we obtain for $\Delta u_3(z,\omega)$ the following result:

$$\Delta u_3(z,\omega)=-\frac{U_3^{(0)}e^{-\sigma l}}{\cos Ql}\cos Qz + \left[\frac{\gamma_0(1+A^{(3)})(1+\beta U_{33})}{f_3^{(3)}+h_3^{(3)}}\Delta T_s+\sigma U_3^{(0)}\right] \times \sin Q(z-l)+U_3^{(0)}e^{-\sigma z}, \tag{17}$$

where

$$U_3^{(0)}=-\frac{\gamma_0\sigma(1+A^{(3)})(1+\beta U_{33})}{(f_3^{(3)}+h_3^{(3)})\sigma^2+\rho_0\omega^2}\Delta T_s,$$

$$Q=\sqrt{\frac{\rho_0\omega^2}{f_3^{(3)}+h_3^{(3)}}}.$$

The mechanical and electrical characteristics of the piezoelement are related by Eqs. (14). Then with the use of the continuity equation for the normal component of the stress at the sample–piezoelement boundary, we obtain the following expression for the voltage signal $V(\omega)$ registered at the output of an open-circuited piezoelement:

$$V(\omega)=-C(f_3^{(3)}+h_3^{(3)})\frac{\partial\Delta u_3}{\partial z}\Big|_{z=l}, \tag{18}$$

where we have introduced the notation

$$C=\frac{e^{(T)}l_1}{C^{(ET)}\varepsilon^{(ST)}+e^{(T)^2}}.$$

Using relations (17) and (18), one can find the piezoelectric signal in explicit form:

$$V(\omega)=-C(f_3^{(3)}+h_3^{(3)})\left\{QU_3^{(0)}e^{-\sigma l}\tan Ql + Q\left[\frac{\gamma_0(1+A^{(3)})(1+\beta U_{33})}{f_3^{(3)}+h_3^{(3)}}\Delta T_s+\sigma U_3^{(0)}\right] - \sigma U_3^{(0)}e^{-\sigma l}\right\}. \tag{19}$$

Expression (19) can be used to determine the piezoelectric signal under rather general conditions. In this paper we shall assume that the sample is quite thick in respect to the propagation of thermal waves, i.e., we can set $e^{-\sigma l}\cong 0$. Here let us consider separately the cases of low and high modulation frequencies of the exciting radiation. We begin with the case of low frequencies, for which $\kappa\rho_0\omega<(f_3^{(3)}+h_3^{(3)})$. Then, using expression (19), we obtain for the signal from the piezoelectric transducer

$$V(\omega)=iC\kappa\rho_0^{3/2}\omega^2\frac{\gamma_0(1+A^{(3)})(1+\beta U_{33})\Delta T_s}{\left\{(1+A^{(3)})\left(K+\frac{4}{3}\mu\right)+[t_{33}^{(0)}+(1+A^{(3)})(2lU_{pp}+(4m+n)U_{33})]\right\}^{3/2}}. \tag{20}$$

In the region of high modulation frequencies of the exciting radiation, for $\kappa\rho_0\omega>(f_3^{(3)}+h_3^{(3)})$, expression (19) leads to the result

$$V(\omega)=-C\rho_0^{1/2}\omega\frac{\gamma_0(1+A^{(3)})(1+\beta U_{33})\Delta T_s}{\left\{(1+A^{(3)})\left(K+\frac{4}{3}\mu\right)+[t_{33}^{(0)}+(1+A^{(3)})(2lU_{pp}+(4m+n)U_{33})]\right\}^{1/2}}. \tag{21}$$

Expressions (20) and (21) can be used to analyze some general behavioral regularities of the piezoelectric signal from objects containing internal stresses. In this paper we shall consider only the physical and not the geometrical nonlinearities,[19,22] i.e., we shall assume that $A^{(3)}\ll 1$. In accordance with the results of Ref. 17 the coefficient β is positive. Therefore, the presence of tensile stresses in the sample enhances the thermoelastic coupling and increases the piezoelectric signal, while the presence of compressive stresses diminishes it.

Besides the quantities which determine the dependence of the piezoelectric signal on the Murnaghan constants and the quantities appearing explicitly in expressions (20) and (21), the initial stress tensor $t_{33}^{(0)}$ also in general has such a dependence. However, in accordance with expression (5), the dependence on the Murnaghan constants in the tensor $t_{33}^{(0)}$ enters only through terms quadratic in the initial strains. Since even under substantially nonlinear conditions the initial strains are usually characterized by quantities much less than unity, the influence of the Murnaghan constants on the piezoelectric signal arises mainly through the terms linear in the initial strains and which are written out explicitly in the denominators of expressions (20) and (21).

Expressions (20) and (21) can also be used to analyze the possibility of determining the Murnaghan constants of a material under the conditions of the problem under discus-

sion. For example, we see from the two expressions that measurement of the dependence of the piezoelectric signal on the initial strain U_{11} (or U_{22}) in the absence of a strain U_{33} will enable one to determine the Murnaghan constant l. For a known l expressions (20) and (21) can be used to determine the coefficient β and the quantity $(4m+n)$ from the dependence of the piezoelectric signal on the initial strain U_{33} for $U_{11}=U_{22}=0$. However, it is not possible here to determine the constants m and n independently within the framework of the problem under discussion.

In Ref. 15 experiments on the excitation of mechanical vibrations in loaded titanium rods by laser radiation were reported. The experiments were in fact done under conditions of low-frequency excitation of acoustic vibrations. In agreement with expression (20) it was found that the piezoelectric signal increases when mechanical vibrations are excited by laser radiation in the zones where tensile stresses exist and decreases in regions of compressive stress. On the basis of data on the value of the coefficient β for metals[17] one can estimate the influence of internal stresses on the piezoelectric signal. For example, for the conditions of Ref. 15 expression (20) shows that the stress dependence of the thermoelastic coupling coefficient leads to roughly a 10% change in the piezoelectric signal.

This value is somewhat smaller than that obtained in Ref. 15. However, it should be kept in mind that for the majority of metals the Murnaghan constants have negative values. Then, in accordance with expressions (20) and (21), the stress dependence of the piezoelectric signal due to mechanical nonlinearities will be analogous to the dependence on the thermoelastic coupling coefficient. Under these conditions the total change in the piezoelectric signal will be somewhat larger than the value due solely to the change in the thermoelastic coupling coefficient. Unfortunately, it is difficult to make a more detailed estimate of the piezoelectric signal on account of the lack of data on the Murnaghan constants for titanium. Strong changes in the piezoelectric signal have also been recorded near the ends of cracks formed by Vickers indentation[13,16,24] in ceramics. It has been shown[24] that the changes in the piezoelectric signal in this case are due to the presence of internal stresses. Here, in accordance with the results of Ref. 16, the size of these changes ordinarily amount to several tens of percent.

In summary, the proposed theory enables one to determine the basic features of the generation of acoustic vibrations by laser radiation in solid objects containing internal stresses in the framework of the nonlinear theory of thermoelasticity and also to explain qualitatively the existing experimental results. To achieve quantitative agreement of theory with experiment it will be necessary to develop the theory further to more fully take into account the peculiarities of the experimental investigations.

[1] Y. H. Pao and W. Sachse, in *Physical Acoustics*, Vol. XVII (edited by W. P. Mason and R. N. Thurston), Academic Press, New York (1984), pp. 61–143.
[2] T. Iwaoka, S. Yokogama, and Y. Osaka, Jpn. J. Appl. Phys. **24**, 112 (1984).
[3] I. I. Vlasov, V. G. Ralchenko, E. D. Obraztsova *et al.*, Appl. Phys. Lett. **71**, 1789 (1997).
[4] B. Eigenmann, B. Scholtes, and E. Macherauch, Mater. Sci. Eng. **118**, 1 (1989).
[5] M. R. Daymond, M. A. M. Bourke, R. B. Von Dreele *et al.*, J. Appl. Phys. **82**, 1554 (1997).
[6] M. Ceretti, C. Braham, J. L. Lebrun *et al.*, Exp. Tech. **20**, 14 (1996).
[7] É. S. Gorkunov and M. V. Tartachnaya, Zavod. Lab. **59**, 22 (1993).
[8] G. N. Chernyshev, A. L. Popov, V. M. Kozintsev, and I. I. Ponomarev, *Residual Stresses in Strained Solids* [in Russian], Nauka, Fizmatlit, Moscow (1996), 240 pp.
[9] M. J. Pechersky, R. F. Miller, and C. S. Vikram, Opt. Eng. (Bellingham) **34**, 2964 (1995).
[10] C. S. Vikram, M. J. Pechersky, C. Feng, and D. Engelhaupt, Exp. Tech. **20**, 27 (1996).
[11] M. Kasai and T. Sawada, *Photoacoustic and Photothermal Phenomena II*, Vol. 63 of Springer Series in Optical Sciences, Springer-Verlag, New York (1990), pp. 33–36.
[12] R. M. Burbelo, A. L. Gulyaev, L. I. Robur *et al.*, J. Phys. (France) **4**, Colloque C7-311 (1994).
[13] H. Zhang, S. Gissinger, G. Weides, and U. Netzelmann, J. Phys. (France) **4**, Colloque C7-603 (1994).
[14] M. Qian, Chin. J. Acoust. **14**, 97 (1995).
[15] R. M. Burbelo and M. K. Znabitenko, in *Progress in Natural Science*, Taylor & Francis, London; Washington (1996), Suppl. Vol. 6, pp. 720–723.
[16] K. L. Muratikov, A. L. Glazov, D. N. Rouz *et al.*, Pis'ma Zh. Tekh. Fiz. **23**(5), 44 (1997) [Tech. Phys. Lett. **23**(3), 188 (1997)].
[17] R. I. Garber and I. A. Gindin, Fiz. Tverd. Tela (Leningrad) **3**(1), 176 (1961) [Sov. Phys. Solid State **3**, 127 (1961).
[18] T. Bateman, W. P. Mason, and H. J. McSkimin, J. Appl. Phys. **32**, 928 (1961).
[19] A. N. Guz', Prikl. Mekh. **2**, 3 (1970).
[20] T. Tokuoka and M. Saito, J. Acoust. Soc. Am. **45**, 1241 (1969).
[21] A. I. Lur'e, *Nonlinear Theory of Elasticity* [in Russian], Nauka, Moscow (1980), 512 pp.
[22] V. V. Novozhilov, *Fundamentals of the Nonlinear Theory of Elasticity* [in Russian], Gostekhizdat, Leningrad–Moscow (1948), 211 pp.
[23] W. Jackson and N. M. Amer, J. Appl. Phys. **51**, 3343 (1980).
[24] J. H. Cantrell, M. Qian, M. V. Ravichandran, K. W. Knowles, Appl. Phys. Lett. **57**, 1870 (1990).

Translated by Steve Torstveit

Quantitative x-ray topographic analysis of the defects of 6*H*-SiC single crystals and homoepitaxial silicon carbide

G. F. Kuznetsov

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 141120 Fryazino, Moscow Region, Russia

(Submitted December 30, 1996; resubmitted December 19, 1997)
Zh. Tekh. Fiz. **69**, 64–67 (July 1999)

Planar dislocation pileups (PDPs) and curvilinear dislocation segments (CDSs) are considered as indicators of the local elastic shear stress fields (LESSFs) that existed in the growing single crystals at the time of stabilization of their dislocation structure. Calculations using the theory of dislocations with the experimental parameters of PDPs and CDSs measured from the x-ray topograms (taken by the Lang and divergent-polychromatic-beam (DPB) methods) give values of the LESSFs in the range $(0.2-1.5)\times 10^6$ Pa for thin single-crystal wafers of SiC ($6H$) grown by sublimation in a graphite container. A strong nonuniform bending of the single-crystal wafers is observed; for the x-ray topographic study of the dislocation structure in these wafers the DPB method is preferable to the Lang method on account of its low sensitivity to bending. © *1999 American Institute of Physics.* [S1063-7842(99)01307-0]

The use of silicon carbide wafers grown by sublimation in the Lely method as substrates for the growth of epitaxial structures for making a wide range of high-power microelectronic and optoelectronic microdevices (see, e.g., Ref. 1) makes it necessary to study the various types of defects in these materials. A practical outcome of such studies might be the optimization of the technology for preventing the formation of certain types of defects both during the growth of the single-crystal substrates and epitaxial structures and also during the fabrication of the microdevices. In addition to the electrophysical, optical, and electron-microscope methods, the development and production of efficiently operating and reliable microdevices based on SiC will also require nondestructive methods of quantitative x-ray topographic (QXTG) and diffractometric (QXD) analysis of the defects of the single-crystal substrates, the grown epitaxial structures, and the actual device structures fabricated on them.

The QXTG and QXD methods occupy a special niche among methods of investigating structural defects in single crystals and epitaxial systems. The quantitative data provided by these methods has played an important role in the laboratory development and industrial production of reliable injection heterolasers emitting cw at room temperature at wavelengths of 0.85, 1.30, and 1.55 μm.[2] The data provided by the QXTG and QXD methods will no doubt provide a physical basis for speeding up the optimization of the growth processes and the development and production of reliable microdevices with the necessary output parameters in the case of SiC technology as well.

In this paper we employ the methods of QXTG analysis to study thin (150–400 μm) wafers of α-SiC of structural type $6H$, grown by sublimation and precipitation on the walls of a graphite container. The wafers had hexagonal faceting with the greatest development of the $\{0001\}$ basal planes, and a diameter of up to 12 mm.

For studying the dislocations we used the direct x-ray topographic methods of Lang[3] in collimated beams of the characteristic $K\alpha_1$ and $K\alpha_2$ radiation of a molybdenum anode and a divergent polychromatic beam from a quasipoint source (DPB)[2,4] from the same anode. For such radiation the value of μt lay in the range 0.07–0.18. This provided extinction contrast of the dislocations recorded on the topograms, both those taken by the Lang method and those taken by the DPB method. In the Lang method the topograms were taken successively at symmetric reflections of the $\{11\bar{2}0\}$ and $\{30\bar{3}0\}$ type.

In contrast to the Lang method, the DPB method is a method of multiple-beam x-ray topography.[2,4–6] In the DPB method from three to five reflections of the $\{11\bar{2}0\}$, $\{3\bar{3}00\}$, and $\{22\bar{4}0\}$ type for the fundamental zonal ellipse, whose axis $[0001]$ is perpendicular to the basal surface of the crystals and around 15 reflections of the other zonal ellipses were taken on the same photographic plate.

Figure 1 shows the a panoramic topogram of one of the silicon carbide crystals, taken by the Lang method in the $(11\bar{2}0)$ reflection. On it are recorded two comparatively narrow curvilinear bands with a clear image of a set of nonintersecting dislocation lines. In the vast region between these bands the images of the dislocation lines are practically invisible or appear with extremely low contrast. This means that the given $6H$-SiC single crystal, and, incidentally, all the other silicon carbide single crystals investigated in this study, has a pronounced nonuniform macroscopic bending. Because of the high sensitivity of the Lang method to bending, a large part of the area of the investigated single crystals in Laue diffraction and linear scanning is out of the reflecting position, and the defect structure of these regions is not revealed, whereas on those parts of the topograms from the regions of the same crystal which are found in the reflecting position for the characteristic $K\alpha_1$ and $K\alpha_2$ lines of the an-

 © 1999 American Institute of Physics

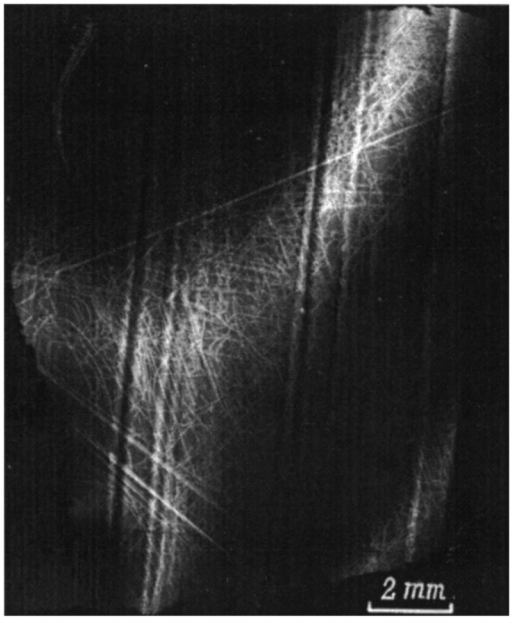

FIG. 1. Panoramic topogram of one of the silicon carbide crystals ($6H$-SiC), taken by the Lang method in the $(11\bar{2}0)$ reflection of strictly collimated characteristic $K\alpha_1$ and $K\alpha_2$ radiation of a molybdenum anode.

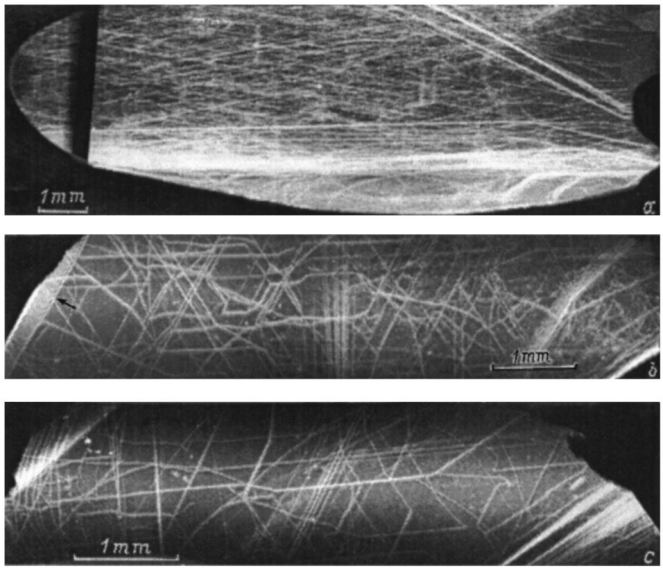

FIG. 2. Topograms of one of the diffractional reflections $(30\bar{3}0)$ of the fundamental zonal ellipse, taken by the DPB method. a: The topogram demonstrates the direction of convexities of the dislocation half-loops and the individual curvilinear dislocations; the arrow indicates the region where the growing single-crystal wafer touched the wall of the graphite crucible; b,c: typical one-sided planar dislocation pileups (PDPs) for curvilinear (b) and rectilinear (c) dislocations squeezed by elastic stresses into the corner region (b) or parallel to the crystallographic planes of the natural lateral faceting of the hexagonal single-crystal wafers (c).

ode, the dislocation structure of the crystals is revealed with excellent contrast and with a high linear resolution (around 3 μm in Fig. 1).

As we see from the topograms of the other $6H-$SiC crystals, the configurations of the reflecting regions are very different. For the given crystal the shape of the reflecting regions and their position relative to the edges of the crystal depend on the type of reflecting planes and the precise orientation of the crystal relative to the incident beam. All of this points to an extremely strong nonuniform bending of the investigated single-crystal wafer of $6H-$SiC. The observed nonuniform bending strongly complicates the study of the defect structure of $6H-$SiC crystals by the Lang method.

The DPB x-ray topographic method, which permits obtaining rather high image contrast for individual dislocations in polychromatic radiation[2,4–6] is, because of this, significantly less sensitive than the Lang method to the effects of a nonuniform bending of the single-crystal wafers. Figure 2 shows a topogram of one of the diffractional reflections $(30\bar{3}0)$ of the fundamental zonal ellipse, taken by the DPB method. It clearly reveals the dislocation structure of this nonuniformly bent $6H-$SiC wafer. This same image was recorded in two other reflections, $(11\bar{2}0)$ and $(2\bar{1}\bar{1}0)$, of the fundamental zonal ellipse and also in certain reflections of other zonal ellipses. The x-ray topograms taken by the DPB method for the other $6H-$SiC samples also quite clearly revealed their dislocation structure. On them one can easily follow the extinction of the images of individual dislocations as the magnitude and direction of the diffraction vectors changed, making it possible to identify the types of the dislocations.

The Lang and DPB methods in $6H-$SiC crystals revealed dislocations localized mainly in the basal planes. They have Burgers vectors of the type $b_1=1/3\langle 10\bar{1}0\rangle$ and $b_2=1/3\langle 11\bar{2}0\rangle$. The maximum linear density of dislocations in the $6H-$SiC samples studied was $N_d\leqslant 6\times 10^2$ cm^{-1} in the basal plane, which corresponds to a dislocation density $N_d\leqslant 6\times 10^4$ cm^{-2} for the $\{hk10\}$ planes, perpendicular to the basal glide planes of the dislocations. Besides curvilinear dislocations one also observes strictly rectilinear (sectorial) dislocations and also rectilinear dislocations with 120-degree bends, the individual rectilinear segments of which coincide with the crystallographic directions of the $\langle 11\bar{2}0\rangle$ type.

My main goal was to elucidate the physical causes of the generation of dislocations in the growing $6H-$SiC single-crystal wafers and to determine the location of the actuated dislocation sources that have generated the dislocations. It is also necessary to ascertain the stage in which the generation and propagation of the dislocations occurs — during growth or during the cooling of the grown crystals.

Judging from the direction of the convexities of the dislocation half-loops and the individual branches of the curvilinear dislocations (Fig. 2a), their sources were localized in the region where the growing single-crystal wafer of $6H-$SiC touched the wall of the graphite crucible. Theoretically the calculation of thermoelastic stresses acting on the region where the crystal is fastened to the wall of the crucible showed that because of the large differences in the coefficients of thermal expansion and in the behavior of these coefficients as functions of temperature, thermoelastic stresses arise in the region where the graphite touches the single-crystal $6H-$SiC wafer, and these stresses, increasing as the grown crystal and crucible are cooled to room tem-

perature, can exceed the yield strength of graphite, $(3-44)\times 10^6$ Pa.[11] This means that the single crystals may pull away from the walls of the container during cooling. The evidence confirms that this happened more than once. This provides grounds for assuming that the generation of dislocation half-loops in single-crystal $6H-$SiC wafers most likely occurs during cooling of the container after the growth of the single crystals.

In a number of samples we recorded typical one-sided planar dislocation pileups (PDPs) of curvilinear dislocations squeezed by elastic stresses in the basal planes into the corner regions of the hexagonal single-crystal wafers (Fig. 2b). We also recorded planar pileups of rectilinear dislocations (Fig. 2c) parallel to the crystallographic planes of the natural lateral faceting of the wafers.

Using the dislocations as indicators, in this paper we have used the parameters of the curvilinear dislocation segments of individual dislocations[7] and groups of planar dislocation pileups (PDPs),[2,8,9] as obtained from measurements on the topograms and from calculations, to calculate, for the first time for SiC $(6H)$, the local elastic stresses τ_l acting in specific growing crystals on account of these dislocations. Here for the PDPs the values of τ_l were calculated using the formula from Ref. 8:

$$\tau=\{[G/(1-\nu)]\sin\alpha_n+G\cos\alpha_n\}bN_n/\pi L_n, \qquad (1)$$

where $G=2.6\times 10^{10}$ Pa is the shear modulus, $\nu=0.20$ is Poisson's ratio,[10] $b=3.1\times 10^{-10}$ m is the Burgers vector of the dislocations for $6H-$SiC, α is the angle of the Burgers vector relative to the dislocation lines, and L_n is the length and N_n the number of dislocations in any of the PDPs.

PDP-based estimates of the elastic stresses acting in $6H-$SiC single-crystal wafers under real growth conditions gave for the local shear stress the value $\tau_{l1}\approx 1\times 10^6$ Pa (or 100 Js/mm^2).

Calculations according to the measured radii of curvature of the segments of individual curvilinear dislocations[7] gave a range of values of the local elastic stresses $\tau_{l2}\approx 2\times 10^5-1.5\times 10^6$ Pa (or 20–150 Js/mm^2). We see that the values of the local elastic stresses obtained by different methods for the same $6H-$SiC crystal agree with one another and should be trustworthy. The range of values obtained for the local elastic shear stresses in real $6H-$SiC crystals is close to the critical values of the shear stresses necessary for excitation of the dislocation sources that generate dislocations in the basal glide planes at high temperatures.

The sinuosity observed for some of the dislocation lines is a sign of their stopping by impurities, which precipitate out, apparently incoherently, in the active glide planes.

In a DPB study of $6H-$SiC single crystals with a thickness of 400 μm and an unetched surface, defects of the surface on the upper basal planes were observed after growth for the natural faceting of the crystals. In Fig. 3 these surface defects are recorded in the form of comparatively small local regions of black-and-white contrast. Their diffraction dimensions are from 50 to 200–1000 μm, and they are distributed with a density of up to 1×10^4 cm^{-2}. Their physicochemical nature can be only hypothesized. Most likely they

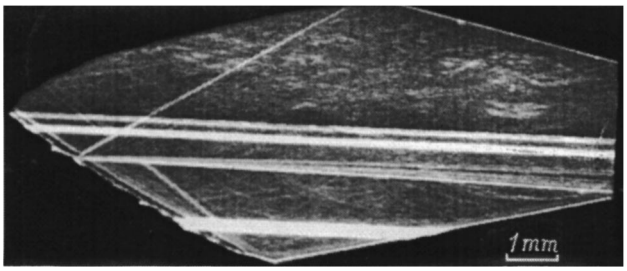

FIG. 3. Topogram of defects of the surface of one of the $6H-$SiC crystals, recorded in the form of different-sized local regions of black-and-white contrast. The topogram was taken by the DPB method in the $(11\bar{2}0)$ reflection of the polychromatic and characteristic line radiation of the molybdenum anode.

are island deposits of some sort of reaction products which interfere with the mechanism of layered growth of the homoepitaxial single-crystal film. However, physically they play the role of local concentrators of the macroscopic elastic stress field due to the nonuniform bending of the single-crystal $6H-$SiC wafers grown and cooled to room temperature.

The topograms of such samples revealed only individual dislocations and with rather low contrast. It is as if they are ''leveled out'' or weakly transilluminated in a fog. The suppression of the image of bulk defects by surface defects is an effect that is generally familiar to x-ray topographers. The author himself has published several papers on this subject.[2,4,6,11,12] After a chemical processing of the surface of the crystals the images of the surface defects vanish from the topograms. Consequently, the surface defects themselves must also vanish from the surface of the single crystals as a result of chemical dissolution, which indirectly confirms their surface island character.

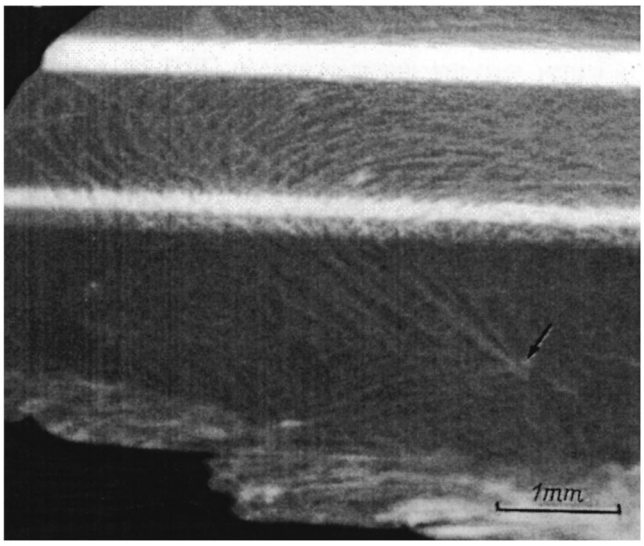

FIG. 4. Defects of the structure of one of the samples of a homoepitaxial $6H-$SiC film–substrate system, revealed by the DPB method in the $(11\bar{2}0)$ reflection of the polychromatic and of the characteristic line radiation of a molybdenum anode.

The DPB method was also used to study the defect structure of one of the samples of a homoepitaxial $6H-$SiC film–substrate system. It became clear that the defect structure of the substrate of this homoepitaxial single-layer structure is characterized by an extremely strong dislocation network, the dislocations of which are emitted both by sources at the crystal–container heterointerface and by an internal source lying in a region of the crystal far from the heterointerface (the arrow in Fig. 4). The contrast of the substrate dislocations is lowered by the influence of the nonuniform elastic field of the epitaxial film.

In some of the DPB topograms the region of the epitaxial film was partially shifted relative to the substrate. In this region one can clearly see that the distribution of the intensity over the area of the topogram, on which the image was produced by polychromatic x radiation, is discrete in the azimuthal and Bragg directions. This contrast is similar to that which we observed in Ref. 13 for mosaic single-crystal films of CdTe grown on mica substrates. For this reason one can draw the preliminary conclusion that the grown single-crystal film is mosaic. To us this means that the growth conditions of the $6H-$SiC homoepitaxial film were such that the operative growth mechanism was not the layer-by-layer growth that is typical for ideal homoepitaxy but rather growth by the heterogeneous formation and growth of a set of three-dimensional nuclei. Some of them, in growing, transform into mosaic blocks with small angles of mutual and maximum misorientation. The three-dimensional nucleus formation during homoepitaxy might be stimulated by surface contaminants on the substrate, which in other samples have been observed in x-ray topography as defects of a surface layer of the single crystals.

[1] A. A. Lebedev, A. N. Andreev *et al.*, in *Proceedings of the VII International Symposium on Power Semiconductor Devices and IC*, Yokohama, Japan (1995), pp. 90–95.
[2] G. F. Kuznetsov, Doctoral Dissertation [in Russian], Moscow (1989), 466 pp.
[3] A. R. Lang, in *Direct Methods of Studying Defects in Crystals*, Mir, Moscow (1965), pp. 205–222; 259–267.
[4] G. F. Kuznetsov, Kristallografiya **21**, 847 (1976) [Sov. Phys. Crystallogr. **21**, 485 (1976)].
[5] G. F. Kuznetsov, Apparatura i Metody Rentgenovskogo Analiza, No. 12, pp. 162–167 (1973).
[6] G. F. Kuznetsov and S. A. Semiletov, Obzory po élektronnoĭ Tekhnike. Ser. Mikroélektronika (TsNII Élektronika, Moscow), No. 1 (1975).
[7] G. F. Kuznetsov, Obzory po Élektronnoĭ Tekhnike. Ser. Mikroélektronika (TsNII Élektronika, Moscow), No. 1 (1975).
[8] G. F. Kuznetsov, IRÉ AN SSSR Preprint No. 2(441) [in Russian], Institute of Radio Engineering and Electronics, Academy of Sciences of the USSR, Moscow (1986), 31 pp.
[9] G. F. Kuznetsov, Kristallografiya **34**, 765 (1989) [Sov. Phys. Crystallogr. **34**, 456 (1989)].
[10] J. R. McLaren, G. Tappin, and R. W. Davidge, Proc. Brit. Ceram. Soc., No. 20, 259–265 (1972).
[11] A. R. Ubbelohde and F. A. Lewis, *Graphite and Its Crystal Compounds* [Clarendon Press, Oxford (1960); Mir, Moscow (1965), 265 pp.].
[12] G. F. Kuznetsov, Élektronnaya Tekhnika. Ser. 8, No. 3, pp. 39–65 (1978).
[13] G. F. Kuznetsov and S. A. Semiletov, Kristallografiya **22**, 664 (1977) [Sov. Phys. Crystallogr. **22**, 381 (1977)].

Translated by Steve Torstveit

Properties and structure of GaAs films grown by molecular beam epitaxy on GaAs substrates with the (100), (111)*A*, and (111)*B* orientations

G. B. Galiev, V. G. Mokerov, Yu. V. Slepnev, and Yu. V. Khabarov

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 103907 Moscow, Russia

A. A. Lomov and R. M. Imamov

Institute of Crystallography, Russian Academy of Sciences, 117333 Moscow, Russia
(Submitted March 9, 1998)
Zh. Tekh. Fiz. **69**, 68–72 (July 1999)

The structural perfection of GaAs epitaxial films grown by molecular beam epitaxy on substrates with the (100), (111)*A*, and (111)*B* orientations is investigated by double-crystal and triple-crystal x-ray diffractometry. It is found that the ratio γ of the molecular fluxes of arsenic and gallium has a strong influence on the structural quality of the epitaxial films. The optimum values of the parameter γ are found for each of the substrate orientations (100), (111)*A*, and (111)*B*. © *1999 American Institute of Physics.* [S1063-7842(99)01407-5]

INTRODUCTION

Silicon is widely used as a doping impurity in the growth of epitaxial structures of III–V semiconductor compounds, in particular, in GaAs and its solid solutions. In molecular beam epitaxy (MBE) on GaAs substrates with the (100) orientation, epitaxial films with *n*-type conductivity are formed. In the case of silicon doping an electron concentration of up to 6×10^{18} cm^{-3} can be achieved in the films. It is also known that when (111)*A* substrates are used for MBE, silicon exhibits amphoteric properties:[1–3] at small values of γ($\gamma=P_{As}/P_{Ga}$, where P_{As} and P_{Ga} are the vapor pressures of As and Ga in the growth zone of the epitaxial films in the MBE apparatus) layer of predominantly *p*-type conductivity form on account of the incorporation of the Si atoms in the As sublattice, while at large γ one obtains films of *n*-type conductivity. This phenomenon opens up new possibilities for fabrication novel semiconductor devices with lateral $p-n$ junctions, buried-heterostructure quantum-well lasers, and other devices. Therefore, the epitaxial growth of III–V layers on substrates of different orientations, with the use of a single dopant — silicon — is of great scientific as well as practical interest.[4–7] However, analysis of the the published data shows that there is significant disagreement in the experimental results obtained by different authors. For example, in Ref. 5 it was found that the change in the conductivity type from *p* to *n* in epitaxial films grown on substrates with the (111)*A* orientation occurs at $\gamma>5$, whereas according to the data of Ref. 6, this occurs at $\gamma>48$. There are no data on the dependence of the quality of the crystal lattices of films grown on substrates of different orientations or under different growth conditions.

In this paper we report a study of the influence of the ratio of the As and Ga fluxes on the electrophysical and structural properties of GaAs epitaxial films grown by MBE on GaAs substrates with the (100), (111)*A*, and (111)*B* orientations, with Si used as a dopant.

FILM GROWTH AND INVESTIGATION PROCEDURES

Epitaxial films of GaAs were grown by MGE on semi-insulating GaAs substrates with the (100), (111)*A*, and (111)*B* orientations. In order that the films be grown under the same conditions, the samples were grown three at a time, on the (100), (111)*A*, and (111)*B* substrates, for each value of γ. A specified value of γ was reached by changing the temperature of the As cell, while the temperature of the Ga cell was held constant ($T_{Ga}=960\,^{\circ}$C) for all the growth processes. The substrate temperature T_g during growth was 600 °C. The thickness of the undoped buffer layer was $\sim0.5\,\mu$m, with an impurity concentration of 5×10^{14} cm^{-3}, and the thickness of the silicon-doped layer was $\sim0.20\,\mu$m. The temperature of the Si molecular source corresponded to the value at which the electron concentration n_e in the doped layer on the (100) substrate was $\sim1\times10^{18}$ cm^{-3}. For the nonconducting samples the thickness of the doped layer was increased to 1–1.5 μm. For all the samples the conductivity type and carrier concentration were measured by the $C-V$ method.

TABLE I.

Sample No.	γ	Substrate orientation	Conductivity type	Carrier concentration, cm^{-3}
1	16	(100)	*n*	1×10^{18}
2	16	(111)*A*	*p*	8.0×10^{16}
3	16	(111)*B*	*n*	$<10^{15}$
4	36	(100)	*n*	9.6×10^{17}
5	36	(111)*A*	*p*	2.2×10^{16}
6	36	(111)*B*	*n*	$\simeq10^{15}$
7	50	(100)	*n*	9.7×10^{17}
8	50	(111)*A*	*n*	4.1×10^{17}
9	50	(111)*B*	*n*	$\simeq10^{15}$
10	77	(100)	*n*	1.1×10^{18}
11	77	(111)*A*	*n*	9.0×10^{17}
12	77	(111)*B*	*n*	8.5×10^{17}

 © 1999 American Institute of Physics

The structural perfection of the grown epitaxial films was studied on a triple-crystal x-ray spectrometer; here slotted single crystals of Ge(111) or Ge(004), with threefold reflection, were used as monochromators, and Ge(111) or Ge(001) single crystals were used as the analyzer crystals. The x-ray beam illuminated a region of dimensions 1 $\times 4$ mm on the sample. Both reflection diffraction curves (RDCs) and triple-crystal x-ray diffraction (TXD) spectra were recorded.

RESULTS AND DISCUSSION

The growth conditions of the epitaxial films, the carrier concentration, and the type of conductivity of the samples are presented in Table I. We see that all of the epitaxial films grown on substrates with the (100) orientation (samples *1*, *4*, *7*, and *10*) have *n*-type conductivity, with an electron concentration $n_e \sim 10^{18}$ cm^{-3} for all values of γ. The films grown on (111)*B* substrates, while also exhibiting *n*-type conductivity, have small values of n_e ($n_e \leqslant 10^{15}$ cm^{-3}) for $\gamma < 50$, and only for $\gamma = 77$ does $n_e = 8.5 \times 10^{17}$ cm^{-3}, which is smaller than for epitaxial films grown on (100) substrates. The situation for epitaxial films grown on (111)*A* substrates is more complicated. In our experiments, as in Ref. 6, at small γ the doping of the epitaxial film by silicon leads to *p*-type conductivity (sample *2*) with $n_p \cong 8.0 \times 10^{16}$ cm^{-3}. As γ is increased from 16 to 36 there is a decrease in n_p to 2.2×10^{16} cm^{-3}. Apparently for $\gamma = 36$ there is substitution of Si atoms in the Ga and As sites simultaneously, which leads to partial compensation of the conductivity. As γ is increased further to 50 there is a change in the type of conductivity (from *p* to *n*), and at $\gamma = 77$ the concentration n_e reaches values $\sim 9.0 \times 10^{17}$ cm^{-3}.

Let us now consider the results of x-ray diffraction studies of the structure of epitaxial films. Figure 1 shows RDCs measured by double-crystal diffractometry. Table II gives values of the reflection coefficients P_r and the half-widths W for both the virgin surfaces and for the epitaxial samples (the sample number can be found in Table I from the parameter γ and the substrate orientation). As we see from Table II, of the samples with the (100) orientation sample *1* had values of P_r and W most closely corresponding to those for the virgin surfaces. This is apparently because that sample was grown at a partial pressure γ closest to the optimum growth conditions for the (100) orientation. As γ was increased, P_r decreased and the half-width W increased; these changes attest to increasing distortions of the structure in the epitaxial film. The largest deviations of the parameters of the RDCs from the values for the virgin surfaces were observed for $\gamma = 36$. Apparently at this ratio of As and Ga pressures there is appreciable formation of structural defects, with a decomposition of the solid solution due to nonuniform growth conditions. At large values of γ the process stabilizes, and the formation of epitaxial films occurs under favorable conditions for stoichiometric growth, as is reflected in the improvement of the parameters of the RDCs (sample *10*). For illustration in Fig. 2 we show the dependence of the reflection coefficients P_r on γ for the samples studied. It follows from Fig. 2 and Table II that the largest changes of P_r are

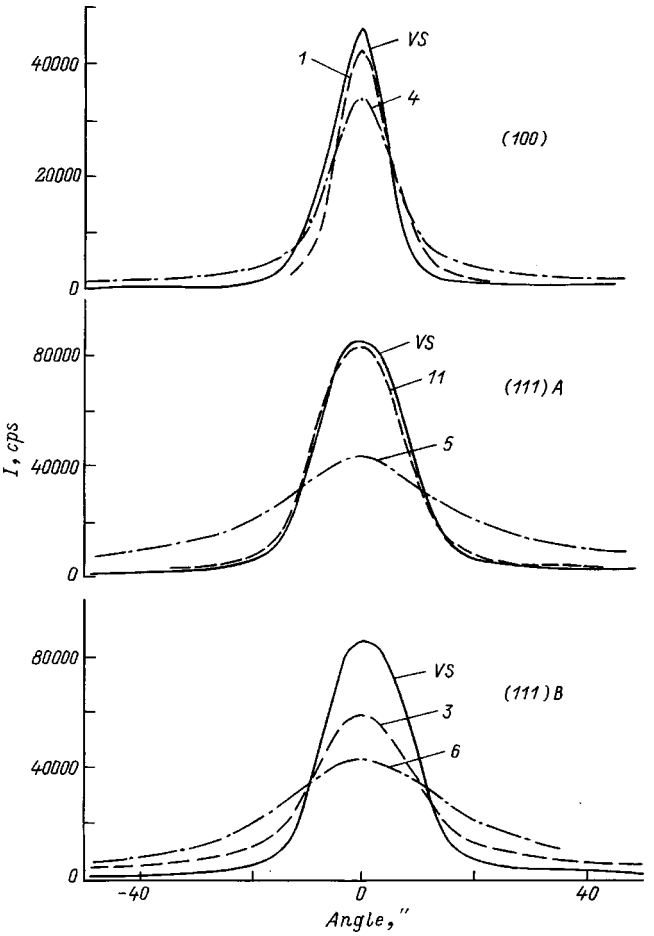

FIG. 1. Reflection diffraction curves (RDCs) for GaAs epitaxial films grown on AaAs substrates with the (100), (111)*A*, and (111)*B* orientations. The curves are labeled with the sample Nos; *VS* is the virgin surface.

observed for samples grown on substrates with the (111)*A* orientation. Since a decrease of P_r and an increase of W are indicative of an increasing degree of structural distortions in the sample,[8] the largest number of defects would be expected to form in the range $20 < \gamma < 40$ (in the case of silicon doping), corresponding to the transition from *p*-type to *n*-type conductivity in epitaxial films with the (111)*A* orientation,.

To check this we analyzed the TXD spectra. The spectra were recorded by scanning the analyzer crystal with the position of the sample held fixed, with an angular detuning from the exact Bragg angle of $\alpha = \theta - \theta_B$ (θ_B is the Bragg

TABLE II.

	Orientation					
	(111)*A*		(111)*B*		(100)	
	P_r, %	W, ''	P_r, %	W, ''	P_r, %	W, ''
Virgin surfaces	85	18.9	86	18.3	66	11.3
$\gamma = 16$	72	23.2	59	22.0	60	11.7
36	43	37.3	43	38.0	49	13.9
50	81	19.5	53	27.2	59	13.6
77	82	18.9	43	45.6	58	13.0

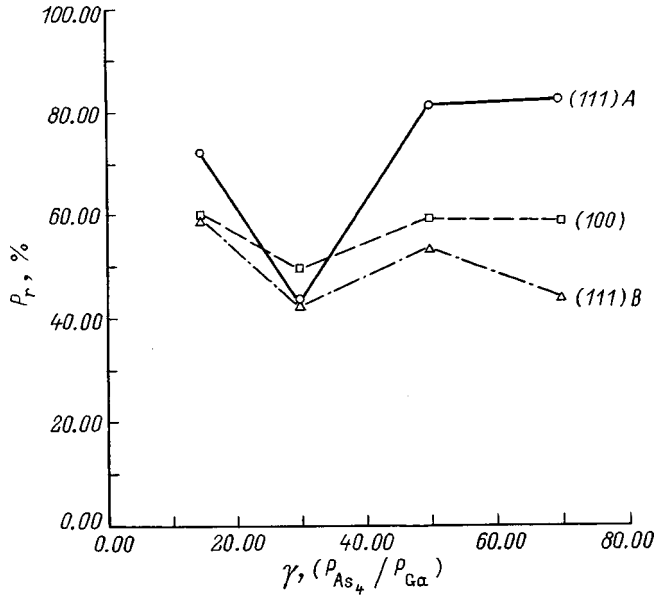

FIG. 2. Coefficient of reflection P_r of GaAs epitaxial films grown on GaAs substrates with the (100) and (111)*A* orientations, plotted versus the ratio of the arsenic and gallium fluxes.

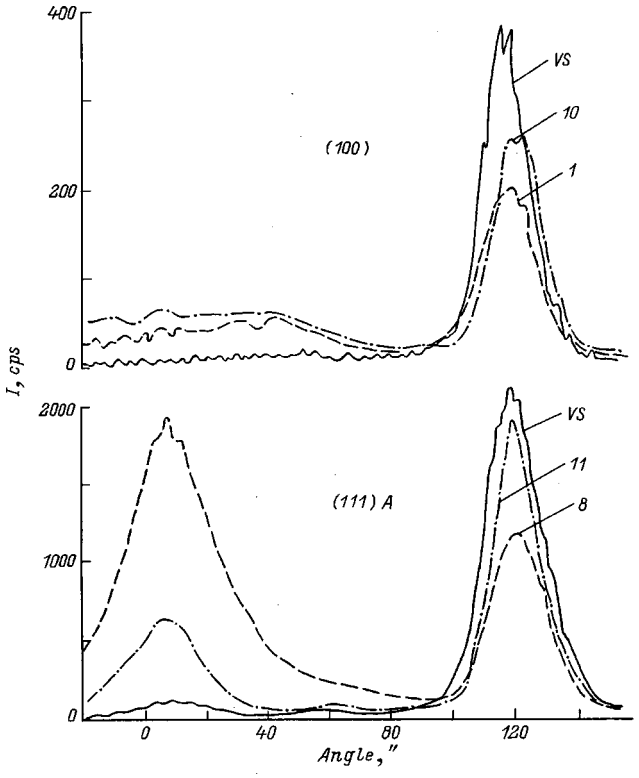

FIG. 3. Triple-crystal x-ray diffraction (TXD) spectra for GaAs epitaxial films grown on GaAs substrates with the (100) and (111)*A* orientations. The curves are labeled by the sample Nos; *VS* is the virgin surface.

angle), as the angle of the analyzer crystal was scanned. In this arrangement, the presence of defects in the sample will give rise to a diffuse peak (a broad hump at $\theta \approx 0$) in addition to the peak due to the Bragg component of the scattering (narrow peak at $\theta = 22''$); the intensity, shape, and angular position of this diffuse peak are determined entirely by the number and type of the defects. The TXD spectra of the GaAs(100) virgin surfaces and the epitaxial films grown on substrates of this orientation are shown in Fig. 3. We see that on the curve *VS* (virgin surface) the intensity near $\theta \cong 0$ is very low (slightly above background) and is constant. At the same time, the diffuse scattering from sample *1* (curve *1*) is easily seen. As the partial pressure is raised ($\gamma = 77$) the diffuse maximum increase (curve *10*). This result indicates that the epitaxial growth conditions on substrates with the (100) orientation are better at $\gamma = 16$. A somewhat different situation is observed for the samples grown on (111)*A* substrates. In the case when the films grown exhibit *n*-type conductivity, a large value $\gamma = 77$ corresponds to epitaxial films with the least distorted structure. In this case the diffuse peak is minimal (curve *11*) and much weaker than for sample *8*, which was grown at $\gamma = 50$. It should also be noted that the electron concentration in sample *11* is significantly higher than in sample *8*. For epitaxial films with *p*-type conductivity the intensity of the diffuse peak in the TXD spectrum is noticeably greater than that of the Bragg scattering component. Thus under the growth conditions investigated we were unable to grow silicon-doped epitaxial films with a structure comparable in perfection to the virgin surface.

CONCLUSION

In summary, the comprehensive studies reported here have shown that when GaAs films are grown by MBE on GaAs substrates in a single technological cycle with Si used as the doping impurity, the films grown on substrates with the (100), (111)*A*, and (111)*B* orientations are quite different both in terms of the degree of perfection of the structure and also in their electrical properties. It was found that the ratio of the molecular fluxes of arsenic and gallium during growth have a strong influence on the structural perfection of GaAs films. The optimum values of the parameter γ, for which the GaAs films form with the most perfect structure, were determined for substrate orientations (100), (111)*A*, and (111)*B*. It was found that the crystal structure of GaAs epitaxial films grown on substrates with the (111)*A* orientation is almost always more perfect than that of films grown on (111)*B* substrates.

[1] S. S. Bose, B. Lee, M. H. Kim, and G. E. Stillman, J. Appl. Phys. **63**, 743 (1988).
[2] W. I. Wang, R. F. Marks, and L. Vina, J. Appl. Phys. **59**, 937 (1986).
[3] S. Subbanna, H. Kroemer, and I. L. Merz, J. Appl. Phys. **59**, 488 (1986).
[4] Y. Okano, H. Seto, H. Katahama *et al.*, Jap. J. Appl. Phys. **28**, 151 (1989).
[5] Y. Okano, M. Shigeta, H. Seto *et al.*, Jap. J. Appl. Phys. **29**, 1357 (1990).
[6] F. Piazza, L. Pavesi, M. Henini, and D. Johnston, Semicond. Sci. Technol., No. 7, 1504 (1992).
[7] L. Pavesi, M. Henini, and D. Johnston, Appl. Phys. Lett. **66**, 2846 (1996).
[8] A. M. Afanas'ev, P. A. Aleksandrov, and R. M. Imamov, *X-Ray Diffraction Diagnostics of Submicron Films* [in Russian], Nauka, Moscow (1989), p. 151.

Translated by Steve Torstveit

Low-temperature diffusion of gold in germanium under the influence of atomic hydrogen

V. M. Matyushin

Zaporozhe State Technical University, 330063 Zaporozhe, Ukraine
(Submitted April 3, 1998)
Zh. Tekh. Fiz. **69**, 73–76 (July 1999)

An investigation is made of the diffusion of gold in germanium under the influence of the energy released by the recombination of hydrogen atoms to form molecules. Crystals of *n*-type germanium with gold films ($d=1\times10^{-7}$ m) are exposed to atomic hydrogen for various times (up to 10^4 s) at temperatures close to room temperature. The diffusion of gold in the germanium is analyzed by laser mass spectrometry, and also by measuring the surface resistance, the minority carrier lifetime, and the infrared transmission spectra. Mechanisms are proposed for the stimulation of heterodiffusion and accompanying processes. © *1999 American Institute of Physics.* [S1063-7842(99)01507-X]

INTRODUCTION

The high-temperature diffusion processes employed to introduce electrically active and recombination impurities into semiconductor crystals to a calculated concentration at a given depth are accompanied by many deleterious effects such as the redistribution of already implanted impurities, the formation and growth of parasitic films, the creation of appreciable mechanical stresses at the heterointerface, and many others. Thus we are faced with the problem of optimizing these heterodiffusion regimes in terms of reducing the process temperature without increasing its duration.

A drastic reduction in the overall temperature in the process system is best achieved by using local stimulation of the impurities in semiconductors. Accelerating the heterodiffusion under the influence of atomic hydrogen is an effective means of achieving this. When hydrogen atoms recombine to form molecules, a comparatively large energy is released (4.5 eV per recombination event),[1] which, when transferred to a surface impurity atom, may stimulate heterodiffusion processes in the subsurface and bulk regions of semiconductor crystals.

The present author and coworkers[2,3] investigated the low-temperature diffusion of copper and nickel in germanium under the influence of atomic hydrogen. However, the diffusion of gold atoms is of greater practical interest, since gold is a more efficient recombination impurity in germanium.

METHOD OF INVESTIGATION

The substrate material was (111)-oriented, *n*-type single-crystal germanium with $\rho=1.5\times10^3\ \Omega\cdot$m and dislocation density 10^6 m^{-2}. The wafers were etched in a 40% HF, 20% HNO_3, 40% H_2O mixture with added $AgNO_3$ to remove the damaged layer and were then ultrasonically cleaned.

Gold films were deposited by thermal evaporation in vacuum. The substrate temperature during the film deposition was of the order of 400 K, so that no significant diffusion of gold into the germanium occurred. The thickness of the films was monitored using an MAR-2 x-ray microanalyzer and was $(1-1.2)\times10^{-7}$ m. The surface state of the samples was monitored by measuring the surface resistance R_s using a four-probe method and also using an MIM-7 microscope at 500× magnification.

Atomic hydrogen was prepared by dissociation of hydrogen molecules in a plasma formed by an rf discharge. The molecular hydrogen was released from a solution of 20% KOH and distilled water by electrolysis. The hydrogen was then passed through a prefilter to collect any KOH droplets and a drying column (filled with granular KOH alkali), before being fed into the working chamber. The atomic hydrogen concentration in the working chamber was measured by a calorimetric method[4] and also by an electron paramagnetic resonance method and reached 5×10^{20} m^{-3} at a pressure of 15–20 Pa in the system.

The samples were placed on an aluminum holder 0.25 m from the discharge zone, which prevented any hydroxyl radicals and H^+ ions from reaching the sample.[4] The temperature of the samples was measured using a Chromel–Copel thermocouple attached to their surface. The samples were only heated by the energy released as a result of recombination, and their temperature did not exceed 330 K.

The distribution of the gold concentration in the germanium was investigated by laser mass spectrometry. The investigations were carried out using an EMAL-2 system in which laser sputtering was followed by separation of the ions in electric and magnetic fields and detection in an ion-sensitive material. The gold concentration was determined from the optical density of the corresponding line in the mass spectrum.

The influence of the gold atoms on the minority carrier lifetime was studied by a point contact modulation method.[5] The infrared transmission spectra of the samples were also measured using an IKS-21 spectrometer.

EXPERIMENTAL RESULTS AND DISCUSSION

The films were sputtered and became thinner under the influence of the energy released by the adsorption (2–2.5 eV)

 © 1999 American Institute of Physics

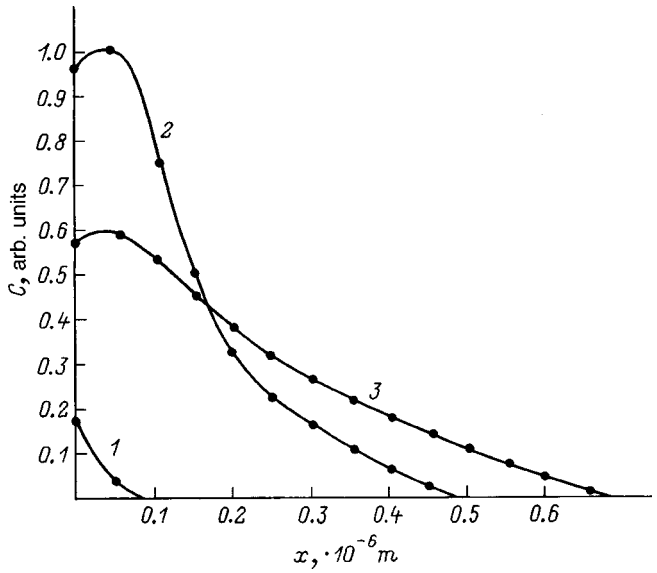

FIG. 1. Distribution of Au concentration in Ge after treatment of the samples in atomic hydrogen for various times: *1* — untreated Ge sample with Au film, *2* — Ge sample with Au film treated in H for 1800 s, and *3* — Ge sample with Au film treated in H for 3600 s.

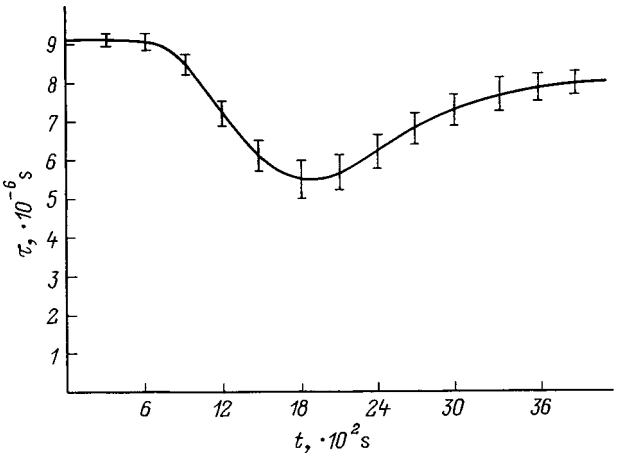

FIG. 2. Lifetime of minority carriers in Ge with Au film as a function of the treatment time in atomic hydrogen.

and subsequent recombination (1.5–2 eV) of hydrogen atoms on the gold surface. The surface resistance R_s of the samples therefore increased and, after the films had been completely removed, became the same as R_s for the initial germanium. Since gold has no volatile compounds with hydrogen, the sputtering mechanism is a purely physical one in which an amount of energy sufficient to detach it from the film is transferred to a surface atom, together with momentum directed from the crystal surface into the gas phase.

In the stage of the hydrogen treatment when the Au films were still continuous, no gold was observed in the germanium. Only after a critical thickness ($d \sim 5 \times 10^{-8}$ m) had been reached, at which discontinuities appeared in the film, did gold atoms begin to penetrate into the subsurface layers of the germanium and then diffuse into the bulk.

Figure 1 shows the gold concentration in germanium after treatment in atomic hydrogen for various times. Curve *1* was obtained for a germanium sample on which a gold film had been deposited and then etched off using a 3:1 HCl–HNO_3 mixture (aqua regia). It can be seen from this curve that no significant diffusion of gold into the germanium takes place when these films are deposited. During treatment in atomic hydrogen for 1800 s a considerable number of atoms penetrated from the film into the bulk of the germanium (curve *2* in Fig. 1). At this stage of the treatment the maximum concentration of gold was observed in the subsurface layers of germanium. During the subsequent treatment process, the germanium layers doped with gold were sputtered and the gold concentration decreased in the subsurface layers. However, the gold concentration in the deeper layers increased as a result of the gold being dispersed in the bulk of the Ge (curve *3* in Fig. 1). Prolonged treatment in H ($t > 10^4$ s) resulted in total sputtering of the Au-doped Ge layers and no gold atoms were observed in the germanium.

The properties of Ge crystals doped with Au were studied by measuring the minority carrier lifetime τ. We know that the addition of gold to germanium accelerates the recombination of electrons and holes and reduces the lifetime τ if Au atoms occupy lattice sites.[6] Figure 2 gives τ as a function of the hydrogen treatment time. It can be seen that initially τ remains constant, since only sputtering of the film has occurred. Under further treatment a decrease in τ is observed as a result of the Au being driven into the Ge. Subsequent treatment almost completely restored τ to its initial value.

Control samples were annealed in order to eliminate any influence on τ from hydrogen atoms adsorbed on the surface or diffusing into the bulk. The samples were annealed at 423 K for 7200 s, during which the values of τ remained constant. This confirms that the change in the lifetime τ for these samples is only caused by the incorporation of gold into the germanium under the influence of hydrogen.

The change in the transmission spectra revealed appreciable absorption at wavelengths in the range $(7.3–10) \times 10^{-6}$ m for Ge samples with Au films treated in atomic hydrogen, whereas the initial germanium and also the Ge samples from which the Au film had been removed by etching in aqua regia were transparent in this part of the infrared (Fig. 3). The absorption peak corresponds to 0.2 eV, which coincides with one of the energy levels created near the bottom of the valence band when a gold atom occupies a Ge crystal lattice site.[6]

It has been confirmed that gold is introduced into germanium under the influence of hydrogen by a two-stage mechanism in which Au atoms are driven into the subsurface layer and then diffuse into the bulk of the Ge. The drive-in of the Au atoms occurs directly under the influence of the released energy. The energy release time for the chemical interaction of hydrogen atoms is of the order of 10^{-10} s (exchange interaction time), so that the process of energy release and dissipation is consistent with a Θ-burst.[6] Since the temperature in the Θ-burst region is fairly high, of the order of the melting point, defect formation and multiphonon scattering mechanisms play an increasing role. The formation of defects and phonon processes promotes the implantation of Au surface atoms in the subsurface layers. An estimate of the

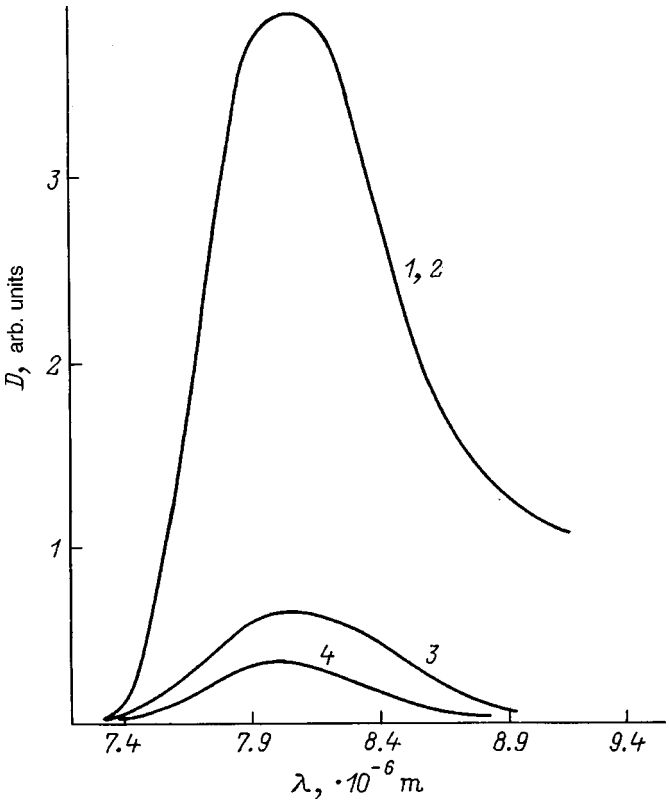

FIG. 3. Transmission spectra of Ge samples with Au film: *1* — initial Ge sample, *2* — Ge sample after removal of Au in aqua regia, *3* — Ge sample with Au film treated in H for 1800 s (Au residue removed in aqua regia), and *4* — Ge sample with Au film.

mean free paths $L_{\parallel}$ and $L_{\perp}$ for longitudinal and transverse phonons, respectively,[7,8] showed that $L_{\parallel} \gg L_{\perp}$ and $L_{\parallel} \sim 10^{-8}$ m, i.e., the depth of the active zone where the hydrogen-stimulated phonon processes take place most effectively is comparable with the depth of the concentration maximum at the middle stage of the hydrogen activation process (curve *2*, Fig. 1).

Gold atoms expelled from the active zone then diffuse into the bulk of the Ge. The diffusion coefficient was calculated using the approximation of diffusion into a semi-infinite crystal with a zero initial concentration across a surface at which the impurity concentration is kept constant. These conditions correspond best to curve *2* in Fig. 1. The diffusion coefficient calculated from this curve is of the order of 3×10^{-17} m^2/s. From a comparison of D_M and D_V, the coefficients of diffusion of Au in Ge by the interstitial and vacancy mechanisms, respectively,[9,10] which at the experimental temperature ($T = 330$ K) have the values $D_M = 8.2 \times 10^{-20}$ m^2/s and $D_V = 1.4 \times 10^{-40}$ m^2/s, we can conclude that the hydrogen-stimulated diffusion of Au atoms into the bulk occurs predominantly by an interstitial mechanism. Under normal conditions this mechanism is realized only at high temperatures ($T > 1000$ K), so that the equilibrium between gold at lattice sites Au_S and in interstitial positions Au_I,

$$Au_I \rightleftarrows Au_S + \Delta Q,$$

is shifted to the left only at high temperatures. Under the influence of atomic hydrogen the concentration Au_I is high even at $T = 330$ K as a result of gold being driven into interstitial positions in the active zone. However, at low, close to room temperatures, when Au_I encounters a vacancy it becomes localized in a lattice site, as is confirmed by measurements of the lifetime τ (Fig. 2) and the transmission spectra (Fig. 3).

CONCLUSIONS

To sum up, the presence of atomic hydrogen effectively stimulates the diffusion of Au into Ge at temperatures close to room temperature. The incorporation of gold into the germanium involves Au atoms being driven into the Ge subsurface layers to a depth of the order of 10^{-8} m, where defect formation mechanisms take place and the released energy is dissipated by multiphonon mechanisms, and this is followed by diffusion of Au into the bulk of the crystals predominantly by an interstitial mechanism. However, a considerable fraction of the gold atoms are localized in Ge lattice sites by interaction with vacancies.

Low-temperature heterodiffusion under the action of hydrogen can be effectively used in electron technology to implant recombination impurities into semiconductor crystals.

[1] V. A. Lavrenko, *Recombination of Hydrogen Atoms at the Surface of Solids* [in Russian], Naukova Dumka, Kiev (1973), 204 pp.
[2] V. M. Matyushin, V. P. Shapovalov, and A. R. Koshman, Int. J. Hydrogen Energy **22**(2), 259 (1997).
[3] V. M. Matyushin, V. I. Lishchenko, and A. N. Gorban', J. Chem. Soc., Faraday Trans. 2 **32**, 1407 (1987).
[4] V. A. Sokolov and A. N. Gorban', *Luminescence and Adsorption* [in Russian], Nauka, Moscow (1969), 187 pp.
[5] L. P. Pavlov, *Methods of Determining the Main Parameters of Semiconductor Materials* [in Russian], Vysshaya Shkola, Moscow (1975), 206 pp.
[6] B. I. Boltaks, *Diffusion in Semiconductors* [in Russian], Fizmatgiz, Moscow (1961), 462 pp.
[7] B. M. Mogilevskiĭ and A. F. Chudnovskiĭ, *Thermal Conductivity of Semiconductors* [in Russian], Nauka, Moscow (1972), 536 pp.
[8] C. I. Classbrenner and G. A. Sluck, Phys. Rev. **134**, 1058 (1964).
[9] *Atomic Diffusion in Semiconductors*, edited by D. Shaw [Plenum Press, London, 1973; Mir, Moscow, 1975, 686 pp.].
[10] W. C. Dunlap, Phys. Rev. **97**, 614 (1955).

Translated by R. M. Durham

Electric fields of dipole structures

M. Yu. Volokobinskiĭ and A. S. Yastrebov

M. A. Bonch-Bruevich St. Petersburg State University of Telecommunications, 191186 St. Petersburg, Russia

(Submitted June 30, 1998)
Zh. Tekh. Fiz. **69**, 77–79 (July 1999)

An analysis is made of the electric field strength of structures formed by point and filamentary dipoles in a dielectric medium. Formulas and the results of calculations are presented.
© *1999 American Institute of Physics.* [S1063-7842(99)01607-4]

Structures and substructures which impart new qualities to instruments and devices are used in electrical engineering and electronics to screen and control carrier fluxes. They are used in electroacoustic transducers in telephone equipment, thyristors, bipolar and field-effect transistors, and in devices to regulate high voltages and high currents.[1–4]

It follows from Refs. 5 and 6, which examined the conversion of energy from a charged particle flux into electromagnetic radiation in a periodic electric field, that substructures are potentially useful in energy converters.

Filamentary and spherical conducting inclusions of small radius are substructure elements at which dipole moments are induced in an electric field. In this context we shall analyze the laws governing the electric field distribution for arrays of spherical inclusions and ''gratings'' of filamentary dipoles.

We shall calculate the electric field E_0 acting on a dipole located at the center of an array and the z component $E_z(z)$ of the electric field on the z axis as a function of the distance from the (x,y) plane, in which the dipoles are distributed at the nodes of a square array, for three cases: 1) the z axis passes through the central node of the array, as shown in Figs. 1a and 1b, $E_z(z)=E_1$; 2) the origin is shifted along the x axis by a half period $d/2$, $E_z(z)=E_2$; 3) the origin is shifted along the x and y axes by a half period, $E_z(z)=E_3$. We postulate that the dipoles are oriented along the z axis and have the same dipole moment μ.

The field E_0 acting on the central dipole is the sum of the fields of all the other dipoles, each elementary dipole of which, being located at a distance $\rho=(x^2+y^2)^{0.5}$ from the center of the array at a point with coordinates $x=md$ and $y=nd$ (where m and n are integers, and d is the period of the array), contributes a field $E_{\mu z}$ given by[7]

$$E_{\mu z}=-\frac{\mu}{4\pi\varepsilon}\frac{1}{\rho^3}=-\lambda(m^2+n^2)^{-1.5}, \tag{1}$$

where $\lambda=\mu/(4\pi\varepsilon d^3)$, and ε is the absolute permittivity.

The field E_0 is given by

$$E_0=\lambda\left\{-2\left[\sum_{m=1}^{M} m^{-3}+\sum_{n=1}^{N} n^{-3}+2\sum_{m=1}^{M}\sum_{n=1}^{N}(m^2+n^2)^{-1.5}\right]\right\}=\lambda R_0, \tag{2}$$

where R_0 is the expression in braces; the upper limits of summation M and N are related to the number of points of the array M_x and N_y along the x and y axes: $M=(M_x-1)/2$, $N=(N_y-1)/2$.

If the number of points along the two axes is the same, $M_x=N_y$, expression (2) simplifies to

$$R_0=-4\left[\sum_{m=1}^{M} m^{-3}+\sum_{m=1}^{M}\sum_{n=1}^{N}(m^2+n^2)^{-1.5}\right]. \tag{3}$$

When constructing expression (3), we assumed that the dipoles located in each of the four quadrants make the same contribution.

The z component of the electric field produced on the z axis at the point $z=kd$ by an elementary dipole with coordinates $x=md$ and $y=nd$ is given by[7]

$$E_z=\lambda\frac{2k^2-(m^2+n^2)}{(k^2+m^2+n^2)^{5/2}}. \tag{4}$$

Summing the fields of all the dipoles, we obtain the z component of the electric field as $E_1=\lambda R_1$, $E_2=\lambda R_2$, or $E_3=\lambda R_3$, which are expressed in terms of the coefficients R_1, R_2, and R_3 given by the following formulas:

$$R_1=\frac{2}{k^3}+4\sum_{m=1}^{M}\frac{2k^2-m^2}{(k^2+m^2)^{5/2}}+4\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{2k^2-(m^2+n^2)}{(k^2+m^2+n^2)^{5/2}}, \tag{5}$$

where the first term gives the contribution of the central dipole, the second gives the contribution of the dipoles located on the x and y axes, and the third gives the contribution of the dipoles loacted in the quadrants;

$$R_2=2\sum_{m=1}^{M}\frac{2k^2-(m-0.5)^2}{[k^2+(m-0.5)^2]^{5/2}}+4\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{2k^2-[(m-0.5)^2+n^2]}{[k^2+(m-0.5)^2+n^2]^{5/2}}, \tag{6}$$

where the first term is due to the dipoles on the axis and the second to the dipoles in the quadrants, and

 © 1999 American Institute of Physics

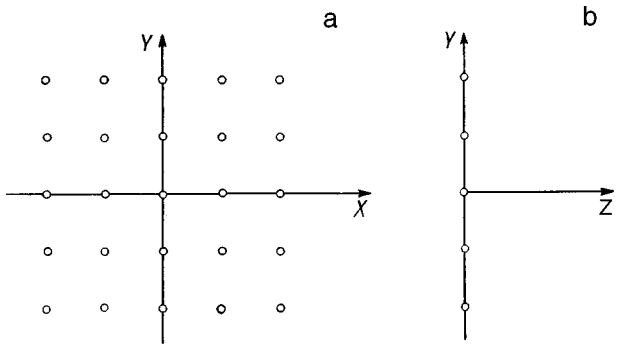

FIG. 1. Planar dipole array: a — from above (in plane), b — in the plane of the array (side view).

$$R_3=4\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{2k^2-[(m-0.5)^2+(n-0.5)^2]}{[k^2+(m-0.5)^2+(n-0.5)^2]^{5/2}} \tag{7}$$

is due to the dipoles in the quadrants.

Expressions (5)–(7) are written assuming that the z axis passes through the center of the rectangular array; in Eq. (5) the number of points along the x and y axes is odd, in Eq. (6) the number of points along the x axis is even and the number along the y axis is odd, and in Eq. (7) the number of points along both axes is even.

To calculate the field of a grating made up of filamentary dipoles extended along the y direction and situated in the (x,y) plane, we can use the expression for the electrical potential of a filamentary dipole[8]

$$\varphi=\frac{\mu}{2\pi\varepsilon}\frac{z}{z^2+x^2}, \tag{8}$$

where $\mu=q\delta$, δ is the distance between the filaments and q is the charge per unit length of a dipole formed by two parallel filaments carrying charges of opposite sign; the z axis passes through the filament and is directed perpendicular to the y axis, from the negative filament to the positive filament.

Expression (8) yields expressions for the components of the electric field produced by a filamentary dipole along the x, y, and z axes,

$$E_x=\frac{\mu}{2\pi\varepsilon}\frac{2xz}{(z^2+x^2)^2},\quad E_y=0,$$

$$E_z=\frac{\mu}{2\pi\varepsilon}\frac{z^2-x^2}{(z^2+x^2)^2}.$$

We introduce the notation $\psi=\mu/(2\pi\varepsilon d^2)$, where d is the distance between the filamentary dipoles, and we express the electric field as $E_4=\psi R_4$, $E_5=\psi R_5$, or $E_6=\psi R_6$. The field E_4 produced at a central dipole filament by all the other filamentary dipoles is given by the coefficient

$$R_4=-2\sum_{m=1}^{M}m^{-2}.$$

For a grating consisting of $2M+1$ filamentary dipoles the field E_5 at a point $z=kd$ on the z axis, which passes through the central filament of the grating, is expressed in terms of R_5, and for a grating of $2M$ filaments the field E_6 on the z axis, which passes through the plane at a point equidistant between the two central dipoles, is expressed in terms of R_6,

$$R_5=k^{-2}+2\sum_{m=1}^{M}\frac{k^2-m^2}{(k^2+m^2)^2},$$

$$R_6=2\sum_{m=1}^{M}\frac{k^2-(m-0.5)^2}{[k^2+(m-0.5)^2]^2}.$$

Results of calculating R_1, R_2, R_3, R_5, and R_6 for short distances from the array ($k\leqslant 5$, $M=N=50$) are presented in Table I. For large distances ($k\geqslant 10$) Table II gives only the values of R_3 and R_6, since here the coefficients R_1, R_2, and R_3 are approximately equal, as are R_5 and R_6.

Let us analyze the data given in Table I. At distances from the array not exceeding one period d, the electric field is highly nonuniform, and the component E_z changes sign as we move along the x and y axes. The total electric field is several times lower than that created by one of the nearest

TABLE I. Values of the coefficients in the expressions for calculating the electric field at small k.

k	R_1	R_2	R_3	R_5	R_6
0.0		−23.24644	−16.40418		−9.82960
0.1	1991.303	−20.45643	−15.37890	96.81367	−8.91620
0.2	241.93217	−14.23478	−12.74208	21.99407	−6.76874
0.3	66.91961	−8.23017	−9.46759	8.37208	−4.47807
0.4	25.14579	−4.16417	−6.42897	3.82603	−2.69656
0.5	10.95109	−1.90894	−4.06327	1.90320	−1.52761
0.6	5.17967	−0.79583	−2.42357	0.99320	−0.83959
0.7	2.58776	−0.28304	−1.37325	0.53731	−0.43383
0.8	1.35425	−0.05637	−0.73576	0.30206	−0.21568
0.9	0.74610	0.04130	−0.36291	0.17875	−0.09724
1.0	0.43942	0.08373	−0.15037	0.11359	−0.03347
1.2	0.20121	0.10735	0.03440	0.06058	0.01901
1.4	0.13668	0.11159	0.09002	0.04555	0.03400
1.6	0.11882	0.11228	0.10634	0.04126	0.03826
1.8	0.11379	0.11237	0.11104	0.04004	0.03946
2.0	0.11235	0.11235	0.11236	0.03968	0.03979
3.0	0.11152	0.11207	0.11262	0.03946	0.03985
5.0	0.11065	0.11119	0.11173	0.03922	0.03960

Note: $k=0$, $R_0=-8.921609$, $R_4=-3.250265$.

TABLE II. Values of the coefficients in the expressions for calculating the electric field at large k.

k	R_3	R_6
10	1.0771×10^{-1}	3.8460×10^{-2}
20	9.3849×10^{-2}	3.4482×10^{-2}
50	4.6189×10^{-2}	2.0000×10^{-2}
100	1.1306×10^{-2}	8.0001×10^{-3}
200	2.2184×10^{-3}	2.3530×10^{-3}
500	1.5686×10^{-4}	3.9604×10^{-4}
1000	1.9900×10^{-5}	9.9751×10^{-5}
2000	2.4969×10^{-6}	2.4984×10^{-5}
3000	7.4033×10^{-7}	1.1108×10^{-5}
4000	3.1240×10^{-7}	6.2490×10^{-6}
5000	1.5997×10^{-7}	3.9996×10^{-6}
1000	1.9999×10^{-8}	9.9997×10^{-7}

dipoles. At distances between approximately $2d$ and $5d$ from the array the electric field is nearly uniform and depends weakly on the coordinates. As z increases between $20d$ and $100d$ the field E_z above an array of point dipoles decreases approximately as $1/z^2$ and that above a grid of filamentary dipoles decreases as $1/z$. At fairly large distances from the array ($k \geqslant 1000$), ten or more times greater than the dimensions of the array, the field above the center of the array is almost the same as that above an elementary dipole multiplied by the number of dipoles forming the array. Thus one can assess the difference in the behavior of the electric field distribution above individual dipoles and arrays and take these into account when designing instruments and devices with substructures, for instance when calculating the forces exerted on electric charges and polarizable particles by a dipole array.

[1] S. M. Sze, *Physics of Semiconductor Devices* [Wiley, New York, 1981; Mir, Moscow, 1984, 456 pp.].
[2] A. Blicher, *Field-effect Bipolar Power Transistor Physics* [Academic Press, New York, 1981; Énergoatomizdat, Leningrad, 1986, 248 pp.].
[3] J. Kaufhold, A. Jastrebov, and M. Volokobinsky, in *Jahresbericht 1995 der Deutsche Telekom FH Dieburg*, Dieburg, Germany (1995), pp. 103–113.
[4] M. Yu. Volokobinskiĭ, I. N. Komarov, V. I. Reshetnikov *et al.*, Peterburg. Zh. Élektron. No. **2**, 19 (1996).
[5] V. A. Buts, Radiotekhnika No. **9**, 9 (1997).
[6] Y.-H. Ohtsuki, *Charged Beam Interaction with Solids* (Taylor and Francis, London, 1983) [Russ. transl., Mir, Moscow, 1985, 280 pp.]
[7] A. R. Von Hippel, *Dielectrics and Waves* [Wiley, New York, 1954; IL, Moscow, 1960, 438 pp.].
[8] V. A. Govorkov, *Electric and Magnetic Fields* [in Russian], GÉI, Moscow–Leningrad (1960), 463 pp.

Translated by R. M. Durham

Surface waves of the potential type at the interface between a metal and an inhomogeneous medium

N. A. Azarenkov, V. K. Galaĭdych, and V. P. Olefir

Kharkov State University, 310077 Kharkov, Ukraine
(Submitted February 14, 1997; resubmitted April 6, 1998)
Zh. Tekh. Fiz. **69**, 80–83 (July 1999)

A study is made of surface waves of the potential type propagating along the interface between a metal and a plasma of nonuniform density, with the thermal motion of the electrons taken into account. Dispersion relation for these waves are derived and solved for a linear plasma density profile. The influence of the nonuniformity of the plasma density on the dispersion properties of the waves is studied. Cases of negative and positive gradients are considered.
© *1999 American Institute of Physics.* [S1063-7842(99)01707-9]

Structures formed by a plasma and a metal surface are currently being studied intensively.[1–3] This interest has been stimulated by the many possible applications of these structures. They are frequently encountered in various plasma technology and fusion facilities (limiters, divertors), in probe diagnostics of plasmas, in studies of the properties of antennas in plasmas, and in semiconductor electronic devices. Surface waves can propagate in these structures.[1] The first theoretical[4–6] and experimental[7] papers on the subject were concerned with surface waves at the interface between the magnetoactive plasma of a semiconductor and a metal. However, the present author and coworkers[8,9] showed that surface waves can also exist at the interface between a metal and a free plasma if the thermal motion of the electrons in the plasma is taken into account. The approximation of a homogeneous plasma was usually considered in all these studies. However, the assumption that a plasma bounding a metal surface is homogeneous cannot always be justified. Under real conditions the plasma is usually inhomogeneous.[1]

Here we consider surface waves at the interface between a metal and an inhomogeneous free plasma with allowance for the thermal motion of the electrons. As we know,[1] unlike the case of volume waves in an inhomogeneous plasma whose frequency is the solution of the local dispersion relation $\omega=\omega(\mathbf{k},x)$, the frequency of the surface waves is an integral function of the density and does not depend on the coordinates.

We shall consider a free (in zero external magnetic field) plasma occupying the half space $x>0$, bounded by a metal surface in the plane $x=0$. The system of equations describing the electromagnetic field of a particular surface wave in the plasma consists of the Maxwell equations and the equations of quasihydrodynamics with allowance for the gaskinetic pressure[1]

$$\operatorname{curl}\mathbf{E}=-\frac{1}{c}\frac{\partial\mathbf{H}}{\partial t},\quad \operatorname{curl}\mathbf{H}=\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}+\frac{4\pi}{c}\sum_{\alpha}\mathbf{j}_{\alpha},$$

$$\frac{\partial\mathbf{V}_{\alpha}}{\partial t}=\frac{e_{\alpha}}{m_{\alpha}}\mathbf{E}-\frac{\nabla p_{\alpha}}{n_{\alpha}m_{\alpha}},\quad \frac{\partial n_{\alpha}}{\partial t}+\operatorname{div}(n_{\alpha}\mathbf{v}_{\alpha})=0,$$

$$\mathbf{j}_{\alpha}=e_{\alpha}n_{\alpha}\mathbf{v}_{\alpha};\quad p_{\alpha}=n_{\alpha}T_{\alpha}, \tag{1}$$

where e_{α}, m_{α}, p_{α}, n_{α}, T_{α}, and $\mathbf{v}_{\alpha}$ are the charge, mass, gaskinetic pressure, density, temperature, and hydrodynamic velocity of particles of species α ($\alpha=e,i$).

The solutions of the system (1) must satisfy boundary conditions at the interface. Since the frequencies of the wave disturbances in the plasma of interest to us are considerably lower than the natural frequencies of the disturbances in the metal, the condition for continuity of the tangential component of the electric field gives the quasistatic boundary condition $\mathbf{E}_{\tau}^{pl}(x=0)=0$ (Ref. 10). Allowance for the thermal motion of the plasma electrons increases the order of the differential equation describing the spatial distribution of the electromagnetic field. Thus an additional boundary condition is required, for which we take the commonly used kinematic condition that the normal component of the hydrodynamic electron velocity is zero,[11] which corresponds to specular reflection of the particles from the plasma boundary.

We shall assume that the unperturbed plasma density varies along the normal to the interface $n_0=n_0(x)$ and that an E wave propagates along the z axis. The dependence of the components of this wave on the coordinates and the time is given by

$$\mathbf{A}(\mathbf{r},t)=\mathbf{A}(x)\exp[i(k_3z-\omega t)],$$

i.e., we are considering surface waves traveling along the interface. We shall analyze undamped waves, neglecting collisions of the plasma electrons and assuming that the metal is a perfect conductor. Without the ion motion, system (1) yields a system of equations for the electric field components E_x, E_z

$$\beta_T^2\frac{d^2E_x}{dx^2}-\frac{\beta_T^2}{n_0(x)}\frac{dn_0(x)}{dx}\frac{dE_x}{dx}-(k_3^2-k^2\varepsilon)E_x=ik_3\left[(1-\beta_T^2)\frac{dE_z}{dx}+\frac{\beta_T^2}{n_0(x)}\frac{dn_0(x)}{dx}\frac{dE_z}{dx}\right],$$

$$\frac{d^2E_z}{dx^2}+(k_3^2-k^2\varepsilon)E_z=ik_3(1-\beta_T^2)\frac{dE_x}{dx}, \tag{2}$$

 © 1999 American Institute of Physics

where $\beta_T = v_{Te}/c$, v_{Te} is the thermal velocity of the plasma electrons, c is the speed of light, and $\varepsilon(x) = 1 - (\Omega_e(x)/\omega)^2$ ($\Omega_e(x) = \sqrt{4\pi e^2 n_0(x)/m_e}$ is the electron plasma frequency).

The magnetic field of the wave is expressed in terms of E_X and E_Z from the following equation

$$ikH_y = ik_3 E_x - \frac{dE_z}{dx}. \tag{3}$$

If the plasma density profile $n_0(x)$ is such that there is no point $x = x_0$ at which the dielectric constant $\varepsilon(x)$ of the inhomogeneous plasma vanishes, the solution of the system (2) may be sought in the form[12]

$$\mathbf{E} = \mathbf{E}^t + \mathbf{E}^l. \tag{4}$$

In this case the single wave process in this system is characterized by the frequency ω and the wave number $\mathbf{k}_3$, but is a superposition of two waves: the electromagnetic polarization wave $\mathbf{E}^t$ and a wave associated with the thermal motion of the electrons $\mathbf{E}^l$ (space charge wave). These waves have different skin depths along the normal to the metal surface. Azarenkov and Kondratenko[8] showed that the inequality $q_1^{-1} \gg q_2^{-1}$ holds for the skin depths of the waves $\mathbf{E}^t$ and $\mathbf{E}^l$ (where q_1^{-1} is the skin depth of $\mathbf{E}^t$ and q_2^{-1} is the skin depth of $\mathbf{E}^l$, respectively)

$$q_1 = \sqrt{k_3^2 - k^2\varepsilon}, \quad q_2 = \sqrt{k_3^2 - k^2\varepsilon/\beta_T^2}. \tag{5}$$

If the nonuniformity of the density is such that

$$q \gg_1^{-1}, \quad a \gg q_2^{-1}, \tag{6}$$

where $a = n_0 |dn_0/dx|^{-1}$ is the inhomogeneity parameter, the inhomogeneity can be considered to be unimportant for a space charge wave but important for the polarization component of the surface wave. The influence of the electron thermal motion on the polarization component $\mathbf{E}^t$ can be neglected. Thus, we can set $\beta_T = 0$ to determine the field $\mathbf{E}^t$. We shall consider waves for which $k_3^2 \gg |k^2\varepsilon|$. If this condition is satisfied, these surface waves will be potential waves.[8] The system of equations describing the components of the electric field of the transverse t and longitudinal l components then has the form:

$$\frac{d^2E_z^t}{dx^2} + \frac{1}{\varepsilon(x)}\frac{d\varepsilon(x)}{dx}\frac{dE_z^t}{dx} - k_3^2 E_z^t = 0, \quad E_x^t = \frac{i}{k_3}\frac{dE_z^t}{dx}, \tag{7}$$

$$\frac{d^2E_z^l}{dx^2} - q_2^2 E_z^l = 0, \quad ik_3 E_x^l = dE_z^l/dx. \tag{8}$$

In general, Eq. (7) for E_z^t cannot be solved for an arbitrary law $\varepsilon(x)$. However, exact solutions of Eq. (7) can be constructed for some particular cases.

Let us analyze a linear plasma density profile with a negative gradient, $n_0(x) = n_0(1 - x/a)$. For this law of variation the solutions of Eq. (7) can be expressed in terms of a zeroth-order modified Bessel function:[13]

$$E_z^t = B \cdot I_0(k_3\xi), \quad \xi = a\frac{\varepsilon(0)}{\varepsilon(0) - 1} - x. \tag{9}$$

The other components of this wave field are expressed in terms of a first-order modified Bessel function:

$$E_x^t = -iB \cdot I_1(k_3\xi), \quad H_y^t = -i\frac{k\varepsilon}{k_3} B \cdot I_1(k_3\xi). \tag{10}$$

We obtain the dispersion relation for the surface wave from the condition that the tangential component of the resultant electric field $\mathbf{E}$ at the metal–plasma interface is zero and from the kinematic boundary condition. This equation reduces to the form:

$$1 - \frac{\Omega_e^2(0)}{\omega^2}\frac{k_3}{q_2}\frac{I_1(k_3\xi(0))}{I_0(k_3\xi(0))} = 0. \tag{11}$$

If the density nonuniformity is weak, $k_3 a \gg 1$, Eq. (10) can be simplified, and we can obtain the following solution:

$$k_3 = k_{30}\left\{1 + \frac{[1 - \varepsilon(0)]^3}{2\varepsilon^2(0)[2 - \varepsilon(0)]k_{30}a}\right\}, \tag{12}$$

where $k_{30} = \omega^2/v_{Te}\Omega_e(0)$ is the solution of the dispersion relation for a surface wave at the interface between a homogeneous plasma and a metal. For the low-frequency range ($\omega^2 \ll \Omega_e^2(0)$) we have

$$k_3 = k_{30}\left[1 + \frac{1}{2k_{30}a}\right]. \tag{13}$$

Let us now analyze a linear plasma density profile with a positive gradient, $n_0(x) = n_0(1 + x/a)$. Using another change of variable,

$$\zeta = a\frac{\varepsilon(0)}{\varepsilon(0) - 1} + x \tag{14}$$

we obtain a solution of Eq. (7) in the form of a zeroth-order Macdonald function[13]

$$E_z^t = C \cdot K_0(k_3\zeta). \tag{15}$$

The other wave components are expressed in terms of a first-order Macdonald function

$$E_x^t = -iC \cdot K_1(k_3\zeta), \quad H_y^t = -i\frac{k\varepsilon}{k_3} C \cdot K_1(k_3\zeta). \tag{16}$$

In this case we obtain the dispersion relation for the surface wave as

$$1 - \frac{\Omega_e^2(0)}{\omega^2}\frac{k_3}{q_2}\frac{K_1(k_3\zeta(0))}{K_0(k_3\zeta(0))} = 0. \tag{17}$$

This equation can also be simplified for cases of weak inhomogeneity, $k_3 a \gg 1$. As a result, we obtain the following equation for the wave vector of the surface wave:

$$k_3 = k_{30}\left\{1 - \frac{[1 - \varepsilon(0)]^3}{2\varepsilon^2(0)[2 - \varepsilon(0)]k_{30}a}\right\}. \tag{18}$$

For the low-frequency range ($\omega^2 \ll \Omega_e^2(0)$) we obtain

$$k_3 = k_{30}\left[1 - \frac{1}{2k_{30}a}\right]. \tag{19}$$

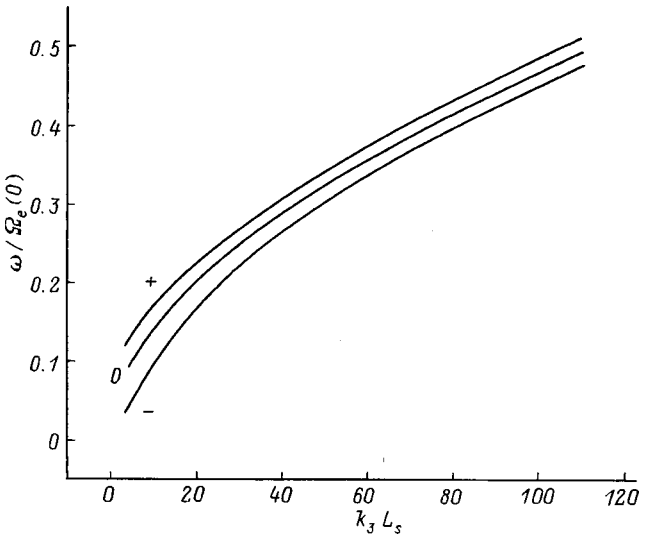

FIG. 1. Frequency of surface wave as a function of wave number for various plasma density profiles: + — increasing density inside plasma, − — decreasing density, 0 — homogeneous plasma, $a/L_s=0.5$, and $v_{\mathrm{Te}}/c=0.002$.

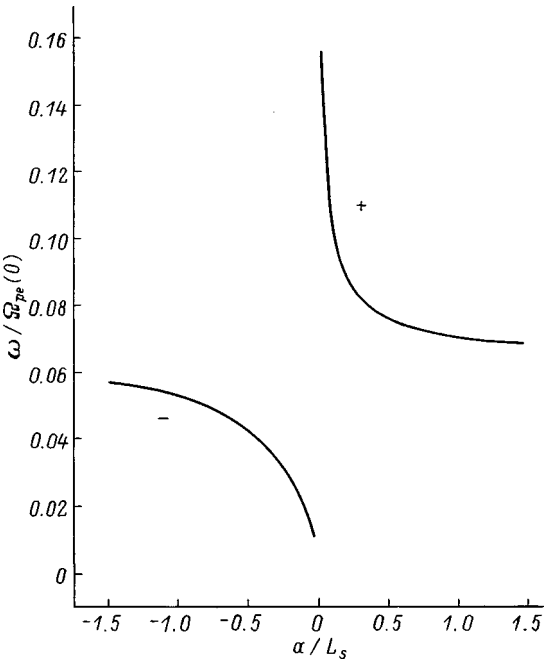

FIG. 3. Wave frequency versus inhomogeneity parameter: + and − as in Fig. 1, $k_3/L_s=2$, $v_{\mathrm{Te}}/c=0.002$.

For arbitrary density nonuniformity, solutions of Eqs. (11) and (17) can be obtained numerically, and the results of these calculations are plotted in Fig. 1. In this case, the skin depth $L_s=c/\Omega_e(0)$ was taken to be the natural unit of length for normalization. It is easy to see that the frequency of a surface wave for the negative-gradient profile $n_0(x)=n_0(1-x/a)$ is always lower and the frequency for the positive-gradient profile $n_0(x)=n_0(1+x/a)$ is always higher than the corresponding frequency for a homogeneous plasma. Figure 2 shows the change in the phase velocity of the wave as a function of its frequency for various density gradients. This behavior of the curves is consistent with the wave dispersion in a homogeneous plasma, where an increase in the plasma density causes a decrease in the wave number and an increase in the phase velocity of the wave. Figure 3 gives the wave frequency as a function of the inhomogeneity parameter a. Low values $a\to 0$ correspond to abrupt inhomogeneity, and for these cases we need to check that the initial assumptions are satisfied. High values $a\to\pm\infty$ correspond to a homogeneous medium, so that the curves have a common limit which is equal to the surface wave frequency in a homogeneous plasma.

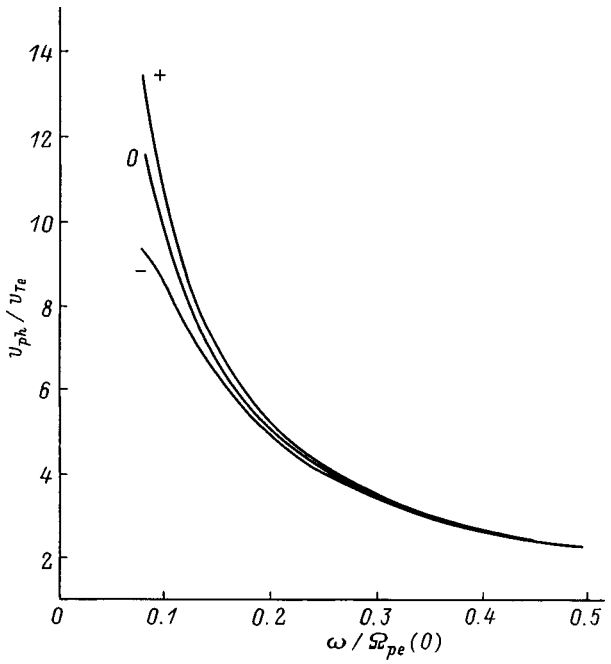

FIG. 2. Phase velocity of surface wave as a function of frequency. The notation is the same as in Fig. 1: $k_3/L_s=2$, $v_{\mathrm{Te}}/c=0.002$.

To conclude, we have established that surface waves of the potential type can propagate along the interface between a metal and an inhomogeneous plasma when the thermal motion of the plasma electrons is taken into account. Dispersion relations were obtained and solved for these waves. Cases of negative and positive plasma density gradients were considered.

The dispersion curves for the surface waves at a metal–inhomogeneous plasma interface do not differ qualitatively from those for an interface between a metal and a homogeneous plasma. The dependence of the natural frequency of the surface wave on its wave vector is shifted to the left (right) along the abscissa for a plasma with a positive (negative) density gradient.

These results show that the nonuniformity of the plasma density influences the surface wave characteristics in an integral manner.

[1] A. N. Kondratenko, *Surface and Volume Waves in a Bounded Plasma* [in Russian], Énergoatomizdat, Moscow (1985), 232 pp.
[2] N. N. Beletskiĭ, V. M. Svetlichnyĭ, D. D. Khalameĭda, and V. M. Yakovenko, *Electromagnetic Phenomena in the Microwave Range in Inhomogeneous Structures* [in Russian], Naukova Dumka, Kiev (1991), 216 pp.

[3] N. A. Azarenkov, A. N. Kondratenko, and K. N. Ostrikov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **36**, 335 (1993).
[4] S. R. Seshadri, IRE Trans. Microwave Theory Tech. **MTT-10**, 573 (1962).
[5] R. J. Hirota, J. Phys. Soc. Jpn. **19**, 1130 (1964).
[6] R. J. Hirota and K. Suzuki, J. Phys. Soc. Jpn. **21**, 1112 (1966).
[7] M. J. Toda, J. Phys. Soc. Jpn. **19**, 1126 (1964).
[8] N. A. Azarenkov and A. N. Kondratenko, Ukr. Fiz. Zh. **30**, 718 (1985).
[9] N. A. Azarenkov, Zh. Tekh. Fiz. **57**, 1165 (1987) [Sov. Phys. Tech. Phys. **32**, 681 (1987)].
[10] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, 2nd ed. with L. P. Pitaevskii [Pergamon Press, Oxford, 1984; Nauka, Moscow, 1982, 620 pp.].
[11] A. N. Kondratenko, *Plasma Waveguides* [in Russian], Atomizdat, Moscow (1976), 232 pp.
[12] A. D. Piliya and V. I. Fedorov, Zh. Éksp. Teor. Fiz. **57**, 1198 (1969) [Sov. Phys. JETP **30**, 653 (1969)].
[13] E. Kamke, *Gewöhnliche Differentialgleichungen* [Acad. Verlag, Leipzig, 1959; 5th ed., Nauka, Moscow, 1976, 576 pp.].

Translated by R. M. Durham

Nonlinear theory of beam excitation of azimuthal surface modes in plasma-filled cylindrical metal waveguides

V. O. Girka and A. M. Kondratenko

Kharkov State University, 310108 Kharkov, Ukraine

A. E. Sporov

Scientific-Technological Center of Electrophysics, National Academy of Sciences of Ukraine, 310002 Kharkov, Ukraine

(Submitted February 23, 1998)
Zh. Tekh. Fiz. **69**, 84–88 (July 1999)

It is shown that electromagnetic surface waves propagating along the azimuthal angle can be excited efficiently by an annular electron beam in a cylindrical metal waveguide partially filled with a magnetoactive plasma. A self-consistent system of differential equations is obtained to describe the nonlinear interaction between the beam particles and an azimuthal surface wave in the single-mode regime. This system of equations is analyzed numerically and the influence of the parameters of this waveguide structure on the development of the resonant beam instability is determined. © *1999 American Institute of Physics.* [S1063-7842(99)01807-3]

INTRODUCTION

Studies of plasma–beam interaction are attracting continuing interest primarily because of the major practical value of the expected results, which are finding extensive fields of application, ranging from beam heating of plasmas in controlled fusion devices and for conducting geophysical experiments in space, to solving problems in plasma electronics. The present theoretical study is devoted to solving a problem relevant to one line of plasma electronics research,[1] which involves studying the interaction between charged particle beams and the natural modes of plasma waveguides. Studies of electromagnetic wave generation processes governed by the parameters of the beam and the plasma waveguide are aimed at producing rf amplifiers or oscillators with the highest possible efficiency, permitting continuous frequency tuning over a wide range, and having comparatively small dimensions, the highest possible radiation power, and so on. Obviously, all these requirements cannot be satisfied simultaneously by a single device. The present study is an attempt to construct a nonlinear theory for a plasma electronics device which could satisfy two criteria to a certain extent, i.e., it would offer continuous tuning of the radiation frequency in a fairly compact device.

At present, the conditions for the onset of beam instabilities and nonlinear interaction between charged particle beams and plasma in a strong external magnetic field and in a finite magnetic field have been studied in fairly great detail. The influence of the spectrum of generated waves on the nature of the beam–plasma interaction has also been analyzed without taking the influence of the plasma boundaries into account.[2–4] The bounded dimensions of the plasma not only lead to changes in the dynamics of excitation of volume waves but also creates conditions for the excitation of surface waves.[5] Surface waves have various distinguishing features compared with volume waves: the energy of the surface waves is transferred along the plasma boundary and can be extracted comparatively easily from the interaction zone.[1]

Here we study the possibility of exciting a surface mode of extraordinary polarization which is a natural mode for a cylindrical metal waveguide partly filled with a cold magnetoactive plasma. The wave propagates along the azimuthal angle at right angles to an external static axial magnetic field and is called[6] an azimuthal surface wave (ASW). However, this wave is only a surface wave in the region of the plasma cylinder; in the insulator separating the metal wall of the waveguide from the plasma, this electromagnetic disturbance propagates as a volume wave. We note that such plasma-filled waveguide structures are currently being studied fairly intensively (see Refs. 7 and 8) and the literature cited therein), so that this choice of subject is highly topical. A linear theory of the beam excitation of ASWs was put forward in Ref. 9, and the case of dissipative ASW instability was studied in Ref. 10.

We propose to excite ASWs using a cylindrical metal waveguide of radius R_2 containing a coaxial plasma column of radius R_1, where $R_2-R_1 \ll R_1$. In the space $R_1<r<R_2$ an annular electron beam rotates about the plasma column (which may be a semiconductor plasma or a gas plasma contained in a thin-walled dielectric tube). The beam density n_b is substantially lower than the plasma density n_p. An external static magnetic field $\mathbf{B_0}$ is oriented along the z axis. We assume that the electrical conductivity of the metal waveguide wall is sufficiently high that we can use the boundary condition for the tangential component of the ASW electric field at the metal surface in the form $E_\tau(R_2)=0$.

We used a hydrodynamic plasma model, Maxwell's equations, and the equations of motion for the beam particles in the gap $R_2>r>R_1$ between the plasma insert and the metal wall of the waveguide to obtain a system of differential equations for describing the nonlinear stage of the beam excitation of ASWs. Bearing in mind the inequality for the

 © 1999 American Institute of Physics

beam particle and plasma densities $n_n \ll n_p$, we neglected the influence of the beam on the ASW dispersion properties and the influence of the self-field of the beam on the electromagnetic field in which it propagates.

DERIVATION OF THE SYSTEM OF EQUATIONS

In the cold plasma approximation, Maxwell's equations can be divided into two independent subsystems, one of which describes the field of an ASW of extraordinary polarization. Assuming that the wave field depends on the time t and the azimuthal angle φ as $E,H \propto \exp(im\varphi - i\omega t)$ and that the space is uniform along the z axis, we obtain a second-order differential equation for the magnetic component H_z of the ASW field and expressions linking the ASW field with H_z in the region of the plasma cylinder $r<R_1$:

$$\frac{\partial^2 H_z}{\partial \xi^2}+\frac{1}{\xi}\frac{\partial H_z}{\partial \xi}-\left(1+\frac{m^2}{\xi^2}\right)H_z=0, \tag{1a}$$

$$E_r=\frac{\varepsilon_2}{k\varepsilon_1\psi^2}\frac{\partial H_z}{\partial r}+\frac{mH_z}{kr\psi^2},$$

$$E_\varphi=\frac{im\varepsilon_2 H_z}{kr\varepsilon_1\psi_2}+\frac{i}{k\psi^2}\frac{\partial H_z}{\partial r}, \tag{1b}$$

where $\xi=kr\psi$, $k=\omega c^{-1}$, c is the speed of light, $\psi^2=(\varepsilon_2^2-\varepsilon_1^2)\varepsilon_1^{-1}$, ε_1 and ε_2 are the components of the permittivity tensor of a cold magnetoactive plasma (see Ref. 5, for example).

The fields of the ASW in the region occupied by the beam are described by the following system of differential equations:

$$\frac{\partial^2 H_z}{\partial \zeta^2}+\frac{1}{\zeta}\frac{\partial H_z}{\partial \zeta}-\left(1-\frac{m^2}{\zeta^2}\right)H_z=F_b, \tag{2a}$$

$$E_r=\frac{4\pi}{i\omega}j_r-\frac{mH_z}{\zeta},$$

$$E_\varphi=\frac{4\pi}{i\omega}j_\varphi-i\frac{\partial H_z}{\partial \zeta}, \tag{2b}$$

where

$$F_b=-\frac{4\pi}{\omega\zeta}\left[\frac{\partial}{\partial\zeta}(j_\varphi\zeta)-imj_r\right], \quad \zeta=kr,$$

$$j_r=-|e|\sum_{i=1}^{N}\delta(r-r_i)\,\delta(\varphi-\varphi_i)\frac{\partial r}{\partial t},$$

$$j_\varphi=-|e|\sum_{i=1}^{N} r\,\delta(r-r_i)\,\delta(\varphi-\varphi_i)\frac{\partial \varphi}{\partial t},$$

and $\delta(x-x_0)$ is a delta function.

The solutions of the system (1a), (1b) for the ASW field in the plasma region are a modified Bessel function $I_m(\xi)$ for the H_z component, and linear combinations of $I_m(\xi)$ and its derivative with respect to the argument $I'_m(\xi)$ for the components E_r and E_φ. We used the boundary condition that the ASW field is bounded on the waveguide axis $(r=0)$. The system of equations (2a) and (2b) was solved by variation of a parameter. The components of the ASW field in the region $R_1<r<R_2$ are expressed in terms of the first-order Bessel functions $J_m(\zeta)$, the Neumann functions $N_m(\zeta)$, their derivatives with respect to the arguments, and the components of the beam current density j_r and j_φ.

We shall use the following boundary conditions to derive equations describing the behavior of the amplitude of the envelope and the wave phase. The tangential electric field of the ASW vanishes at the surface of the waveguide metal wall:

$$E_\varphi(R_2)=0. \tag{3}$$

The field E_φ is continuous at the plasma boundary $r=R_1$:

$$\{E_\varphi(R_1)\}=0. \tag{4}$$

The ASW magnetic field H_z is also continuous at the plasma boundary $r=R_1$

$$\{H_z(R_1)\}=0. \tag{5}$$

We also assumed that there are no currents at the surface of the metal wall of the waveguide and the plasma surface, $j_\varphi(R_1)=j_\varphi(R_2)=0$.

Using standard procedures for averaging and separating out the slow time (see, for instance, Ref. 1), we find equations for the amplitude of the envelope and the wave phase for the case where natural waveguide modes are excited and the dissipative processes in the plasma are neglected:

$$\frac{\partial A}{\partial t}=-\frac{\alpha D_p}{NzPL}\sum_{i=1}^{N}\left[\frac{m}{w}\frac{\partial R_i}{\partial t}L_1(\zeta_i)\sin(m\varphi_i+\Theta-\omega t)\right.$$
$$\left.+R_i^2\frac{\partial\varphi_i}{\partial t}L_2(\zeta_i)\cos(m\varphi_i+\Theta-\omega t)\right],$$

$$\frac{\partial\Theta}{\partial t}=\frac{\alpha D_p}{NzPLA}\sum_{i=1}^{N}\left[R_i^2\frac{\partial\varphi_i}{\partial t}L_2(\zeta_i)\sin(m\varphi_i+\Theta-\omega t)\right.$$
$$\left.-\frac{m}{w}\frac{\partial R_i}{\partial t}L_1(\zeta_i)\cos(m\varphi_i+\Theta-\omega t)\right], \tag{6}$$

where $A=E_y B_0^{-1}$ is the dimensionless envelope amplitude, Θ is the phase, $\alpha=n_b n_p^{-1}$, $z=|\omega_e|\Omega_e^{-1}$, $w=\omega\Omega_e^{-1}$, $R_i=r_i\Omega_e c^{-1}$, ω_e and Ω_e are the electron cyclotron and plasma frequencies, respectively, $L=J_m(\zeta_1)N'_m(\zeta_2)-J'_m(\zeta_2)N_m(\zeta_1)$, $\zeta_1=kR_1$, $\zeta_2=kR_2$, $L_1(\zeta_i)=J_m(\zeta_i)N'_m\times(\zeta_2)-J'_m(\zeta_2)N_m(\zeta_i)$, $L_2(\zeta_i)=J'_m(\zeta_i)N'_m(\zeta_2)-J'_m(\zeta_2)N'_m\times(\zeta_i)$,

$$D_p=\frac{I'_m(\psi\zeta_1)}{\psi I_m(\psi\zeta_1)}+\frac{m\varepsilon_2}{\varepsilon_1\psi^2\zeta_1}, \quad P=\frac{d}{dw}\left[D_p+\frac{L_2(\zeta_1)}{L}\right].$$

The equation of motion for the beam electrons can be conveniently written in terms of the particle momentum $\mathbf{p}=\gamma m_e\mathbf{V}$ (γ is the relativistic factor)

$$\frac{d\mathbf{p}}{dt}=e\mathbf{E}+\frac{e}{c}[\mathbf{V}\times(\mathbf{H}+\mathbf{B}_0)]. \tag{7}$$

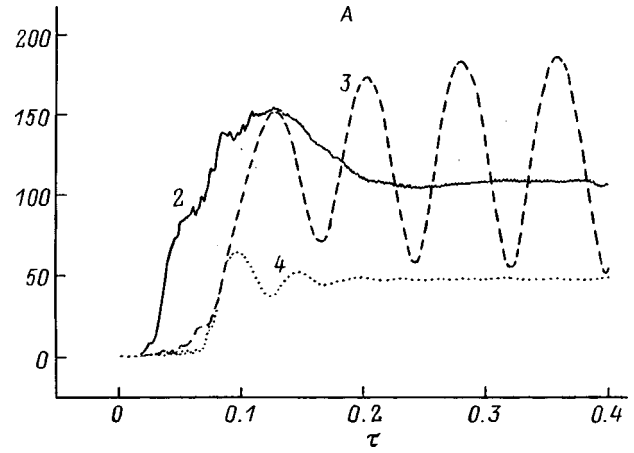

FIG. 1. Evolution of the dimensionless amplitude A as a function of the dimensionless time τ for $|\omega_e|=0.098572\,\Omega_e$, $R_1=5.0$, $R_2=5.5$.

Substituting into Eq. (7) the ASW field for the region $R_1<r<R_2$, we obtain the following system of equations for the ith particle:

$$\frac{\partial R_i}{\partial t}=|\omega_e|\frac{\nu_i}{z\gamma_i},$$

$$\frac{\partial \varphi_i}{\partial t}=|\omega_e|\frac{u_i}{z\gamma_i R_i},$$

$$\frac{\partial \nu_i}{\partial t}=-|\omega_e|\frac{u_i}{\gamma_i}\left(1-\frac{u_i}{zR_i}\right)+AR_1\left(\frac{m}{R_1}-w\frac{u_i}{\gamma_i}\right)\times\left(\frac{1}{\zeta_1 D_p}-R_\alpha\right)\sin\,(m\varphi_i+\Theta-\omega t),$$

$$\frac{\partial u_i}{\partial t}=|\omega_e|\frac{\nu_i}{\gamma_i}\left(1-\frac{u_i}{zR_i}\right)+\frac{AwR_1}{\zeta_1}\left[R_\alpha-1+\frac{R_\alpha}{D_p}\times\left(\frac{m^2}{\zeta_1}-\zeta_1\right)\right]\cos\,(m\varphi_1+\Theta-\omega t)+\frac{AwR_1\nu_i}{\gamma_i}\left(\frac{1}{\zeta_1 D_p}-R_\alpha\right)\sin\,(m\varphi_i+\Theta-\omega t),\tag{8}$$

where $R_\alpha=r_iR_1^{-1}-1$, $\nu=p_rm_e^{-1}c^{-1}$, $u=p_\varphi m_e^{-1}c^{-1}$, and ν and u are the dimensionless momenta of the beam particles (radial and azimuthal, respectively).

In order to derive the last two equations in the system (8), we used the initial assumption that the region occupied by the beam is relatively narrow, i.e., $R_2-R_1\ll R_1$. This allowed us to substantially simplify the expressions contained in the system (8), by using asymptotic expansions of cylindrical Bessel functions and their Wronskians (see Ref. 11), and consequently significantly reduced the computing time. The time required for direct calculations of cylindrical functions increases drastically, creating additional difficulties. A joint solution of the systems of equations (6) and (8) was obtained numerically.

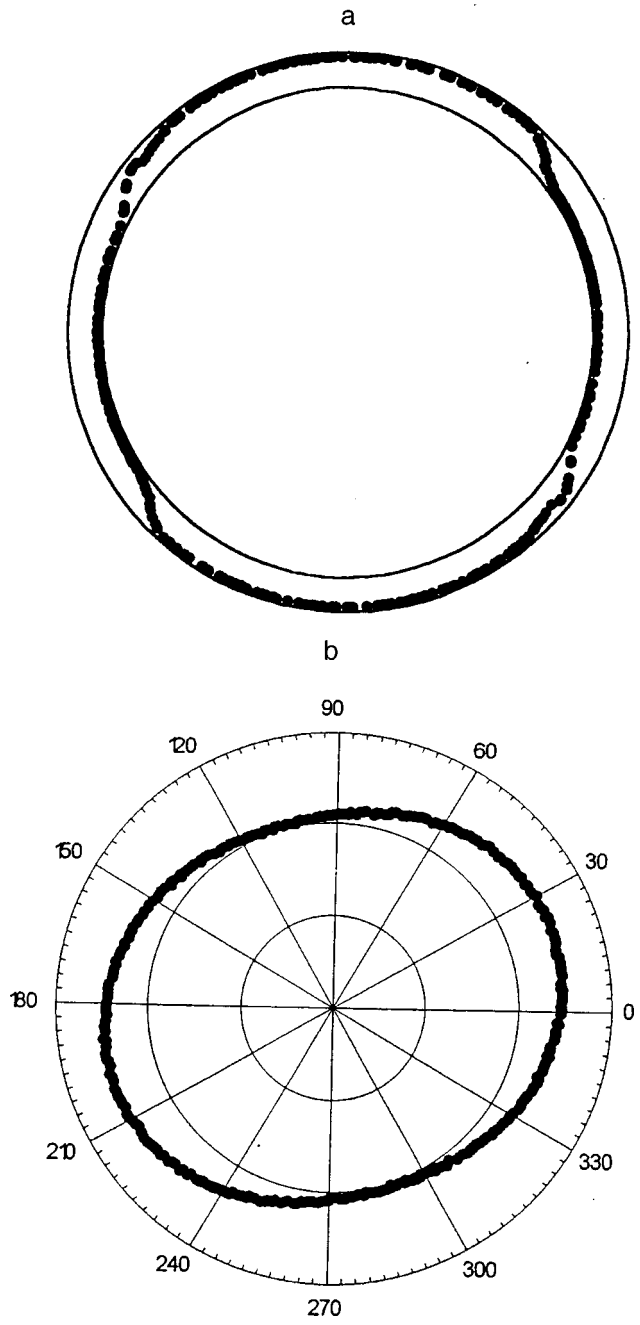

FIG. 2. Distribution of beam particles in coordinate (a) and phase (b) space for the same values of the parameters as in Fig. 1 and $\tau=0.02$.

RESULT OF THE NUMERICAL SIMULATION

A fourth-order Runge–Kutta method was used to obtain a numerical solution of the joint system (6) and (8). The number of macroparticles used to simulate the electron beam was $N=450$. A specular reflection model (the particles did not disappear as a result of the interaction) was used for the interaction between the beam particles and the plasma boundary or the metal waveguide wall. The results of a numerical investigation of the development of the resonant beam instability of an ASW are plotted in Figs. 1–4. For the numerical simulation we used the following values of the waveguide and beam parameters: initial wave amplitude $A=10^{-3}$, phase $\Theta=0$, radial momentum of the beam par-

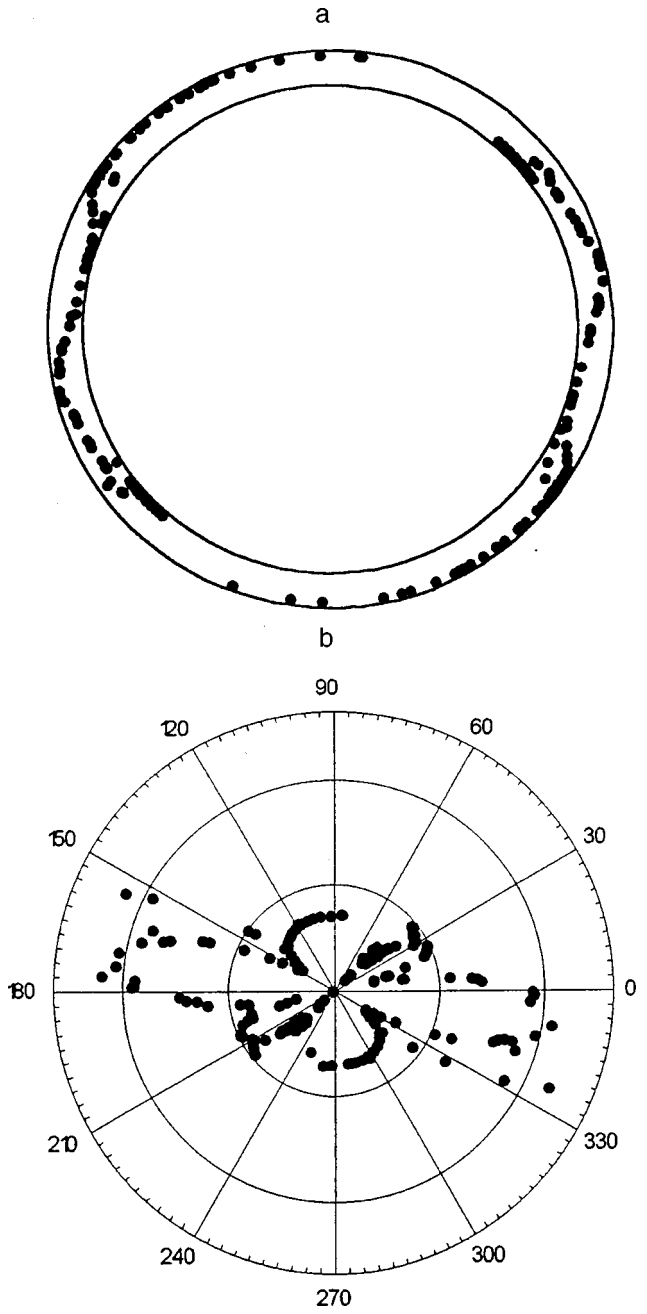

FIG. 3. a,b — As Fig. 2a,b, respectively, for $\tau=0.6$.

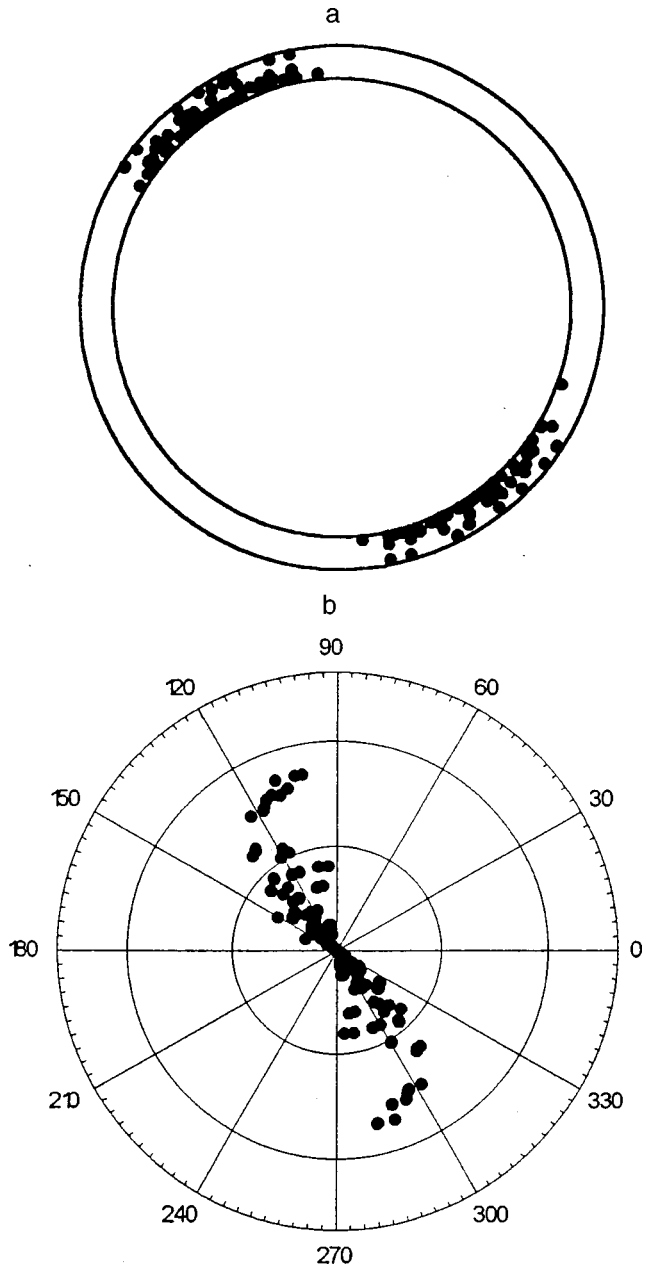

FIG. 4. a,b — As Fig. 2a,b, respectively, for $\tau=0.03$.

ticles $\nu_i=0$, angular momentum of the particles $u_i=zR_i$ ($\pm 2\%$), and ratio of the electron beam density to the plasma density $\alpha=10^{-2}$. The initial distribution of the beam particles over the angle φ was approximately uniform with a small random deviation ($\Delta\varphi=\pm 1\%$), and the initial radial distribution of the beam particles was assumed to be random in the region $R_1<r<R_2$. The other waveguide parameters were: $R_2-R_1=0.1R_1$, $|\omega_e|=0.098572\Omega_e$. In Fig. 1 the labels *2*, *3*, and *4* are the azimuthal mode indices m of the ASW.

It was shown in Ref. 9 that there is a region of effective wave number $k_{\text{eff}}=mcR_1^{-1}\Omega_e$ where the highest values of the ASW beam instability growth rate are localized. This corresponds to $k_{\text{eff}}\approx 0.4$, so that for the computations the value of R_1 was selected so that $k_{\text{eff}}=0.4$. A study of the influence of the choice of k_{eff} on the development of beam instability confirms the results obtained in Ref. 9. The ASW amplitude did in fact increase more slowly if k_{eff} had values of 0.3, 0.2 or 0.5, 0.6. A decrease in k_{eff} caused a greater slowing of the growth of the ASW amplitude than did an increase, a finding which is also consistent with the results of Ref. 9. A change in the sign of the azimuthal mode index, which determines the direction of propagation of the ASW, leads to cutoff of the instability. A decrease in the ratio $|\omega_e|\Omega_e^{-1}$ and an increase in the parameter α lead to a reduction in the time interval over which the ASW amplitude increases from the initial value to the maximum.

Figures 2–4 show the development of the beam particle distribution in coordinate space (angular and radial coordinates) and phase (angular coordinate and angular momentum) space when an ASW is excited with the azimuthal

mode index $m=2$. We selected the most characteristic moments in dimensionless time τ, when the initial ($\tau=0$), almost uniform beam particle distribution (in both coordinate and phase space this resembled a continuous undeformed ring) begins to become distorted ($\tau=0.02$), breaks up into bunches ($\tau=0.06$), and is converted into two (according to the mode index $m=2$) incipient bunches ($\tau=0.3$). Figures 2a, 3a, and 4a show the distribution of the beam particles in coordinate space for an ASW with the azimuthal mode index $m=2$ at dimensionless times $\tau=0.02$, 0.06, and 0.3. The change in the distribution of the beam particles in phase space as a function of time can be seen from Figs. 2b, 3b, and 4b. A study of these figures reveals that as the resonant beam instability of the ASW develops, the particle distribution over the angle φ in coordinate space evolves from an initial approximately uniform distribution over the azimuthal angle to form two particle bunches for $m=2$ (the number of these bunches is equal to the azimuthal mode index of the ASW). In phase space, as a result of being trapped by the wave field, the beam particles form patterns similar to the spokes in a wheel (the number of these spokes being the same as the mode index m).

CONCLUSIONS

We have investigated the excitation of an ASW having a frequency in the range $|\omega_e|<\omega<\sqrt{\Omega_e^2+\omega_e^2/4}-|\omega_e|/2$, by an annular electron beam rotating about a plasma column which partially fills a cylindrical metal waveguide. This waveguide structure is placed in an axial static magnetic field. The resonant beam instability of the ASW was investigated in the single-mode approximation. Nonlinear interaction between the electron beam and the natural waveguide modes causes the beam to break up into bunches which are gradually trapped in the potential wells of the wave. After this trapping, the instability reaches a stage of nonlinear saturation. This saturation state is characterized by a slight variation of the amplitude of the envelope about a certain equilibrium value. The azimuthal mode index has the strongest influence on the character of the development of the ASW beam instability.

This work was partially supported by the Ministry of Science and Technology of the Ukraine, Grant No. WTZ UKR 010-97.

[1] A. N. Kondratenko and V. M. Kuklin, *Principles of Plasma Electronics* [in Russian], Énergoatomizdat, Moscow (1988), 320 pp.
[2] A. A. Rukhadze, L. S. Bogdankevich, S. E. Rosinskiĭ, and V. G. Rukhlin, *Physics of High-Current Relativistic Electron Beams* [in Russian], Atomizdat, Moscow (1980), 167 pp.
[3] R. B. Miller, *Introduction to the Physics of Intense Charged-Particle Beams* [Plenum Press, New York, 1982; Mir, Moscow, 1994, 432 pp.].
[4] M. V. Kuzelev and A. A. Rukhadze, *Electrodynamics of Dense Electron Beams in Plasma* [in Russian], Nauka, Moscow (1990), 432 pp.
[5] A. N. Kondratenko, *Surface and Volume Waves in a Bounded Plasma* [in Russian], Énergoatomizdat, Moscow (1985), 208 pp.
[6] V. A. Girka, I. A. Girka, A. N. Kondratenko *et al.*, Radiotekh. Elektron. **33**, 1031 (1988).
[7] I. N. Onischenko, V. A. Balakirev, A. M. Korostelev *et al.*, in *Proceedings of the 11th International Conference on High Power Particle Beams*, Prague, 1996, Contributed Papers. Vol. 1. pp. 426–429.
[8] M. V. Kuzelev, O. T. Loza, A. A. Rukhadze *et al.*, *ibid.*, pp. 1–55.
[9] V. A. Girka, I. A. Girka, V. P. Olefir *et al.*, Pis'ma Zh. Tekh. Fiz **17**(1), 87 (1991) [Sov. Tech. Phys. Lett. **17**(1), 35 (1991).
[10] V. A. Girka, I. A. Girka, and V. I. Tkachenko, Zh. Tekh. Fiz. **66**(4), 114 (1996) [Tech. Phys. **41**, 357 (1996)].
[11] D. S. Kuznetsov, *Special Functions* [in Russian], Vysshaya Shkola, Moscow (1968), 422 pp.

Translated by R. M. Durham

Theory of absorption of electromagnetic radiation by highly inhomogeneous two-component systems

S. O. Gladkov

N. N. Semenov Institute of Chemical Physics, Russian Academy of Sciences, 117977 Moscow, Russia
(Submitted April 28, 1997; resubmitted January 18, 1999)
Zh. Tekh. Fiz. **69**, 89–94 (July 1999)

Calculations are made of the permittivity of a composite formed by a mixture of a dielectric matrix and macroscopic magnetic inclusions. It is shown that the dielectric loss tangent in this material is a complex function of their concentration x, and this function is determined. The dependence of the absorption maximum on x is determined and it is shown that the maximum is only observed in the presence of foreign inclusions. Conditions for the existence of an absorption maximum are obtained for the two most interesting physical cases. A method of calculating the components of the depolarization tensor is proposed for a thin-film composite.
© *1999 American Institute of Physics.* [S1063-7842(99)01907-8]

Studies of the properties of complex composite compounds have recently attracted major interest for various reasons. The first and main reason from the physical point of view is that these compounds exhibit anomalous properties compared with ''normal'' materials of homogeneous composition. In particular, some physical parameters, such as the thermal conductivity $\varkappa$ (see Refs. 1–4, for instance), because of its strong dependence on the concentration x of the disperse impurity phase, have somewhat unusual temperature dependences which are only manifested when $\varkappa$ passes through an extremum. Another reason is that these compounds can be (and are!) considerably cheaper than homogeneous structures while at the same time possessing various physical characteristics that are identical to or superior to those of the homogeneous materials in a specific range of parameters (such as temperature, frequency, and applied field).

In the present paper an investigation will be made of a composite having a dielectric matrix (say, polypropylene) and we shall select macroscopic magnetic particles as the impurity phase. The dielectric constant and dielectric loss tangent, important in practical applications, will then be calculated as a function of the frequency of the applied electric field, the concentration of the magnetic phase x, and the temperature T.

Let us assume that the composite is a very thin film. An alternating electric field is directed along the plane of the film parallel to the z axis (Fig. 1). To determine the dielectric constant of this medium, we shall proceed as follows. Since the electric displacement vector $\mathbf{D}$ can be determined from the relation $\mathbf{D}=-\delta F/\delta \mathbf{E}_1$, where F is the free energy per unit volume of the composite and $\mathbf{E}_1$ is the electric field inside the composite, we write, neglecting interaction between the host matrix and the impurities,

$$F=(1-x)F_0+xF_1, \tag{1}$$

where F_0 is the free energy of the dielectric matrix and F_1 is the free energy of the magnetic particles.

Consequently, the components of the electric displacement vector are

$$D_\alpha=(1-x)D_{0\beta}\partial E_{0i\beta}/\partial E_{i\alpha}+xD_{1\beta}\partial E_{1i\beta}/\partial E_{i\alpha}, \tag{2}$$

where the subscripts are $\alpha,\beta=x,y,z$, $\mathbf{D}_1$ is the electric displacement inside the magnetic particles, $\mathbf{D}_0$ is the electric displacement in the pure dielectric (in the host matrix), and here and everywhere below a summation over repeated Greek indices is implied.

We adopt the following procedure to calculate the derivatives in expression (2). We express the internal fields E_{0i} and E_{1i} in the form

$$\mathbf{E}_{0i}=\mathbf{E}+4\pi\alpha_0\mathbf{E}_{0i}(N_0-L_0),$$
$$\mathbf{E}_{1i}=\mathbf{E}+4\pi\alpha_1\mathbf{E}_{1i}(N_1-L_1), \tag{3}$$

where N_0 and N_1 are coefficients of the depolarization tensor and depend on the shape of the objects, the coefficients $L_{0,1}$ are given by $E_{0,1L}=4\pi L_{0,1}\mathbf{P}_{0,1}$, where $\mathbf{E}_{0,1L}$ is the Lorentz field, and $\alpha_{0,1}$ are the polarizabilities of the two phases (dielectric and magnetic).

We have a similar relation for the ''total'' field $\mathbf{E}_i$:

$$\mathbf{E}_i=E+4\pi\alpha\mathbf{E}_i(N-L). \tag{4}$$

Here N are the coefficients of the depolarization tensor of the entire composite and the coefficients L are determined by

$$\mathbf{E}_L=4\pi\alpha L\mathbf{E}_i.$$

It now follows from expressions (3) and (4) that

$$\partial E_{0i\beta}/\partial E_{i\alpha}=(\partial E_{0i\beta}/\partial E_\gamma)(\partial E_\gamma/\partial E_{i\alpha})$$
$$=[1+(\varepsilon-1)(L-N)]/[1+(\varepsilon_0-1)\times(L_0-N_0)], \tag{5}$$
$$\partial E_{1i\beta}/\partial E_{i\alpha}=(1+(\varepsilon-1)(L-N))/(1+(\varepsilon_1-1)\times(L_1-N_1)), \tag{6}$$

Then, since

 © 1999 American Institute of Physics

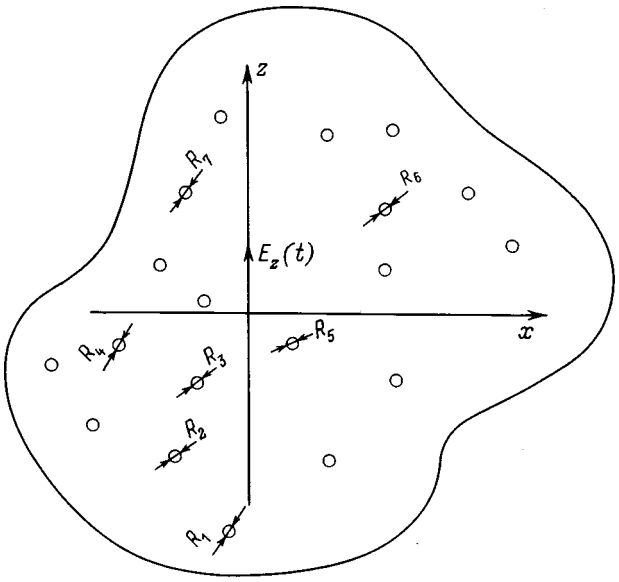

FIG. 1. Geometric configuration of coordinate axes and applied electric field.

$$\mathbf{D}_0=\varepsilon_0\mathbf{E}_{0i},\quad \mathbf{D}_1=\varepsilon_1\mathbf{E}_{1i},\quad \mathbf{D}=\varepsilon\mathbf{E}_i,$$

in accordance with expressions (2) and (3), ε is given by

$$\varepsilon=[(1-x)\varepsilon_0/(1+(\varepsilon_0-1)(L_0-N_0))^2+x\varepsilon_1/(1+(\varepsilon_1-1)(L_1-N_1))^2][1+(\varepsilon-1)(L-N)]^2.$$

Solving this equation for ε, we find

$$\varepsilon=(1-2\zeta(L-N)(1+N-L))/2\zeta(L-N)^2-\{[(1-2\zeta(L-N)(1+N-L))/2\zeta(L-N)^2]^2-(1+N-L)^2/(L-N)^2\}^{1/2},\tag{7}$$

where

$$\zeta=(1-x)\varepsilon_0/(1+(\varepsilon_0-1)(L_0-N_0))^2+x\varepsilon_1/\times(1+(\varepsilon_1-1)(L_1-N_1))^2.\tag{8}$$

Let us assume that the magnetic inclusions are spheres of radius R. We then have $L_1=N_1=1/3$, and expression (8) is simplified substantially:

$$\zeta=(1-x)\varepsilon_0/(1+(\varepsilon_0-1)(L_0-N_0))^2+x\varepsilon_1.\tag{9}$$

It can be seen from formula (7) that in the limiting cases when the sample is purely magnetic or purely dielectric, we obtain the natural expressions $\varepsilon=\varepsilon_1$ and $\varepsilon=\varepsilon_0$.

Formula (7) can be used to calculate the dependence of ε on the frequency ω of the applied electric field $\mathbf{E}=\mathbf{E}_0\exp(i\omega t)$, where $\mathbf{E}_0$ is the amplitude of the field. For this purpose we assume that

$$\varepsilon_0=1+i\varepsilon_0'',\quad \varepsilon_1=\varepsilon_1'+i\varepsilon_1''.$$

Extremely cumbersome but simple algebraic calculations lead us to the following expressions for ε' and ε'':

$$\varepsilon'=\{\zeta_1+2a(a-1)(\zeta_1^2+\zeta_2^2)-\zeta_1[d_1+(d_1^2+d_2^2)^{1/2}]^{1/2}/\sqrt{2}-\zeta_2[(d_1^2+d_2^2)^{1/2}-d_1]^{1/2}/\sqrt{2}\}/2a^2(\zeta_1^2+\zeta_2^2),\tag{10}$$

$$\varepsilon''=\{\zeta_2[d_1+(d_1^2+d_1^2)^{1/2}]^{1/2}-\zeta_1[(d_1^2+d_2^2)^{1/2}-d_1]^{1/2}-\zeta_2\sqrt{2}\}/2\sqrt{2}a^2(\zeta_1^2+\zeta_2^2),\tag{11}$$

where

$$d_1=1+4a(a-1)\zeta_1,\quad d_2=4a(a-1)\zeta_2,$$

$$\zeta_1=x+(1-x)A/C,\quad \zeta_2=(1-x)B/C+x\varepsilon_0'',$$

$$A=\varepsilon_0'[(1-b+b\varepsilon_0')^2-b^2(\varepsilon_0'')^2]+2b(\varepsilon_0'')^2(1-b+b\varepsilon_0'),$$

$$B=\varepsilon_0''[(1-b+b\varepsilon_0')^2-b^2(\varepsilon_0'')^2]-2b\varepsilon_0'\varepsilon_0''(1-b+b\varepsilon_0'),$$

$$C=[(1-b+b\varepsilon_0')^2-b^2(\varepsilon_0'')^2]^2+4b^2(\varepsilon_0'')^2(1-b+b\varepsilon_0')^2,$$

$$a=L-N,\quad b=L_0-N_0,\quad A>0,\quad \zeta_1>0,\quad C>0.\tag{12}$$

Using formulas (11) and (12) we can easily calculate $\tan\delta$. In fact, since the inequality $d_1\gg d_2$ is satisfied for any frequencies, we find

$$\tan\delta=\varepsilon''/\varepsilon'=(xB+(1-x)\varepsilon_0'')/(xA+(1-x)C)(1+4a(a-1)\zeta_1)^{1/2}.\tag{13}$$

Formula (13) can be simplified still further if we assume that the composite is a very thin film. In this case we have $N_0=1$, $L_0=0$, and thus $a=1$, so that

$$\tan\delta=[xB+(1-x)\varepsilon_0'']/[xA+(1-x)C].\tag{14}$$

This formula shows that $\tan\delta$ has no extrema as a function of the magnetic inclusion concentration x.

We shall consider the case when b is small or $b=0$ (the coefficients b will be calculated exactly a little later). Then, we obtain

$$\tan\delta=[(1-x)\varepsilon_0''+x\varepsilon_1'']/[(1-x)\varepsilon_0'+x].\tag{15}$$

Substituting

$$\varepsilon_1''=\omega\tau_1/(1+\omega^2\tau_1^2),\quad \varepsilon_0''=|\varepsilon_0|\omega\tau_0/(1+\omega^2\tau_0^2),$$

$$\varepsilon_0'=|\varepsilon_0|/(1+\omega^2\tau_0^2),$$

we find

$$\tan\delta=\frac{[(1-x)\tau_0 s(1+s^2)/\tau_1+xs(1+s^2\tau_0^2/\tau_1^2)]}{(1+s^2)[x(1+s^2\tau_0^2/\tau_1^2)+(1-x)|\varepsilon_0|]},\tag{16}$$

where $s=\omega\tau_1$.

It therefore follows from expression (16) that when the frequency dependence is taken into account, the dielectric loss tangent has a well-defined maximum at a specific frequency. Note that $\tan\delta$ for a completely homogeneous material is simply equal to $\omega\tau_0$, as expected. The qualitative difference between the absorptivities of the composite and a homogeneous dielectric can be seen clearly from Fig. 2.

In order to find the extremum frequency, we need to solve the cubic equation

$$y^3+A_1y^2+3B_1y-3C_1=0,\tag{17}$$

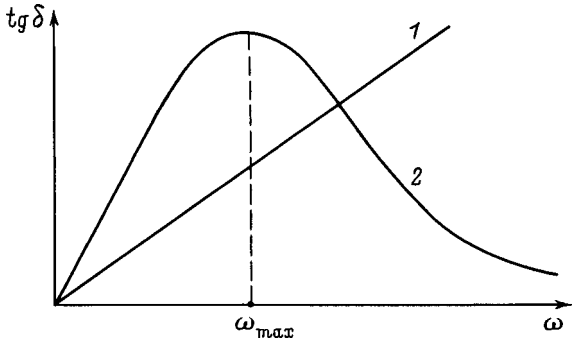

FIG. 2. Dependence of $\tan\delta$ on the frequency of the applied alternating electric field in two physical cases: *1* — homogeneous (single-phase) material, and *2* — two-phase structure.

where

$$y=s^2, \quad A_1=(9R_1/R_2)-1-R_0/R_3,$$

$$B_1=(R_1R_0/R_2R_3)+(R_1/R_2)-R_0/R_3,$$

$$C_1=R_0R_1/R_2R_3,$$

$$R_0=(1-x)|\varepsilon_0|+x, \quad R_1=(1-x)q+x,$$

$$R_2=3(xq^2+(1-x)q), \quad R_3=xq^2, \quad q=\tau_0/\tau_1. \quad (18)$$

Equation (17) has a single solution (since $y>0$!), which is determined by the condition

$$A<0. \quad (19)$$

Expressions (19) and (12) yield the following inequality

$$x^2g_1+2xg_2-|\varepsilon_0|<0, \quad (20)$$

where

$$g_1=(q-1)(|\varepsilon_0|-1-q^2-3q),$$

$$g_2=q^2-0.5q|\varepsilon_0|+0.5.$$

If $g_1>0$, which is satisfied by conditions (21) or (22), we find

$$1<q<(|\varepsilon_0|+5/4)^{1/2}-1.5, \quad (21)$$

where $|\varepsilon_0|>5$, or

$$(|\varepsilon_0|+5/4)^{1/2}-1.5<q<1, \quad (22)$$

where the inequality $1<|\varepsilon_0|<5$ must be satisfied.

If conditions (21) and (22) are satisfied, the concentration of magnetic additives satisfies

$$x<[(g_2^2/g_1^2)+|\varepsilon_0|/g_1]^{1/2}-g_2/g_1. \quad (23)$$

If $g_1<0$, which, unlike conditions (21) and (22), is satisfied by a different ''ballpark'' of given (exactly solvable!) times τ_0 and τ_1, we obtain

$$0<q<(|\varepsilon_0|+5/4)^{1/2}-1.5, \quad q<1, \quad (24)$$

or

$$q>1, \quad q>(|\varepsilon_0|+5/4)^{1/2}-1.5. \quad (25)$$

The system of inequalities (24) is not satisfied under real conditions (unless $|\varepsilon_0|=1$). Thus, there remains only one possibility

$$q>(|\varepsilon_0|+5/4)^{1/2}-1.5. \quad (26)$$

This inequality corresponds to concentrations

$$x>[(g_2^2/g_1^2)+|\varepsilon_0|/g_1]^{1/2}-g_2/g_1. \quad (27)$$

We shall now give numerical estimates of the right-hand sides of inequalities (23) and (27). Assuming that q is of the order of unity, we find $g_1\sim 1$, $g_2\sim -|\varepsilon_0|$, and as a result, we obtain two possibilities: x is greater than or smaller than $2|\varepsilon_0|$. In reality only $x<2|\varepsilon_0|$ is possible, and this means that

$$x<1. \quad (28)$$

In accordance with inequality (23), if conditions (21) and (22) are satisfied, we find

$$x<x_0,$$

where $x_0=[(g_2^2/g_1^2)+|\varepsilon_0|/g_1]^{1/2}-g_2/g_1$.

We write this inequality in expanded form

$$\tau_1<\tau_0<[(|\varepsilon_0|+5/4)^{1/2}-1.5]\tau_1, \quad \text{if } (|\varepsilon_0|>5), \quad (29)$$

$$[(|\varepsilon_0|+5/4)^{1/2}-1.5]\tau_1<\tau_0<\tau_1,$$

$$\text{if } (1<|\varepsilon_0|<5), \quad (30)$$

It should be noted that in principle, inequalities (29) and (30) can be satisfied by selecting suitable magnetic and dielectric materials. For magnetic structures at temperatures $T<T_c$, where T_c is the magnetic phase transition temperature (and this is the temperature range we are dealing with), the times τ_1 are determined by: a) magnon–magnon, b) magnon–phonon, and c) phonon–phonon relaxation mechanisms.[5–8] One particular mechanism will predominate in different temperature ranges. For a dielectric the relaxation time τ_0 should be determined by the relaxation of the polarization **P**. It should be noted that in the case of a phonon relaxation mechanism in both the dielectric and the magnetic substance, the behavior of the times τ_1 and τ_0 is qualitatively the same but differs quantitatively. This is because the interaction constants between phonons in the dielectric and the magnetic substance differ substantially, since they depend on the symmetry of the material. If this mechanism is taken to be the main one, q will simply be equal to the ratio of the appropriate striction constants K_1 and K_2, i.e.,

$$q=K_2/K_1. \quad (31)$$

Thus, the condition $q\sim 1$ is quite realistic. Conversely, if the magnon interaction mechanism predominates in the magnetic substance ($\tau_{1\text{magn}}<\tau_{1\text{phon}}$), the inequality $\tau_1<\tau_2<[(\varepsilon_{20}+5/4)^{1/2}-1.5]\tau_1$ is satisfied, and $|\varepsilon_0|>5$ [see Eq. (29)]. In this case, $\tan\delta$ has a well-defined maximum. If inequality (29) is not satisfied, no absorption maximum occurs, and $\tan\delta$ increases linearly with frequency.

We shall now analyze the particular case when $b=0$, which is achieved if $L_0=N_0$. The expression for $\tan\delta$ [see formula (14)] simplifies substantially, and, taking formula (12) into account, we find

$$\tan\delta=[(1-x)\varepsilon_0''+x\varepsilon_1'']/[x+(1-x)\varepsilon_0']. \quad (32)$$

The analysis of formula (32) is quite simple if $(1-x)\varepsilon_0'' \geqslant x\varepsilon_1'$, which can be achieved for low concentrations of the magnetic phase.

Setting

$$\varepsilon_0''=|\varepsilon_0|\omega\tau_0/(1+\omega^2\tau_0^2), \quad \varepsilon_1''=\omega\tau_1/(1+\omega^2\tau_1^2),$$

$$\varepsilon_0'=|\varepsilon_0|/(1+\omega^2\tau_0^2),$$

we find

$$\tan\delta=[|\varepsilon_0|(1-x)\omega\tau_0]/[(1-x)|\varepsilon_0|+x(1+\omega^2\tau_0^2)]. \quad (33)$$

The extremum of this function is localized at

$$\omega=\omega_{\max}=[1+(1-x)|\varepsilon_0|/x]^{1/2}/\tau_0. \quad (34)$$

For a given frequency of the external field, the concentration of the magnetic phase corresponding to the region of maximum absorption is found from Eq. (34) as

$$x_{\text{opt}}=|\varepsilon_0|/(\omega^2\tau_0^2+|\varepsilon_0|-1). \quad (35)$$

Briefly summarizing this analysis, we note that the absorption of electromagnetic radiation by dielectric composites can be regulated by adding magnetic, metallic, or other foreign substances to the host matrix. At this point, particular attention should be drawn to the fact that for metal inclusions the range of optimum absorption of electromagnetic radiation should be shifted toward shorter wavelengths. This is because the relaxation times in the electronic subsystem are generally appreciably shorter than those in the magnon and phonon subsystems (see Ref. 9, for example), and, consequently, the absorption maximum undergoes a hierarchical shift toward high frequencies.

CALCULATION OF DEPOLARIZING FACTORS

We shall consider the problem of calculating the depolarizing factors N_0 (the coefficients b) in the case of practical importance when the dielectric is a film and the foreign inclusions may be considered to be voids. We shall assume (Fig. 3) that the geometry of the configuration of spherical voids is an axisymmetric ''disk'' figure, i.e., along the z axis the voids are distributed along concentric circles with no voids at the center of the circle, and we shall calculate N_0 for this figure. According to the definition of the components of the depolarization tensor,[10] we have

$$N_{ik}=(1/4\pi)\int(\partial^2/\partial x_i\partial x_k)(1/r)dV$$

$$=(1/4\pi)\int\int_S(\partial/\partial x_i)(1/r)dS_k. \quad (36)$$

Beginning our calculations with N_{zz}, we write N_{0zz} in the form

$$N_{0zz}=(1/4\pi)\int\int dxdy\int_{-\delta}^{\delta}(\partial^2/\partial z^2)(1/r)dz. \quad (37)$$

To find the inner integral, we envisage a one-dimensional structure and represent this schematically as in Fig. 4, where the spheres (the voids are assumed to be spheres with some size spread) are denoted by two transverse

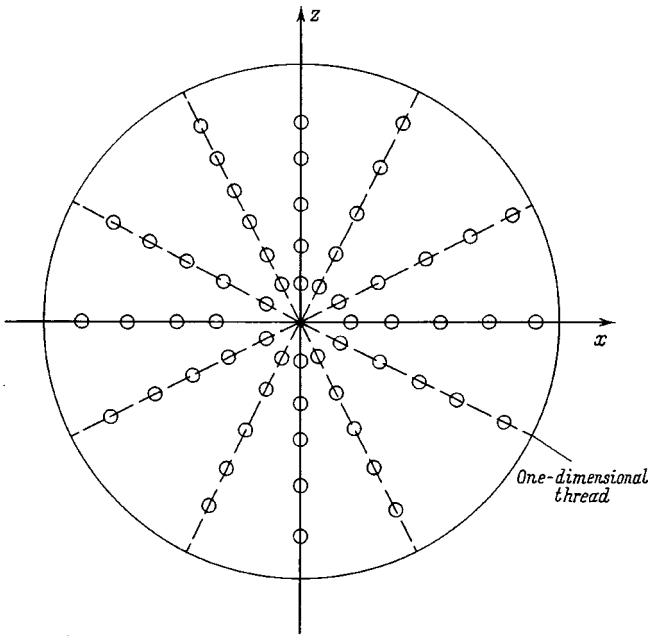

FIG. 3. Schematic diagram of a disk structure of voids. The concentric circles lie in the $y-z$ plane.

bars separated by a distance $2\varkappa$, and neighboring bars are a distance d apart. If the length of this thread is δ^*, it is easy to see that the total number of voids is

$$n_{\max}=(\delta^*-\varkappa)/(d+2\varkappa). \quad (38)$$

Then, since the concentration $x=\varkappa(1+2n_{\max})/\delta^*$, and $1-x=dn_{\max}/\delta^*$, formula (38) can be used to find a relation between d and $\varkappa$:

$$\varkappa=xd/2(1-x). \quad (39)$$

With this in mind, the inner integral is written as

$$J_{in}=-z/r^3\big|_{-\delta^*/2}^{\delta^*/2}=-\int_{\varkappa}^{d+\varkappa}(\partial/\partial z)(z/r^3)dz$$

$$-\int_{d+3\varkappa}^{2d+3\varkappa}(\partial/\partial z)(z/r^3)dz-\ldots$$

(the ellipsis stands for terms with $d\Rightarrow -d$, $\varkappa\Rightarrow-\varkappa$). For low concentrations of the magnetic phase as far as $x\leqslant 1/2$ the functions obtained as a result of integrating can be expanded in powers of $\varkappa$. As a result, we find

$$N_{0zz}=(1/4\pi)\int\int_{-M}^{M}dxdy\left\{2\varkappa/(x^2+y^2+\varkappa^2)^{3/2}\right.$$

$$\left.+4\varkappa\sum_{n=1}^{n_{\max}}(x^2+y^2+n^2d^2)^{-3/2}\right\}. \quad (40)$$

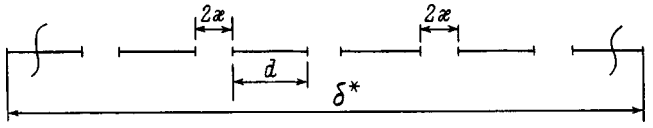

FIG. 4. Arbitrary representation of spherical voids in the form of linear gaps.

Converting to polar coordinates we obtain

$$N_{0zz}=(\varkappa/2)\left\{-(\rho^2+\varkappa^2)^{-1/2}-2\sum_{n=1}^{n_{\max}}(\rho^2+n^2d^2)^{-1/2}\right\}\Big|_0^M. \tag{41}$$

Having divided the radial threads into regions

$$\rho\in[0;d_1+\varkappa]\cup[d_1+3\varkappa;2d_1+3\varkappa]$$
$$\cup[2d_1+5\varkappa;3d_1+5\varkappa]\cup\ldots,$$

where $d_1=\sqrt{2d}$ and

$$N_{\max}=(M-\varkappa)/(d_1+2\varkappa), \tag{42}$$

as a result (as above), by expanding expression (41) in powers of $\varkappa$, we obtain the approximation

$$N_{0zz}=1-5\varkappa^2/2d_1^2\sum_{N=1}^{N_{\max}}[1/Nd-2d_1\varkappa/(d_1^2+N^2d^2)^{3/2}].$$

Substituting expression (39) and taking into account that $d_1=\sqrt{2d}$, we finally obtain

$$N_{0zz}=1-5x^2/16(1-x)^2+\{x/2(1-x)\}\sum_{N=1}^{N_{\max}}1/N. \tag{43}$$

We now have only to estimate the sum of the terms of the divergent harmonic series. We have the approximation

$$\sum(1/N)=N_{\max}\ln N_{\max},$$

and, allowing for Eqs. (42) and (39), we find

$$N_{0zz}=1-5x^2/16(1-x)^2+[x^2\ln(x/2)]/4(1-x). \tag{44}$$

For the components $N_{0xx}=N_{0yy}$, we have

$$N_{0xx}=N_{0yy}=(1-N_{0zz})/2=5x^2/32(1-x)^2$$
$$-[x^2\ln(x/2)]/8(1-x). \tag{45}$$

Thus, these formulas clarify the dependence of b on the concentration of voids. At this point, it should be reemphasized that the magnetic macroscopic additives have been replaced purely formally by voids. Note that in our notation $b=N_{0xx}$.

It can be seen from formula (45) that the dependence of the coefficients b on the concentration of the magnetic phase for small x is extremely weak (proportional to x^2), but even so, when the dependence $b(x)$ is taken into account, the functional behavior of $\tan\delta$ becomes very complex and can be represented schematically as in Fig. 5. It should be noted that this nonstandard behavior of $\tan\delta$ is characteristic of any two-component structure when a foreign substance having the concentration x is added to the host matrix, provided that this substance is not a dielectric!

Calculations of the dielectric constant in the form given above do not exhaust the possibilities of using other formulas to find ε. In particular, we can use the nonequilibrium density matrix method to determine the polarization coefficient α and therefore ε. In fact, we have

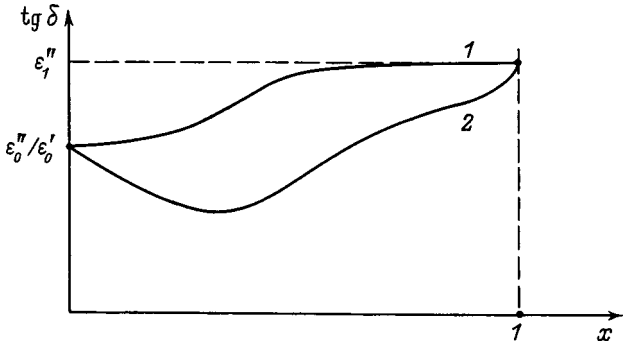

FIG. 5. Dependence of $\tan\delta$ on the concentration of the magnetic phase in two real cases: *1* — N_2 does not depend on x, *2* — N_2 is a function of x.

$$\alpha_{ik}(\omega)=\int\langle[d_i(0,x_1),d_k(t,x_2)]\rangle$$
$$\times\exp(i\omega t)d^3x_1d^3x_2dt/\hbar V^3, \tag{46}$$

where $d_i(t,x)$ is the atomic dipole moment operator, the brackets as usual indicate a commutator, and the angle brackets denote averaging over the equilibrium density matrix.

Direct calculations of this expression are fairly complicated (see Refs. 2 and 3, for example, where the Kubo formula was used to calculate the thermal conductivity), and cannot be made in a general form. However, this formula works in simple particular cases, and if we define the specific form of the operator $\mathbf{d}_i$ and the interaction Hamiltonian of the various subsystems, it can be simplified and the calculations brought to a logical conclusion. If we neglect the interaction between the two phases, dielectric and magnetic (at this point we note that any quantity of the phases may exist[3,11] but we shall then need to make some modification to the following formula), formula (46) may be rewritten somewhat differently, i.e.,

$$\alpha_{ik}(\omega)=x\int_{V_1}\langle[d_{1i}(0),d_{1k}(t)]\rangle\exp(i\omega t)dt/\hbar V+(1-x)$$
$$\times\int_{V_0}\langle[d_{2i}(0),d_{2k}(t)]\rangle\exp(i\omega t)dt/\hbar V, \tag{47}$$

which now contains the dipole moment operators $\mathbf{d}_1$ and $\mathbf{d}_2$ corresponding to the two phases, $x=V_1/V$, $1-x=V_0/V$, $V_1+V_0=V$, where V_1 is the volume of the magnetic phase, and V_0 is the dielectric component of the host matrix.

CONCLUSIONS

The dielectric constant of two-phase systems (dielectric + macroscopic magnetic particles) calculated above can be used to predict various characteristics of the absorption of external electromagnetic radiation by these substances.

1. The value of $\tan\delta$ for these structures has a well-defined maximum which is determined only by the impurity phase concentration x (if $x=0$ there is no maximum), and the absorption maximum is localized in a specific wavelength range.

2. In these composites the depolarizing factors begin to depend strongly on the impurity-phase concentration.

3. These composites (with a specific concentration x) can be used for shielding from a specific type of electromagnetic radiation in a particular frequency range.

This work was partly supported by the Russian Fund for Fundamental Research, Grant No. RFBR-96-03-03237.

[1] G. N. Dul'nev and Yu. P. Zarichnyak, *Thermal Conductivity of Mixtures and Composites* [in Russian], Leningrad (1974).
[2] S. O. Gladkov, Physica B **167**, 159 (1990).
[3] S. O. Gladkov, Fiz. Tverd. Tela (St. Petersburg) **39**, 1622 (1997) [Phys. Solid State **39**, 1446 (1997)].
[4] S. O. Gladkov, Zh. Tekh. Fiz **67**(7), 8 (1997) [Tech. Phys. **42**, 724 (1997)].
[5] A. I. Akhiezer, V. G. Bar'yakhtar, and S. V. Peletminskii, *Spin Waves* [North-Holland, Amsterdam, 1968; Nauka, Moscow, 1967].
[6] A. Clogston, H. Suhl, L. Walker, and P. Anderson, Phys. Rev. **101**, 903 (1956).
[7] M. Sparks, R. London, and C. Kittel, **122**, 791 (1961).
[8] S. O. Gladkov, Zh. Éksp. Teor. Fiz. **83**, 806 (1982) [Sov. Phys. JETP **56**, 452 (1982)].
[9] S. O. Gladkov, Phys. Rep. **182**, 211 (1989).
[10] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, 2nd ed. with L. P. Pitaevskii [Pergamon Press, Oxford, 1984; Nauka, Moscow, 1982].
[11] S. O. Gladkov, Pis'ma Zh. Tekh. Fiz **24**(10), 29 (1998) [Tech. Phys. Lett. **24**(5), 383 (1998)].

Translated by R. M. Durham

Perturbed machine phase ellipse and particle tracking in ultrahigh-energy linear accelerators

V. M. Tsakanov

Erevan Physics Institute, 375036 Erevan, Armenia
(Submitted March 4, 1998)
Zh. Tekh. Fiz. **69**, 95–98 (July 1999)

An investigation is made to determine how inaccuracies in mounting the quadrupole lenses influence the dilution of the transverse emittance of the beam. The study is made for two specific designs of 500 GeV linear colliders, thermal and superconducting. The analytical prediction is compared with the results of numerical calculations on the averaging of the beam emittance dilution effects for various sets of random deviations of the lenses from the accelerator axis. © *1999 American Institute of Physics.* [S1063-7842(99)02007-3]

A meticulous study of the beam behavior along the main linac with allowance for all the tolerances on the magnetic and accelerating elements is required in order to preserve the nominal emittance of the electron beam in a linear collider at energies of 0.5–1 TeV. For numerical simulations of the beam dynamics, various codes for particle tracking in the main linac[1–3] have been developed with comprehensive allowance for dilution of the transverse emittance of the electron beam,[4] and these continue to be refined. At this point, it should be noted that the enlargement of the normalized beam emittance in the main linac as compared with the nominal normalized beam emittance should not exceed a few percent.

The main beam broadening effects can be attributed to perturbation of the central trajectory caused by inaccurate installation of the quadrupole lenses along the accelerator. Even with tolerances of the order of 100 μm for the displacement of the lenses relative to the axis, the perturbed trajectory may exceed the beam dimensions by several orders of magnitude, leading to transverse beam instability caused by the excitation of transverse modes in the accelerating sections.[5] The only method of allowing for inaccurate mounting of the quadrupole lenses along the accelerator is to consider the deviations of the lenses relative to the axes as a random set of uncorrelated displacements with a mean-square alignment precision. However, the perturbed central beam trajectory may differ by orders of magnitude for two different sets so the beam emittance dilution effects must be averaged over a large set of deviations of the lenses from the axis. The important question then arises as what criterion should be used for the reliability of the particle tracking results relative to the number of averagings. The particle tracking results are usually confined to averaging over 20–30 sets of random deviations of the lens center from the axis. This is because, up to energies of around 50 GeV, the mean-square dimensions of the beam in the main linac vary negligibly as the number of averagings is increased further. However, as will be shown below, this cannot serve as a guarantee of the reliability of the results at higher accelerator energies, since it does not take into account the larger number of lenses and the accumulation of averaging errors.

Here we introduce the mean-square perturbed phase ellipse of the transverse oscillations of the centroid of the beam due to inaccurate installation of the quadrupole lenses along the accelerator. The area of the phase ellipse is expressed analytically in terms of the accelerator parameters. The agreement between the numerical and analytical results for the particle tracking will then serve as the criterion for the reliability of the particle tracking results along the entire accelerator. The perturbed phase ellipse is determined by the focusing system along the accelerator and the rate of particle acceleration, and is an important characteristic of the accelerator from the point of view of tolerances for the accelerator elements. In addition, a knowledge of the perturbed phase ellipse can be applied to optimize the focusing system and select a strategy for correcting the perturbed orbit. We present results of an investigation for the thermal (SBLC, frequency of rf system 3 GHz) and superconducting (TESLA, frequency of rf system 1 GHz) designs of 500 GeV linear colliders.[6] The equation describing the central beam trajectory, allowing for random deviations of the quadrupole lenses from the accelerator axis, is given by

$$\frac{1}{\Gamma(z)}\frac{d}{dz}\Gamma(z)\frac{d}{dz}x(z)-K_x(z)[x(z)-x_{qk}]=0, \tag{1}$$

where z and x are the longitudinal and transverse coordinates, x_{qk} are the random displacements of the lenses from the axis, K_x is the power of the quadrupole lenses, and Γ is the Lorentz factor of an equilibrium particle along the accelerator.

The particular solution of the equation of motion corresponding to the perturbed central beam trajectory may be expressed in the form

$$x_c(z)=\sqrt{\frac{\beta(z)}{\Gamma(z)}}\sum_k K_k L_{qk} x_k \sqrt{\beta_k \Gamma_k}\sin[\phi(z)-\phi(z_k)], \tag{2}$$

where $\beta(z)$ and $\phi(z)$ are the instantaneous values of the amplitude function and the phase of the transverse betatron oscillations, L_q is the length of a quadrupole lens, and terms with the subscript k correspond to their values in the kth quadrupole.

 © 1999 American Institute of Physics

We assume that the random deviations of the lenses from the axis are uncorrelated, and consequently the cross terms make no contribution to the mean-square displacement of the central trajectory ($\langle x_k x_l\rangle = 0$ for $k \neq l$)

$$\langle x_c^2\rangle = \langle x_q^2\rangle \frac{\beta(z)}{\Gamma(z)} \sum_k K_k^2 L_{qk}^2 \beta_k \Gamma_k \sin^2[\phi(z) - \phi(z_k)], \tag{3}$$

where $\langle x_q^2\rangle^{1/2}$ is the root-mean-square displacement of the centers of the lenses relative to the axis.

At this stage, by analogy with the free betatron oscillations of the particles, we can introduce the mean-square perturbed instantaneous phase ellipse of the beam

$$\gamma\langle x_c^2\rangle + 2\alpha\langle x_c x_c'\rangle + \beta\langle x'^2_2\rangle = A^2, \tag{4}$$

where the area of the ellipse (divided by π) is

$$A^2 = \frac{\langle x_q^2\rangle}{\Gamma(z)} \sum_k K_k^2 L_{qk}^2 \beta_k \Gamma_k, \tag{5}$$

and α, β, and γ are the parameters of the Twiss matrix.

We shall assume that, as usual, the focusing system of the accelerator is a symmetric FODO structure (where F is a focusing lens, D is a defocusing lens, and O is a free gap), which includes accelerating sections. Then, replacing the summation over lenses by a summation over periodic cells and using the relations for a symmetric FODO structure,

$$\beta_{\max} + \beta_{\min} = \frac{2L_c}{\sin\mu}, \quad KL_q L_c = 4\sin\frac{\mu}{2}, \tag{6}$$

we obtain

$$A^2 = 16\frac{\langle x_c^2\rangle}{L_c\Gamma(z)} \sum_n \Gamma_n \tan\frac{\mu_n}{2}, \tag{7}$$

TABLE I.

Accelerator	Cell length L_c, m	Energy increment ΔE, eV/m	Phase shift μ	Lens displacement a_q, μm
SBLC	12	17	$\pi/2$	100
TESLA	48	25	$\pi/3$	500

where μ_n is the phase shift of the betatron oscillations in the nth cell, and L_c is the length of a FODO cell.

If the number of cells is large and the and energy increment per cell is comparatively small, we can convert from summation over cells to an integral over energy, which in the particular case of a constant phase shift gives the following analytical expression for the area of the perturbed phase ellipse:

$$A^2 = 8\frac{\langle x_q^2\rangle}{L_c^2}\frac{E_0}{\Delta E}\tan\frac{\mu}{2}\frac{E_0}{E(z)}\left[\left(\frac{E(z)}{E_0}\right)^2 - 1\right], \tag{8}$$

where E_0 and $E(z)$ are the initial and instantaneous equilibrium beam energies and ΔE is the energy increment per unit length.

The mean-square displacement of the lenses is related to the tolerance a_q by $a_q^2 = 3\langle x_q^2\rangle$. We note immediately that the commonly used phase shift per period $\pi/2$ (Ref. 4) is far from optimum from the standpoint of the perturbed central beam trajectory, and, in view of the fact that the amplitude of the free betatron oscillations is determined by a beta function, this shift can be reduced to values of the order of $\pi/6$ without materially affecting the maximum amplitude of the free betatron oscillations. Table I gives the main parameters for the thermal (SBLC) and superconducting (TESLA) designs of linear colliders.[6]

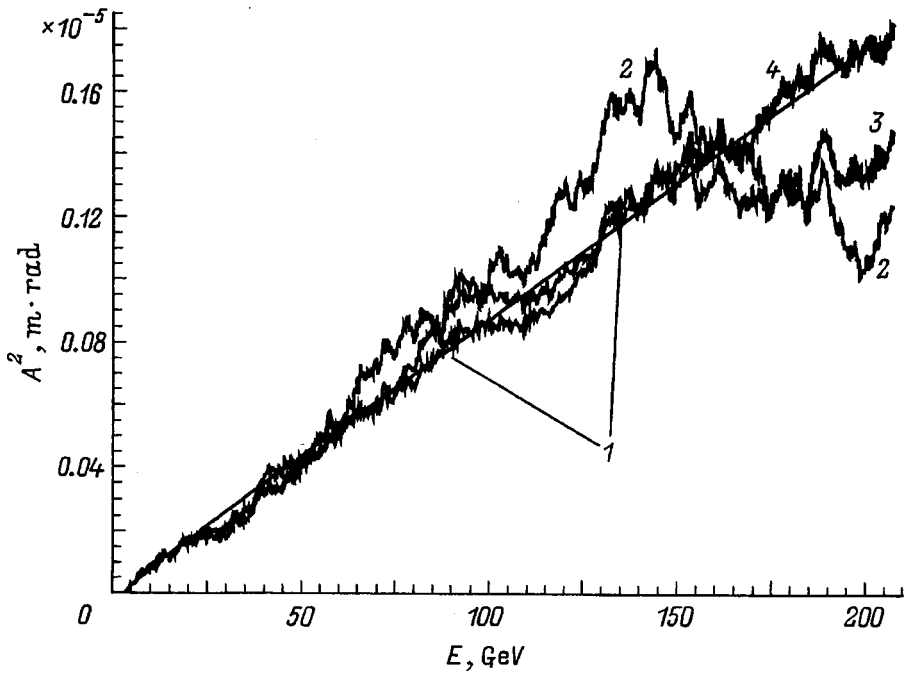

FIG. 1. Evolution of the area of the mean-square perturbed phase ellipse along the SBLC main linac with a 100 μm tolerance for the installation of the quadrupole lenses: *1* — analytical prediction, *2 –4* — averaging over 25, 50, and 100 sets of random deviations, respectively.

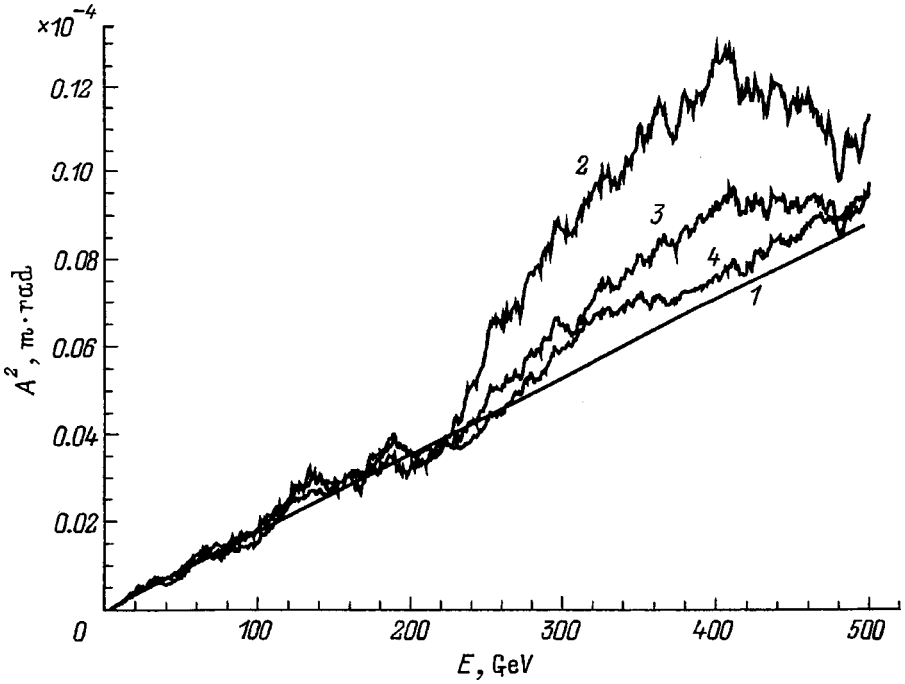

FIG. 2. Evolution of the area of the mean-square perturbed phase ellipse along the TESLA main linac with a 500 μm tolerance for the installation of the quadrupole lenses: *1–4* — as in Fig. 1.

The $\sim$1 GHz superconducting accelerator is distinguished by a beam tube of comparatively large aperture whose transverse wakefields are considerably weaker than those of the $\sim$3 GHz thermal accelerator for the same parameters of the accelerated beam. This means that one can have four (rather than two) accelerating modules per period of the FODO focusing system, substantially reducing the perturbed phase ellipse and consequently increasing the tolerances for the installation accuracy of the quadrupole lenses.

Figures 1 and 2 show the change in the area of the mean-square perturbed phase ellipse for the thermal and superconducting accelerators for 25, 50, and 100 sets of random equally probable deviations of the lenses from the axis, and also the analytical prediction. As we can see, the particle tracking results show good agreement with the analytical calculations when the mean-square perturbed trajectory is averaged over 100 or more sets of lens displacements. We also note that even when averaged over 50 sets of deviations the mean-square phase ellipse differs appreciably from the real one at electron energies higher than 100 GeV for the thermal variant and 250 GeV for the superconducting variant, which must be borne in mind when interpreting the particle tracking results to determine the mean-square dilution of the beam emittance at the accelerator exit. If the accelerator also has a system for guiding and final focusing of the particles at the collision point of the accelerated beams, we can also see how important it is to have an accurate idea of the central trajectory and the mean-square beam emittance at the accelerator exit.

We have derived a simple analytical formula which describes the perturbed mean-square phase ellipse in linear accelerators at high energies. The area of the ellipse may serve as a measure of the accuracy of the particle tracking along the accelerator when determining the mean-square parameters of the beam. In addition, the perturbed phase ellipse is an important characteristic of an accelerator from the point of view of tolerances for the installation of the quadrupole lenses and from the point of view of optimizing the entire focusing system.

The author would like to thank R. Brinkmann and R. Wanzenberg for worthwhile discussions which provided the stimulus for this work.

[1] M. Drevlak, M. Timm, and T. Weiland, in *Proceedings of the 18th International Linac Conference*, Geneva, 1996, pp. 621–623.
[2] A. Mosnier and A. Zakharian, in *Proceedings of the Fourth European Particle Accelerator Conference*, London, 1994, pp. 1111–1113.
[3] R. Assman *et al.*, SLAC Report No. AP-103, (1997), p. 114.
[4] T. Raubenhaimer, PhD Thesis, SLAC-387 (1991), pp. 111–252.
[5] A. Chao, *Physics of Collective Beam Instabilities in High Energy Accelerators* (Wiley New York, 1993), p. 286.
[6] R. Brinkmann, DESY Report No. DESY1997-048 (1997) (2 Vols.), pp. 275–861.

Translated by R. M. Durham

Manifestation of empty electron states in the total current spectra of silver

V. A. Novolodskiĭ, O. M. Artamonov, and S. A. Komolov

St. Petersburg State University, Research Institute of Physics, 198904 St. Petersburg, Russia
(Submitted April 27, 1998)
Zh. Tekh. Fiz. **69**, 99–102 (July 1999)

A combined analysis is made of the low-energy total current and secondary electron emission spectra at the initial stage of formation of a silver coating on a W(110) surface. It is shown that the features observed in the spectra are related to the energy structure of the density of empty states lying above the vacuum level. The dynamics of changes in these features in the total current spectra as the film thickness increases indicates that at the initial stage the silver coating is characterized by the Ag(111) orientation, but as the thickness increases, the Ag(110) orientation predominates. © *1999 American Institute of Physics.*
[S1063-7842(99)02107-8]

INTRODUCTION

The electron spectra of the surface of silver were studied by two methods of electron spectroscopy in which the energy structure of the solid was studied in the same range of electron energies. The first method was secondary electron spectroscopy (SES), in which we studied the spectra of the true secondary electron emission near the cascade maximum in the energy range between 0 and 30 eV (Ref. 1). The primary electron energy and current are kept constant during recording of the SES spectra. The secondary electron emission was recorded in a narrow solid angle using a hemispherical energy analyzer. This SES method was described in detail in Refs. 2 and 3, and the emission spectra are given in the form $N(E)$.

The other method was total current spectroscopy (TCS),[4,5] in which the current of electrons propagating in the sample circuit is recorded when its surface is exposed to a constant current of primary electrons whose energy varies uniformly between 0 and 30 eV. The total current spectrum is identical to the spectrum of electron reflection into an inverted hemisphere, provided that the primary current is the same and that all electrons leaving the solid are collected, i.e.,

$$I_{\mathrm{TCS}}(E) = -I_{\mathrm{refl}}(E).$$

An advantage of total current spectroscopy is that the experimental equipment is simple compared with that required to record the reflection spectra. The total current spectra are expressed as the derivative $dI(E)/dE$ of the current flowing in the sample circuit.

In the present study we analyze silver films on single-crystal tungsten W(110) and also polycrystals of silver. The surface was analyzed in a vacuum $P = 6 \times 10^{-8}$ Pa. The silver films were deposited *in situ* by thermal evaporation. The silver films and polycrystals were studied by total current spectroscopy. The surface of the tungsten substrate was prepared by a standard method of oxidation followed by heating. The silver sample was purified by heating in vacuum for several hours until the surface recrystallized.

EXPERIMENTAL RESULTS

The secondary electron emission spectra near the cascade maximum were studied at primary electron energies E_p between 10 and 400 eV. From energies $E_p = 25$ eV and higher (the energy is measured from the Fermi level), the spectra exhibited fundamentally the same profile, as shown in Fig. 1 (curve *3*). This curve has two peaks, one being a cascade peak recorded at 7.5 eV and the other an emission peak *A* recorded at $E = 17.6$ eV, which is a characteristic feature of the emission spectrum of silver. This feature only appeared in the emission spectrum at primary electron energies above $E_p = 25$ eV. As E_p increases further, the amplitude of this peak increases in proportion to the increase in the amplitude of the cascade peak. For thin (0.8–2 ML) silver films on the surface of tungsten this peak is located at 16.3–16.5 eV, but as the film thickness is increased from 2 to 6 ML, the peak shifts toward higher energies as far as 17.6 eV, and then its energy position remains unchanged with further growth of the film.

The total current spectra of the film surface and the silver polycrystal were identical for films more than 6 ML thick. This spectrum is shown in Fig. 1 (curve *2*). At a primary electron energy of 4.5 eV relative to the Fermi level we observed a primary peak typical of this type of spectrum. It is associated with the energy threshold for the entry of electrons into the solid and corresponds to the vacuum level for silver. The value of 4.5 eV is the electron work function for a silver surface.

As the primary electron energy increases, the total current spectra reveal other features characteristic of silver at energies of 7.2, 9.7, 17.3, and 23 eV denoted as *B*, *C*, *A*, and *D*, respectively, in Fig. 1 (curve *2*).

For thin (0.5–2 ML) silver films the amplitudes of features *B*, *C*, and *D* are small. Peak *A* predominates in the total current spectrum of the thin film, and is shifted 1 eV to the left relative to its position in Fig. 1 (curve *2*).

DISCUSSION OF RESULTS

Figure 1 (curve *1*) gives the calculated integral density of electron energy states for silver.[6] The amplitude of the

 © 1999 American Institute of Physics

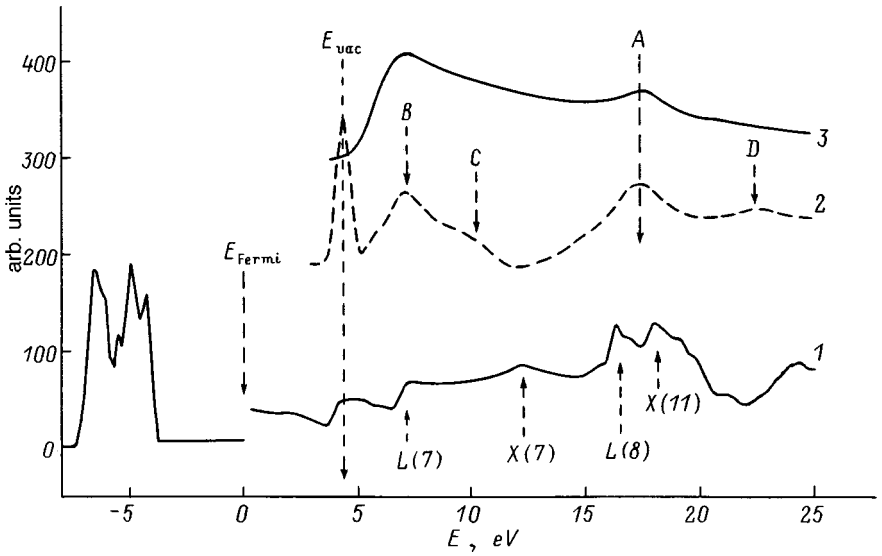

FIG. 1. Spectra of silver: *1* — density of electron energy states, *2* — total current spectrum, and *3* — secondary electron emission spectrum.

empty states is four times that of the filled states. The lower arrows indicate the density-of-states peaks corresponding to different directions of the dispersion relation for the electron energy. For instance, the features denoted by *X* indicate the $\Gamma-X$ directions perpendicular to the Ag(110) face whereas those denoted by *L* indicate the direction perpendicular to the Ag(111) face. The numbers in parentheses after *X* and *L* correspond to different branches of the same direction on the dispersion relation.

The nature of the emission peaks in the SES spectra has been fairly well studied.[7–9] For instance, peak *A* in the SES spectra of silver can be attributed to a maximum of the empty electron energy states above the vacuum level at 17.6 eV. Primary electrons with energy E_p sufficient to ionize filled valence states below the Fermi level (Fig. 1, curve *1*) excite electrons from these states. These secondary electrons occupy empty states between the excited state and the level of E_p. They can enter vacuum from these states if the energy of these states is higher than the vacuum level. The maximum density of empty states corresponds to the emission peak in the SES spectrum. In Fig. 1 (curve *1*) the group of density-of-states peaks at 16.3–17.7 eV shows good correspondence with emission peak *A* in the SES spectrum (Fig. 1, curve *3*). This shift of the emission peak from 16.3 to 17.6 eV with increasing thickness of the silver film is clearly caused by a change in the crystalline orientation of the surface in the thin and thick silver films on W(110). We know[10] that thin silver films have the orientation Ag(111), whereas thick ones become polycrystalline. The position of peak *A* in the SES spectrum of the thin film coincides with the 16.3 eV density-of-states peak characterizing the Ag(111) face. The shift of peak *A* toward 17.6 eV as the film becomes thicker matches the 17.7 eV density-of-states peak in Fig. 1 (curve *1*), which corresponds to the Ag(110) face.

The matching between the energy position of peak *A* on the TCS and SES spectra requires some explanation. The emission peak *A* in the SES spectrum indicates that 17.6 eV electrons are emitted from silver, while the TCS peak *A* indicates that primary electrons of this energy are absorbed by the silver film. Both these peaks coincide with an extremum of the density of empty electron states in silver, i.e., the same density-of-states peak is responsible for the emission and absorption of electrons having the same energy in the silver film.

No contradictions arise here because peak *A* appears in the emission spectra only when the primary electron energy E_p reaches a certain value which is higher than the energy of the 17.6 eV emission peak. Figure 2 shows the behavior of the emission spectra as the energy of the primary electrons decreases from 33.5 to 21.5 eV. The cascade peak background has been subtracted from the curves. It can be seen that the emission peak *A* decreases with decreasing E_p. At $E_p=21.5$ eV (curve *7* in Fig. 2), the emission spectrum barely reveals any peak. As E_p decreases, beginning at $E_p=27.5$ eV (curve *4* in Fig. 2) the suppressed peak *A* overlaps with the left wing of a group of peaks characterizing the electron energy losses. This group consists of energy loss peaks which are attributed to the energy lost by primary electrons in the excitation of various plasma oscillations that are characteristic of silver. We know[11] that this group consists of high-intensity peaks with maxima at 3.6, 3.8, and 7.5–8 eV. Unlike the emission peaks, whose energy position remains constant in all the emission spectra of a particular material, the characteristic losses keep their position constant relative to the peak of elastically reflected electrons. For instance, for the emission spectra of silver shown in Fig. 2, the right-hand energy boundary for the loss band is 3.2 eV from the elastic reflection peak, whereas the left-hand boundary is 7–8 eV from this peak. As E_p decreases, this group of losses shifts with the elastic reflection peak, as shown by the dotted arc *P*.

The energy position of the *A* emission peak remains constant in Fig. 2, while its amplitude decreases with decreasing E_p. This variation of the amplitude of peak *A* shows that as E_p decreases, there is a lower probability of filled valence states becoming ionized and a lower probability of empty states being filled with excited electrons, which reduces the emission. It can be seen from Fig. 2 that even for $E_p=25.5$ eV (curve *5* in Fig. 2) the SES spectrum contains no emission peak *A*, although the characteristic loss peaks are still visible. For $E_p=17.6$ eV the emission spectrum

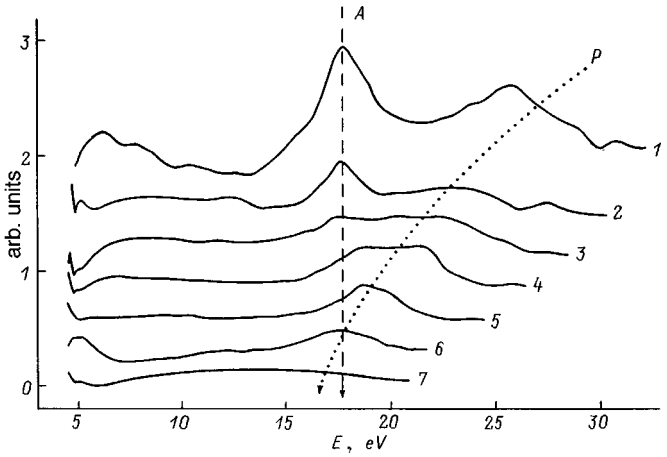

FIG. 2. Secondary electron emission spectra with the cascade maximum subtracted: $E_p=33.5$ (*1*), 31.5 (*2*), 29.5 (*3*), 27.5 (*4*), 25.5 (*5*), 23.5 (*6*), and 21.5 (*7*).

should change qualitatively because at this energy the coefficient of elastic reflection of the electrons decreases. This is indeed observed for the total current spectrum (Fig. 1, curve *2*) at an electron energy of 17.3 eV. This spectrum reveals a peak *A* which indicates increased absorption of electrons having this energy. For $E_p=17.5$ eV primary electrons are evidently ''trapped'' at empty energy states of silver, and a reduction in the coefficient of elastic reflection is observed.

The total current spectra are expressed as the derivative of the current $dI(E)/dE$ flowing through the sample. Therefore, the maximum in the total current spectrum should correspond not to the density maximum but to the maximum upward slope of the density of empty states, as is observed experimentally (Fig. 1, curves *1* and *2*). This corresponds to a shift of peak *A* by 0.3 eV to the left in the total current spectrum relative to peak *A* in the SES spectrum.

In the total current spectra for thin silver films the shift of peak *A* at 16.5 eV, as was noted above, matches the shift of peak *A* in the SES spectrum and also indicates the predominant crystalline orientation of the thin Ag(111) films.

The total current spectra of thick films and silver polycrystals exhibit peaks *B*, *C*, and *D* (Fig. 1, curve *2*). These extrema show good agreement with the features in the density of states (Fig. 1, curve *1*). All these features (Fig. 1, curve *1*) corresponding to peaks *B*, *C*, and *D* are characteristic of the Ag(110) face. The peaks *B*, *C*, and *D* are also present in the total current spectrum, with correspondence being observed for these in the density-of-states spectrum but not in the SES spectrum. The more complex structure of the total current spectra as compared with the SES spectra between 0 and 30 eV can be attributed to the absence of the cascade maximum in this method of recording. In particular, the high-intensity *B* peak, which has the same energy position as the cascade maximum, is obscured by it and is therefore not recorded in the emission spectra. Features *C* and *D* are weak and only appear in the SES spectra at the noise level, whereas they are quite significant in the total current spectra.

CONCLUSIONS

The use of a combination of low-energy secondary electron spectroscopy methods (SES and TCS) has allowed us to identify the main features of the density of empty electron states located in the 0–20 eV range above the vacuum level at the initial stage of the formation of thin silver films. The energy position of the experimental points in the energy dependence of the density of empty states was determined and was found to be in good agreement with the results of theoretical calculations.

The behavior of the main features in the total current spectrum confirms that thin silver films on a W(110) tungsten surface possess the Ag(111) orientation, whereas for thick films the Ag(110) orientation predominates.

[1] I. M. Bronshteĭn and B. S. Fraĭman, *Secondary Electron Emission* [in Russian], Nauka, Moscow (1969), 386 pp.
[2] M. V. Kremkov, *Corpuscular Low-Energy Diagnostics of a Solid Surface* [in Russian], FAN, Tashkent (1986), 163 pp.
[3] O. M. Artamonov, O. M. Smirnov, and A. N. Terekhov, Izv. Akad. Nauk SSSR, Ser. Fiz. **46**, 1383 (1982).
[4] S. A. Komolov, *Integral Methods of Secondary Electron Spectroscopy* [in Russian], Leningrad State University Press, Leningrad, (1986), 180 pp.
[5] S. A. Komolov, *Total Current Spectroscopy of Surfaces* (Gordon and Breach, Philadelphia, 1992), p. 257.
[6] R. Lasser, N. V. Smith, and R. L. Benbow, Phys. Rev. B **21**, 1895 (1981).
[7] O. M. Artamonov and A. N. Terekhov, Fiz. Tverd. Tela (Leningrad) **28**, 862 (1986) [Sov. Phys. Solid State **28**, 479 (1986)].
[8] G. Chiarello, R. G. Agostino, A. Amoddeo *et al.*, J. Electron Spectrosc. Relat. Phenom. **70**, 75 (1994).
[9] S. A. Komolov, N. B. Gerasimova, A. O. Morozov *et al.*, Phys. Low-Dim. Struct. **10**, 35 (1997).
[10] E. Bauer, H. Poppa, G. Todd *et al.*, J. Appl. Phys. **48**, 3773 (1977).
[11] C. Kunz, in *Optical Properties of Solids. New Developments*, edited by B.O. Seraphin (American Elsevier, New York, 1976), p. 499.

Translated by R. M. Durham

Influence of lubricant on the motion of a body in an electromagnetic railgun accelerator. I. Electric current distribution in the accelerated body and the rails

É. M. Drobyshevskiĭ, É. N. Kolesnikova, and V. S. Yuferev

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted January 21, 1998)
Zh. Tekh. Fiz. **69**, 103–111 (July 1999)

An investigation is made of the influence of a liquid conducting film inserted in the gap between the accelerated metal projectile and the rails in a railgun accelerator on the distribution of the current density and the magnetic induction in the rails and the projectile. The film is assumed to be fairly thin, so that the variation of the magnetic induction across the film can be neglected. The problem is considered in a plane geometry. Approximate expressions are derived for the thickness of the velocity skin layer as a function of the thickness and conductivity of the film. The equations for diffusion of the magnetic field in the rails and in the projectile are integrated numerically. It is shown that the presence of a liquid conducting film can significantly suppress the velocity skin effect. Nevertheless, even when fairly thick films of fairly high resistivity are used, the nonuniformity of the electric current density distribution in the projectile still remains very high for the magnetic Reynolds numbers of practical interest. © *1999 American Institute of Physics.* [S1063-7842(99)02207-2]

INTRODUCTION

As we are well aware, velocities significantly in excess of 1–2 km/s have yet to be achieved in electromagnetic accelerators with a metal projectile (armature). The reason for this is that at these velocities the metal contact between the rails and the projectile disappears and is replaced by an arc discharge, with the result that most of the electric current energy is not expended in accelerating the body but in sustaining the electric arc and eroding the contacting surfaces. According to present thinking there are two main reasons for this transition. The first involves the onset of the velocity skin effect,[1–18] and the second is related to the characteristics of current flow across the contact interface between the two conductors and to wear of the sliding contact.[5,6,19–23] In the velocity skin effect, when an accelerated body moves at high velocities, the magnetic field and current flowing through it are concentrated in its rear section, causing an abrupt increase in the Joule heat release in the contact zone and giving rise to a melting wave which propagates forward over the surface of the accelerated body in the direction of motion. When this wave reaches the front edge of the moving body, the metal contact between the body and the rails disappears and is replaced by an arc discharge.

Various approaches have been put forward to describe the velocity skin effect. Only a few of these (Refs. 14 and 15, for instance) are based on direct numerical integration of the equations for diffusion of the magnetic field and heat transfer in the rails and the moving body and in most cases, approximate semianalytical models were used, based on various assumptions. Calculations of the velocity at which the metal contact goes over to arcing made using various models give values of 0.5–2 km/s, depending on the projectile design and the conductivity of the rails. It was shown in Refs. 4,11,13, and 17 that when the projectile is made of a high-resistivity material, the transition velocities in the contact zone may exceed 2.5 km/s. The transition velocity can also be increased by depositing a high-resistivity layer on the surface of the rails.[7,9,15] However, Glinov *et al.*[7] noted that the resistivity of this layer must be between three and four orders of magnitude higher than that of copper.

In studies of the velocity skin effect it is usually assumed that the contact between the rails and the moving body is ideal. However, the real pattern of current flow across the interface of a sliding contact between solid surfaces is far from ideal, and this is the second reason for the transition to the plasma regime. In addition, as a result of the roughness of the contact surfaces, current flows across the contact not over the entire nominal contact zone but in so-called alpha spots, where the two surfaces approach each other at distances comparable to the interatomic distance and the electric current lines become concentrated. At high current densities the alpha spots will undergo heating and melt, which may impair the contact and cause a transition to arcing. In particular, Basilevskiĭ *et al.*[6] showed on the basis of experimental data that at least up to velocities of 1850 m/s, this transition is almost independent of velocity and only depends on the nominal contact area, the contact compression force, and the electric action ($\int I^2 dt$).

Another factor influencing the quality of the contact is frictional heating and wear of the contacts. The influence of frictional heating in railgun accelerators was studied in Refs. 5, 6, and 22–24. At high sliding velocities the contact surfaces of the bodies will melt, giving rise to a molten film[25–27] and additional wear of the contact surfaces. Estimates show that melting of the contact boundary of a projectile due to frictional heating may begin at approximately the same velocities as does the melting due to Ohmic heating as a result of the velocity skin effect.

 © 1999 American Institute of Physics

Consequently, as our analysis shows, in existing designs of electromagnetic accelerators the motion of bodies at high velocities is limited by the simultaneous action of the two factors mentioned: 1) the velocity skin effect, and 2) local damage to the contact surface as a result of the flow of electric current and frictional heating. The velocity skin effect is a fundamental constraint arising from the principle of the electromagnetic acceleration of bodies, and special projectile designs must be employed required to overcome this. However, we can hardly expect any significant success in this direction if we do not solve the second problem and ensure good electrical contact between the projectile and the rails. Moreover, we can consider that the constraint of the velocity skin effect only begins to become the limiting factor in this case.

A substantial improvement in the electrical contact between the rails and the metal projectile can be achieved by specifically inserting a liquid conducting lubricant in the gap between them. First, this eliminates the problem of current concentration in alpha spots, and second, the local frictional heating at the rails–armature interface can be reduced appreciably because dry friction is replaced by viscous dissipation. At first glance, it seemed that the molten film formed as a result of the melting of the contacting surfaces during motion of the projectile in the railgun could be used as the liquid film. However, it would be almost impossible to ensure a stable state and maintain a continuous liquid film between the projectile and the rails by melting the projectile. In addition, this film could not suppress the velocity skin effect because its resistivity would differ very little from that of the projectile. Therefore, in order to achieve a real effect from the use of lubricant, a foreign, electrically conducting liquid must be intentionally inserted in the gap from outside. Quite obviously, the presence of a liquid film may have negative as well as positive influences. In particular, the high electrical resistance of the films leads to additional Joule heating.

Bearing these factors in mind, we attempt to make a detailed theoretical and numerical study of the action of a liquid conducting lubricant on various aspects on the acceleration of a metal projectile. The following problems were considered: the influence of the liquid film on the distribution of the magnetic field and electric current in the armature and rails, the flow of lubricant in the gap between the rails and the accelerated body, and the influence of the film on heating of the accelerated body. The first problem is considered in the present paper and the others will be examined in our next two studies of this problem. The specific method of inserting the lubricant in the gap between the rails and the projectile is not discussed. In principle, this can either be accomplished by depositing a suitable layer on the surface of the rails or by supplying lubricant from the projectile. Here we simply assume that the rails and the metal projectile are separated by a gap filled with a liquid conducting lubricant. All these problems are considered in a plane geometry.

FORMULATION OF THE PROBLEM

In order to simplify the problem we confine ourselves to the simplest projectile shape, i.e., a rectangle. The problem is shown schematically in Fig. 1. The coordinate system is related to the projectile such that the rails move at the velocity U in the positive direction of the x axis. Since the problem is planar, only the y component of the magnetic induction is nonzero, its distribution in the armature and the rails being given by the following equations and boundary conditions: in the armature

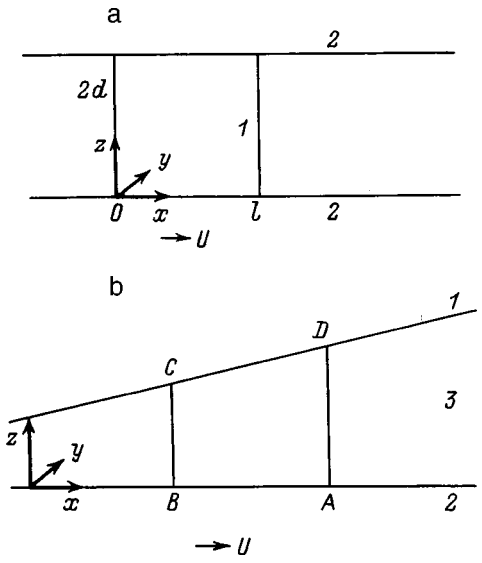

FIG. 1. Schematic of railgun accelerator (a) and gap between rails and projectile (b): *1* — projectile (armature), *2* — rails, and *3* — liquid film.

$$\frac{\partial B_a}{\partial t}=\frac{\partial}{\partial x}\left(D_a\frac{\partial B_a}{\partial x}\right)+\frac{\partial}{\partial z}\left(D_a\frac{\partial B_a}{\partial x}\right), \tag{1a}$$

$$\text{for}\quad x=0\quad B_a=0,\quad \text{for}\quad x=l\quad B_a=B_0,$$

$$\text{for}\quad z=d\qquad \frac{\partial B_a}{\partial z}=0, \tag{1b}$$

in the rails

$$\frac{\partial B_r}{\partial t}+U\frac{\partial B_r}{\partial x}=\frac{\partial}{\partial x}\left(D_r\frac{\partial B_r}{\partial x}\right)+\frac{\partial}{\partial z}\left(D_r\frac{\partial B_r}{\partial z}\right), \tag{2a}$$

$$\text{for}\quad x=0\quad B_r=0,\quad \text{for}\quad z=-\infty\quad B_r=0,$$

$$\text{for}\quad z=0\quad \text{and}\quad x\geqslant l\quad B_r=B_0, \tag{2b}$$

where B_0 is the magnetic induction in the space between the rails behind the accelerated body, which in this geometry depends only on time, d is half the distance between the rails, l is the projectile length, D is the diffusion coefficient of the magnetic field, and the subscripts a and r refer to the armature (projectile) and the rails, respectively.

Problems (1) and (2) must be supplemented by two coupling conditions at the rails–armature contact line (i.e., for $z=0$ and $0\leqslant x\leqslant l$), taking into account that a liquid conducting film is present in the gap between the rails and the armature. When deriving these conditions, we assumed that the film is so thin that any change in the magnetic induction over the film thickness can be neglected. This assumption holds as long as the thickness of the velocity skin layer is substantially greater than the thickness of the liquid film. To a first approximation, this last condition may be written in the form

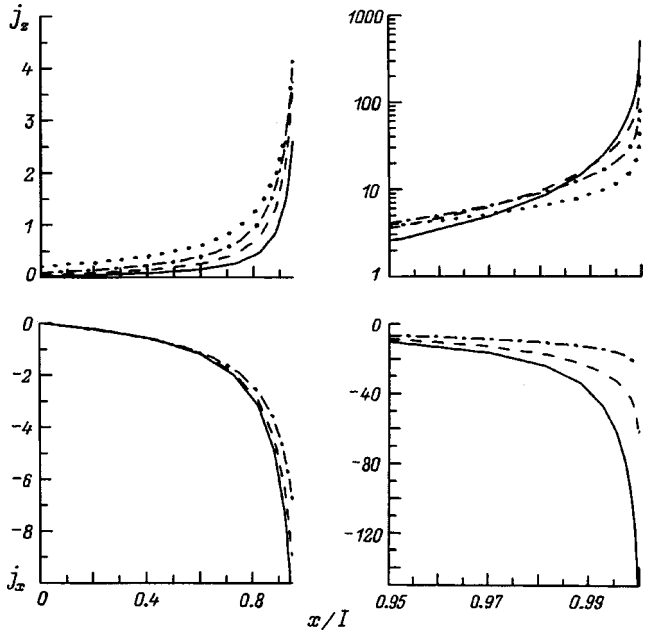

FIG. 2. Distribution of the x and z components of the electric current density at the rails–projectile interface without a liquid film (variant 1A): Re = 1000 (solid curves), 300 (dashed curves), 100 (dot-dash curves), and 30 (dotted curves).

$h \ll D_a^2/D_r U$, where h is the film thickness. Hence, the first condition at the rails–accelerated body interface has the form

$$\text{for} \quad z=0 \quad \text{and} \quad 0 \leqslant x \leqslant l \quad B_r = B_a . \tag{3}$$

We adopt the usual procedure to obtain the second condition. We integrate the Maxwell equation

$$\frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E}$$

over the region $ABCD$ in Fig. 2. Since

$$\int_{ABCD} (\nabla \times \mathbf{E}) ds = -\oint \oint_{ABCD} (E_z dz + E_x dx)$$

then, bearing in mind that the tangential component of the electric field at the rails–film and film–armature interfaces is continuous and that the integral $\int E_z dz$ over the section CD (film–armature interface) is nonzero and given by $-E_{z,a}(h(x_A)-h(x_B))$, we obtain

$$-(E_{x,r}-E_{x,a})dx + (h(x_A)-h(x_B))E_{z,a} + \int_0^{h(x_B)} E_{z,f} dz - \int_0^{h(x_A)} E_{z,f} dz = -\left(h \frac{\partial B}{\partial t} - B \frac{\partial h}{\partial t} \right) dx. \tag{4}$$

However, we find

$$E_{x,r} = \frac{j_{x,r}}{\sigma_r} = -D_r \frac{\partial B_r}{\partial z}, \quad E_{z,a} = \frac{j_{x,a}}{\sigma_a} = -D_a \frac{\partial B_a}{\partial z},$$

$$E_{z,a} = \frac{j_{z,a}}{\sigma_a} = D_a \frac{\partial B_a}{\partial x},$$

$$E_{z,f} = \frac{j_{z,f}}{\sigma_f} = -uB = D_f \frac{\partial B}{\partial x} - uB, \tag{5}$$

where the subscript f refers to the film, σ is the conductivity, $D = 1/\mu_0 \sigma$, and u is the flow rate of the liquid in the film.

Substituting Eqs. (5) into (4), dividing by dx, and letting the difference $|x_A - x_B|$ tend to zero, we have

$$D_r \frac{\partial B_r}{\partial z} - D_a \frac{\partial B_a}{\partial z} = \frac{\partial}{\partial x}\left(D_f h \frac{\partial B_a}{\partial x} \right) - Q \frac{\partial B_a}{\partial x} - D_a \frac{\partial h}{\partial x} \frac{\partial B_a}{\partial x} - h \frac{\partial B_a}{\partial t} + B_a \frac{\partial h}{\partial t}. \tag{6}$$

On the right-hand side of expression (6) the magnetic induction in the film will be replaced by the magnetic induction in the armature, since it is assumed that the change in the induction over the film thickness is small. The last two terms in Eq. (6) can be neglected, since the time taken for diffusion of the magnetic field in the film is substantially shorter that the characteristic time of variation of the magnetic field under the influence of external factors such as the acceleration of the armature. The third terms on the right-hand side of expression (6) can also be neglected, since it is assumed that the resistance of the film is much higher than that of the rails and the armature. If this were not the case, the influence of the film would be utterly insignificant.

Thus, the second condition at the rails–armature boundary finally has the form

$$\text{for} \quad z=0 \quad \text{and} \quad 0 \leqslant x \leqslant l$$

$$D_r \frac{\partial B_r}{\partial z} - D_a \frac{\partial B_a}{\partial z} = \frac{\partial}{\partial x}\left(D_f h \frac{\partial B_a}{\partial x} \right) - Q \frac{\partial B_a}{\partial x}, \tag{7}$$

where

$$Q(x) = \int_0^h u \, dz$$

is the liquid flow rate in the film.

Condition (7) differs from the conditions normally used at the rails–armature interface because of the second term on the right-hand side, which takes into account the velocity distribution in the cross section of the film.

Since the acceleration of the armature is fairly high, the problem (1)–(3), and (7) is generally nonsteady-state. Nevertheless, for our calculations of the magnetic field and electric current distribution we neglected the time derivatives, assuming the problem to be quasisteady. This assumption holds if the characteristic magnetic field diffusion time t_{diff} is substantially shorter than the characteristic time t_{mov} required for the velocity of the accelerated body to change. Estimates show that this condition is fairly well satisfied for the rails but is not generally satisfied in the armature. We have used this approximation anyway, for the following reasons. First, the main idea of the calculations was to study the influence of the liquid film on the motion of the body in the railgun accelerator. The use of a quasisteady approximation does not prevent us from addressing this problem, since it does not distort the character of this influence. Second, the quasisteady approach drastically simplifies the numerical solution

of the problem. In fact, in order to obtain a correct description of the velocity skin layer we need to use meshes which become extremely dense toward the rear edge of the armature. Thus, in the nonsteady-state approach it would be necessary to use nonsteady-state meshes whose density is increased as the problem is solved. Third and last, in the vast majority of studies dealing with the velocity skin effect, the velocity of the body has been assumed to be constant, and nonsteady effects have also been neglected. Therefore, in the next Section we consider the steady-state version of problem (1) and (2).

CONVERSION TO DIMENSIONLESS FORM

In Eqs. (1) and (2) and in the boundary conditions (3) and (7), we transform to dimensionless variables (the tilde will be omitted subsequently)

$$\tilde{B}=B/B_0, \quad \tilde{D}=D/D_*, \quad \tilde{h}=h/h_0, \quad \tilde{x}x=x/l,$$

$$\tilde{\mathbf{j}}=\mathbf{j}\frac{l}{\mu_0 B_0},$$

in the armature

$$\tilde{z}=z/l, \quad \tilde{j}_x=-\frac{\partial\tilde{B}}{\partial\tilde{z}}, \quad \tilde{j}_z=\frac{\partial\tilde{B}}{\partial\tilde{x}},$$

and in the rails

$$\tilde{z}=z\sqrt{\mathrm{Re}}/l, \quad \tilde{j}_x=-\sqrt{\mathrm{Re}}\,\frac{\partial\tilde{B}}{\partial\tilde{z}}, \quad \tilde{j}_z=\frac{\partial\tilde{B}}{\partial\tilde{x}}, \tag{8}$$

where Re is the magnetic Reynolds number $\mathrm{Re}=Ul/D_*$, h_0 is the height of the gap for $x=0$, and D_* is the typical value of the magnetic field diffusion coefficient.

As a result, we have in the armature

$$\frac{\partial}{\partial x}\left(D_a\frac{\partial B_a}{\partial x}\right)+\frac{\partial}{\partial z}\left(D_a\frac{\partial B_a}{\partial z}\right)=0, \tag{9a}$$

$$\text{for} \quad x=0 \quad B_a=0, \quad \text{for} \quad x=1 \quad B_a=1,$$

$$\text{for} \quad z=d \quad \frac{\partial B_a}{\partial z}=0, \tag{9b}$$

in the rails

$$\frac{\partial B_r}{\partial x}=\frac{1}{\mathrm{Re}}\frac{\partial}{\partial x}\left(D_r\frac{\partial B_r}{\partial x}\right)+\frac{\partial}{\partial z}\left(D_r\frac{\partial B_r}{\partial z}\right), \tag{10a}$$

$$\text{for} \quad x=0 \quad B_r=0, \quad \text{for} \quad z=-\infty \quad B_r=0,$$

$$\text{for} \quad z=0 \quad \text{and} \quad x\geqslant 1 \quad B_r=1, \tag{10b}$$

at the interface between the rails and the armature (for $z=0$ and $0\leqslant x\leqslant 1$)

$$B_r=B_a \qquad \text{and} \qquad \sqrt{\mathrm{Re}}D_r\frac{\partial B_r}{\partial z}-D_a\frac{\partial B_a}{\partial z}$$

$$=\frac{\partial}{\partial x}\left(\varepsilon_1 D_f h\frac{\partial B_a}{\partial x}\right)-\varepsilon_2\frac{\partial B_a}{\partial x}, \tag{11}$$

where $\varepsilon_1=h_0/l$ and $\varepsilon_2=Q/D_*$.

It should be noted that Eq. (11) not only describes the case when the rails and the armature are separated by a thin film but also cases where solid thin coatings are applied to the rails or the accelerated body. In the first case we must set $\varepsilon_2=\varepsilon_1\,\mathrm{Re}$ and in the second $\varepsilon_2=0$. For a liquid film the flow rate depends on the velocity and acceleration of the moving body, the shape of the gap between the rails and the armature, the viscosity of the lubricant, and the magnetic pressure. However, for fairly thin films of the order of a few microns thick the velocity profile in the cross section of the films is close to linear. In this case, a good approximation for ε_2 may be the expression $\varepsilon_2=\varepsilon_1\,\mathrm{Re}/2$, which is obtained if the liquid flow rate in the film is taken to be $Q=Uh/2$. It should be noted that for lubricants whose specific gravity is substantially higher than that of the accelerated body, this expression underestimates the flow rate and therefore overestimates the capacity of the film to suppress the velocity skin effect.

In order to solve Eq. (10A) we generally require another boundary condition for $x\to\infty$. However, since for large magnetic Reynolds numbers this equation has the characteristic features of a boundary-layer equation, this condition is in fact unnecessary for its numerical solution.

METHOD OF SOLUTION

The main feature of the problem (9)–(11) is that the solution of the Laplace equation (9) must be joined with the solution of Eq. (10), which contains the convective term. This has the result that at the point $x=1$, $z=0$ a singularity appears in the solution and a velocity skin layer forms around this point. For accurate calculations of this layer we shall seek a solution of the problem (10) in the form $B_r=B_1+B_2$, where the first term B_1 is determined by the boundary-layer equation

$$\frac{\partial B_1}{\partial x}=\frac{\partial}{\partial z}\left(D_r\frac{\partial B_1}{\partial z}\right) \tag{12a}$$

with the boundary conditions

$$\text{for} \quad x=0 \quad B_1=0, \quad \text{for} \quad z=-\infty \quad B_1=0,$$

$$\text{for} \quad z=0 \quad \text{and} \quad x\geqslant 1 \quad B_1=1, \tag{12b}$$

and the second term B_2 is given by

$$\frac{\partial B_2}{\partial x}=\frac{1}{\mathrm{Re}}\frac{\partial}{\partial x}\left(D_r\frac{\partial B_2}{\partial x}\right)+\frac{\partial}{\partial z}\left(D_r\frac{\partial B_2}{\partial z}\right)$$

$$+\frac{1}{\mathrm{Re}}\frac{\partial}{\partial x}\left(D_r\frac{\partial B_1}{\partial x}\right) \tag{13a}$$

with the boundary conditions

$$\text{for} \quad x=0 \quad B_2=0, \quad \text{for} \quad z=-\infty \quad B_2=0,$$

$$\text{for} \quad z=0 \quad B_2=0. \tag{13b}$$

In this case, the following conditions are satisfied at the rails–armature interface $z=0$ and $0\leqslant x\leqslant 1$:

$$B_1 = B_a \quad \text{and} \quad D_r\sqrt{\mathrm{Re}}\left(\frac{\partial B_1}{\partial z}+\frac{\partial B_2}{\partial z}\right)-D_a\frac{\partial B_a}{\partial z}$$

$$=\frac{\partial}{\partial x}\left(\varepsilon_1 D_f h\frac{\partial B_a}{\partial x}\right)-\varepsilon_2\frac{\partial B_a}{\partial x}. \tag{13c}$$

Problem (12) describes a boundary-layer solution, and problem (13) describes a correction to the boundary-layer solution near the point $x=1$, $z=0$, where the velocity skin layer forms, since the rate of change of the magnetic field near this point is of the same order along the x and z coordinates, and consequently, the second derivative with respect to x cannot be neglected near this point. Hence, Eqs. (9), (12), and (13) determine the magnetic field distribution in the rails and the moving body.

If the coefficient of magnetic field diffusion in the rails is assumed to be constant, we can write the following expression for the solution of problem (12):

$$B_1=\int_0^x \frac{\partial B_a}{\partial x}\bigg|_{z=0}\operatorname{erfc}\left(-\frac{z}{2\sqrt{D_r(x-x')}}\right)dx'$$

for $0\leqslant x\leqslant 1$, $z\leqslant 0$,

$$B_1=\int_0^1 \frac{\partial B_a}{\partial x}\bigg|_{z=0}\operatorname{erfc}\left(-\frac{z}{2\sqrt{D_r(x-x')}}\right)dx'$$

for $x>1$, $z\leqslant 0$. (14)

Condition (13C) then has the form

$$\sqrt{D_r\,\mathrm{Re}}\left(\frac{1}{\sqrt{\pi}}\int_0^x \frac{\partial B_a}{\partial x}\bigg|_{z=0}\frac{dx'}{\sqrt{x-x'}}+\frac{\partial B_2}{\partial z}\right)-D_a\frac{\partial B_a}{\partial z}$$

$$=\frac{\partial}{\partial x}\left(\varepsilon_1 D_f h\frac{\partial B_a}{\partial x}\right)-\varepsilon_2\frac{\partial B_a}{\partial x}. \tag{15}$$

If we neglect the correction B_2, calculation of the magnetic field in the armature can be reduced to solving the problem (9) with the condition (15) at $z=0$ (Ref. 2).

Equations (9), (12), and (13) or (9), (14), and (15) were integrated numerically by a finite-difference method. For accurate calculations of the velocity skin effect we used a nonuniform mesh with increasing density toward the boundaries $x=1$ and $z=0$. The mesh-point coordinates were calculated using formulas from Ref. 28. The minimum step was selected in the range 0.0001–0.0004 with 21 or 41 mesh points per unit length.

QUALITATIVE ANALYSIS OF THE INFLUENCE OF THE LIQUID FILM IN THE VELOCITY SKIN LAYER

Before analyzing the results of the numerical calculations, we shall assess the influence of the liquid film on the velocity skin layer using Eq. (15). We take δ to be the characteristic scale of variation of the magnetic field in the velocity skin layer in the armature (this dimension is the same along the x and y axes). Then, assuming that the film thickness is constant ($h=1$) and neglecting the correction B_2 in Eq. (15), we obtain

$$\frac{\sqrt{D_r\,\mathrm{Re}}}{\sqrt{\pi\delta}}-\frac{D_a}{\delta}=\frac{\varepsilon_1 D_f}{\delta^2}-\frac{\varepsilon_2}{\delta}. \tag{16a}$$

The left-hand side of Eq. (16A) vanishes for

$$\delta_1=\frac{\pi}{\mathrm{Re}}\frac{D_a^2}{D_r}. \tag{16b}$$

This value of δ determines the thickness of the velocity skin layer in the absence of a liquid film.[2,10] The right-hand side of Eq. (16A) vanishes for

$$\delta_2=\varepsilon_1 D_f/\varepsilon_2=2D_f/\mathrm{Re}. \tag{16c}$$

This implies, first, that the liquid film will increase the thickness of the velocity skin layer provided that $\pi D_a^2/D_r \leqslant 2D_f$. Second, it is easy to see that as the film thickness increases, i.e., as ε_1 increases, the thickness of the velocity skin layer tends toward δ_2. Let us compare this result with the case where the resistive layer is located on the surface of the armature. In this case, as has been noted, we find $\varepsilon_2=0$. Then, for fairly large ε_1 for which $\varepsilon_1 \gg D_a\delta/D_f$, we obtain the well-known expression[10]

$$\delta=\left(\frac{\pi D_f^2\varepsilon_1^2}{D_1\,\mathrm{Re}}\right)^{1/3}.$$

Thus, from the point of view of increasing the thickness of the velocity skin layer, a liquid layer is less effective than resistive coatings deposited on the surface of the armature.

The limiting value of the velocity skin-layer thickness was obtained assuming that D_f is fixed and the film thickness increases. If the film thickness is fixed and D_f tends to infinity, it follows from Eq. (15) that the magnetic field distribution in the armature will tend toward linear, $j_z\to 1$, and $j_x\to 0$. However, the magnetic field distribution in the rails will tend toward expression (14). Consequently, the x component of the electric field density in the rails at $z=0$ will be $j_x=2\sqrt{x\,\mathrm{Re}/\pi D}$ and can thus be fairly large. In this case, the main Joule dissipation will therefore be concentrated in the rails and not in the armature. It is also clear that each value of Re will correspond to a particular value of D_f for which this limiting magnetic field distribution is nearly attained. This value of D_f can be estimated using expression (16), with δ set equal to 1 or by seeking a solution as a series in reciprocal powers of εD_f. As a result, we have

$$\varepsilon_1 D_f > \frac{8}{3\sqrt{\pi}}\sqrt{D_r\,\mathrm{Re}}+\frac{\varepsilon_1}{2}\mathrm{Re}.$$

Thus if we take $D_*=D_a$, we find, for $D_r=0.7$ and $\mathrm{Re}=1000$ that the contact resistance of the film is forty times the resistance of the metal projectile (in terms of its height) per unit area.

RESULTS OF NUMERICAL CALCULATIONS

As we know, the conductivity of metals depends appreciably on temperature. However, in order to make the magnetic field calculations independent of the temperature distribution and thereby significantly simplify the problem, the temperature dependence of the diffusion coefficients was ne-

TABLE I.

Variant	Film thickness ε_1	D_f	Dependence of film thickness on x
1	0		
2	0.0001	7.2	No
3	0.001	7.2	-‘‘-
4	0.001	30	-’’-
5	0.001	100	-‘‘-
6	0.01	100	-’’-

glected in the present study. The projectile was assumed to be aluminum and the rails copper. The diffusion coefficients of aluminum and copper were taken to be their values at 600 °C, i.e., $0.049\,m^2/s$ for aluminum and $0.035\,m^2/s$ for copper. The coefficient of magnetic field diffusion in the rails used was D_*. The diffusion coefficients in the rails and the liquid film were always assumed to be constant, while the diffusion coefficient in the armature was assumed to be piecewise-constant so that we could estimate the influence of resistive inserts in the armature. Thus, the dimensionless diffusion coefficients in the rails and the armature are $D_r=1$ and $D_a=1.4$, and a hypothetical diffusion coefficient of 14 was used for the resistive inserts in the armature. Three types of armature were considered: uniform aluminum with $D_a=1.4$ (A), having a resistive layer (B)

$$D_a=\begin{cases}1.4, & \text{at} \quad d/2>z>0.01,\\ 14 & \text{at} \quad z<0.01;\end{cases}$$

and having a resistive insert in the rear part of the armature (C)

$$D_a=\begin{cases}1.4, & \text{at} \quad x<1-\sqrt{3}z,\\ 14, & \text{at} \quad x>1-\sqrt{3}z.\end{cases}$$

This last variant simulates a chevron armature for which the rear walls form an angle of 30° with the surface of the rails. In all three variants the armature was assumed to be square. The film material was liquid indium with a diffusion coefficient of $0.253\,m^2/s$ (dimensionless value $D_f=7.2$). However, we also used hypothetical values of the film diffusion coefficient substantially higher than this value and unrelated to any specific material. All the films considered are listed in Table I.

TABLE II. Maximum z component of current density and thickness of velocity skin layer for a uniform armature.

Variant	Reynolds number							
	30		100		300		1000	
	$j_{z,\max}$	δ_{VSL}	$j_{z,\max}$	δ_{VSL}	$j_{z,\max}$	δ_{VSL}	$j_{z,\max}$	δ_{VSL}
1A	30.9	0.114	64.7	0.046	121	0.019	245	0.0065
2A	28.1	0.115	56.5	0.046	108	0.019	219	0.0067
3A	17.6	0.116	34.3	0.049	64.7	0.021	131	0.008
4A			20.3	0.058	35.4	0.029	65.2	0.014
5A			11.9	0.08	19.3	0.046	33.1	0.024
6A			3.9	0.19	6.1	0.124	11.3	0.066

TABLE III. Maximum z component of current density and thickness of velocity skin layer for an armature with a resistive layer (Re = 1000).

Variant	$j_{z,\max}$	δ_{VSL}
3B	33.1	0.039
4B	26.5	0.041
5B	19.5	0.047
6B	10.0	0.074

The calculations were made for four magnetic Reynolds numbers 30, 100, 300, and 1000. If the projectile length is 0.01 m, these Reynolds numbers correspond to projectile velocities of 105, 350, 1050, and 350 m/s. Note that in all these variants the constant resistance of the film was substantially lower than the value at which the limiting magnetic field distribution is established in the armature and the rails. Two parameters were used for the nonuniformity characteristic of the electric current density distribution: the maximum z component of the current $j_{z,\max}$ and the extent of that part of the rails–armature interface δ_{VSL} across which 50% of the electric current flows. This last value may be taken as an estimate of the velocity skin-layer thickness. Table II shows that in the absence of a film this value is almost the same as the estimated thickness of the velocity skin-layer (16B) for magnetic Reynolds numbers of 300 and 1000.

In this geometry the maximum z component of the current always occurs at the point $(x=1,\ z=0)$. However, it was found that the value of j_z at this point depends strongly on the mesh step. It is possible that in our particular model this value is in general infinite. At least, we were unable to determine this with a sufficient degree of reliability for any reasonable crowding of the mesh. Thus, for $j_{z,\max}$ we took the value of j_z at the rails–armature interface at the point $x=0.9997$. On the one hand, this point is fairly close to $x=1$ and in this range of parameters is always inside the velocity skin layer and on the other, the current at this point can be calculated with a sufficient degree of reliability. Obviously this choice of $j_{z,\max}$ is somewhat arbitrary but it can nevertheless effectively demonstrate the influence of the film thickness on the electric current density distribution.

The calculations have shown that near the rear edge of the armature j_z is substantially higher than the x component of the electric current density in all variants except the armature with a resistive insert (variant C). In this last case, since the diffusion coefficient in the armature undergoes a jump

TABLE IV. Maximum z component of the current density and thickness of velocity skin layer for an armature with a resistive insert in the rear section.

Variant	Reynolds numbers			
	100		1000	
	$j_{z,\max}$	δ_{VSL}	$j_{z,\max}$	δ_{VSL}
1C	22.4	0.18		
3C	8.9	0.18	44.5	0.039
4C	6.8	0.18	30.5	0.043
5C	5.3	0.18	19.6	0.052
6C	2.8	0.27	9.1	0.087

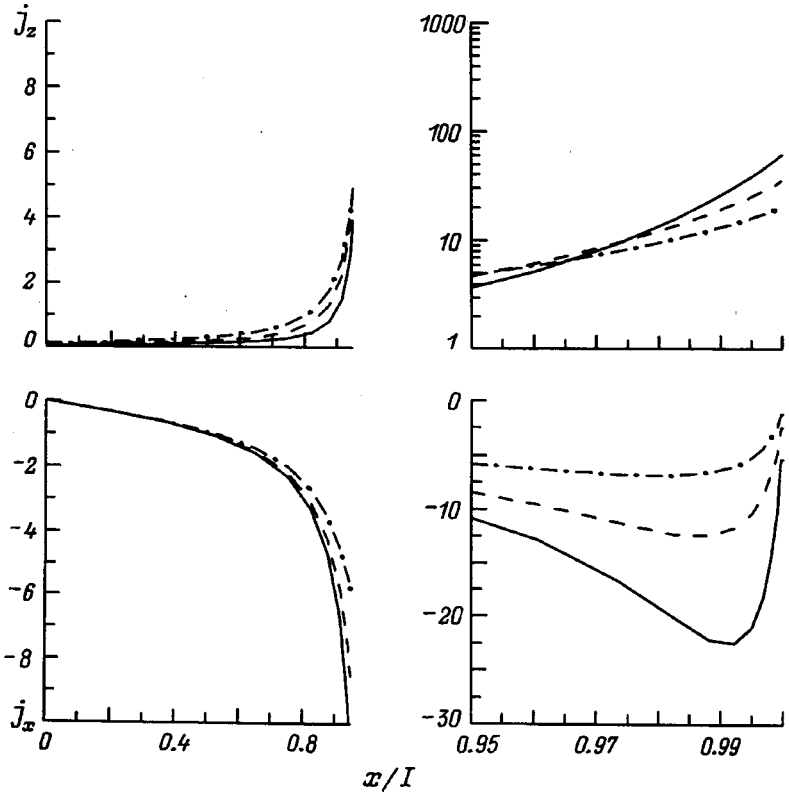

FIG. 3. Distribution of the x and z components of the electric current density at the rails–armature interface with a liquid film (variant 4A): $\varepsilon_1 = 0.001$, $D_f = 30$; Re = 1000 (solid curves), 300 (dashed curves), and 100 (dot-dash curves).

along an interface inclined at an angle of 30° to the surface of the rails, the x component of the current becomes very appreciable and may even exceed the z component. Thus, for this variant $j_{z,\max}$ is not a sufficiently representative characteristic of the nonuniformity of the current density distribution. The values of $j_{z,\max}$ and δ_{VSL} for the different variants are given in Tables II–IV, and Figs. 3 and 4 show the distributions of the x and z components of the electric current density at the rails–armature contact interface. The x component of the electric current density shown in these figures was calculated from the armature side. However, the x component of the current density at the contact interface on the

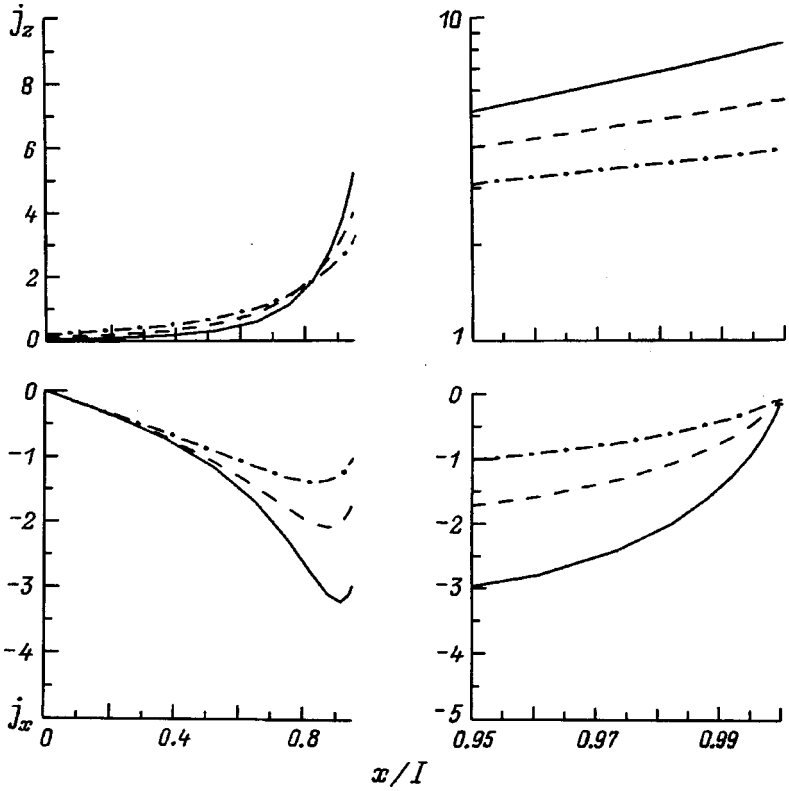

FIG. 4. Distribution of the x and z components of the electric current density at the rails–armature interface with a liquid film (variant 6A): $\varepsilon_1 = 0.01$, $D_f = 100$; Re = 1000 (solid curves), 300 (dashed curves), and 100 (dot-dash curves).

TABLE V. Maximum x component of the current density in the rails at the interface with the armature.

Variant	Re = 300	Re = 100
1A	92.3	215
2A	112	258
3A	115	269
4A	94.2	215
5A	73.2	164
6A	37.7	102

rails side is much greater. In the absence of any film this may be because copper has a higher conductivity than aluminum. The presence of a resistive film increases this discontinuity still further. In the limit that the contact resistance of the film tends to infinity, as we showed in the previous section, j_x will tend toward zero in the armature and toward $2\sqrt{x\,\mathrm{Re}/\pi}$ in the rails. This is, as it were, in compensation for the reduction in the velocity skin effect. Table V gives the maximum x component of the current density $j_{z,\max}$ in the rails at the interface with the projectile. It can be seen that as the contact resistance of the film increases, the maximum x component of the current density in the rails initially increases as a result of the influence of the film and then begins to decrease as a result of a general reduction in the current density caused by the suppression of the velocity skin effect.

These results allow us to draw the following conclusions.

In the absence of a film, the maximum current density increases substantially more slowly with increasing magnetic Reynolds number than the directly proportional relationship predicted from formula (16A) for the thickness of the velocity skin layer.

The influence of the film is primarily observed as a reduction in the maximum current density. Moreover, the skin-layer thickness (the value of δ_{VSL}) increases far more slowly as the contact resistance of the film increases.

A combination of a liquid film and resistive inserts in the armature can significantly suppress the velocity skin effect. Nevertheless, even when fairly thick films are used with a fairly high resistivity, the nonuniformity of the electric current density distribution in the armature for the Reynolds numbers of practical interest is still very high. For instance, in variant 6A the thickness of the velocity skin layer is only 7% of the armature length, although the contact resistance of the film is 1.5 times the resistance of the armature (in terms of its height) per unit length. In this case, for a current of 300 kA and a projectile size of 0.01 m the voltage drop across the liquid film will be 9 V. In order to completely eliminate the velocity skin effect at high armature velocities the contact resistance of the film should be ten times the resistance of the projectile. In this case, the voltage drop across the film will be tens of volts.

This work was partly supported by DERA (Contract No. SMC/4C2061).

[1] G. C. Long, IEEE Trans. Magn. **MAG-25**, 347 (1989).
[2] P. B. Parks, J. Appl. Phys. **67**, 3511 (1990).
[3] L. E. Thurmond, B. K. Ahrens, and J. P. Barber, IEEE Trans. Magn. **MAG-27**, 326 (1991).
[4] T. E. James, IEEE Trans. Magn. **MAG-27**, 482 (1991).
[5] B. A. Uryukov, A. D. Lebedev, and K. K. Milyaev, in *Proceedings of the Second All-Union Seminar on Dynamics of a High-Current Arc Discharge in a Magnetic Field* [in Russian], Novosibirsk (1991) pp. 33–71.
[6] V. P. Bazilevskiĭ, R. M. Zayatdinov, and Yu. A. Kareev, *ibid.*, pp. 285–303.
[7] A. P. Glinov, N. M. Kolyadin, A. E. Poltanov *et al.*, *ibid.*, pp. 315–339.
[8] V. S. Yuferev, M. L. Gnedina, and N. Yu. Gnedin, Zh. Tekh. Fiz **62**(1), 83 (1992) [Sov. Phys. Tech. Phys. **37**, 41 (1992)].
[9] Y. A. Dreizin, IEEE Trans. Magn. **MAG-29**, 798 (1993).
[10] J. P. Barber and Y. A. Dreizin, IEEE Trans. Magn. **MAG-31**, 96 (1995).
[11] T. E. James and D. C. James, IEEE Trans. Magn. **MAG-31**, 162 (1995).
[12] Yu. A. Kareev and R. M. Zayatdinov, IEEE Trans. Magn. **MAG-31**, 180 (1995).
[13] T. E. James, IEEE Trans. Magn. **MAG-31**, 622 (1995).
[14] J. D. Powell and A. E. Zielinski, IEEE Trans. Magn. **MAG-31**, 645 (1995).
[15] A. J. Schoolderman, IEEE Trans. Magn. **MAG-31**, 651 (1995).
[16] M. Angeli, E. Gardelli, and B. Azzerboni, IEEE Trans. Magn. **MAG-33**, 37 (1997).
[17] T. E. James and D. C. James, IEEE Trans. Magn. **MAG-33**, 86 (1997).
[18] L. C. Woods, IEEE Trans. Magn. **MAG-33**, 152 (1997).
[19] J. P. Barber, A. Challita, B. Maas, and L. E. Thurmond, IEEE Trans. Magn. **MAG-27**, 228 (1991).
[20] J. P. Barber and A. Challita, IEEE Trans. Magn. **MAG-29**, 733 (1993).
[21] I. S. Glushkov, Yu. A. Kareev, L. G. Kotova *et al.*, IEEE Trans. Magn. **MAG-33**, 549 (1997).
[22] K. T. Hsieh and B. K. Kim, IEEE Trans. Magn. **MAG-33**(1), 237 (1997).
[23] D. J. Hildenbrand, J. R. Rapka, and B. J. Long, IEEE Trans. Magn. **MAG-33**(1), 74 (1997).
[24] B. A. Uryukov, A. D. Lebedev, and C. C. Milyaev, in *Proceedings of the Fourth European Symposium on Electromagnetic Launch Technology*, Gelle, Germany, 1993, Paper No. P1511.
[25] R. Crawford, J. Taylor, and D. Keefer, IEEE Trans. Magn. **MAG-31**, 138 (1997).
[26] C. Persad, A. Yeoh, G. Prabhu *et al.*, IEEE Trans. Magn. **MAG-33**, 140 (1997).
[27] S. R. Kerwien, IEEE Trans. Magn. **MAG-33**, 104 (1997).
[28] C. J. Quon, Fluid Mech. **181**, 233 (1987).

Translated by R. M. Durham

Characteristics of a thin-film sensor for a scanning SQUID microscope

D. E. Kirichenko, A. B. Pavolotskiĭ, I. G. Prokhorova, and O. V. Snigirev

M. V. Lomosonov Moscow State University, 119899 Moscow, Russia
(Submitted July 30, 1998)
Zh. Tekh. Fiz. **69**, 112–117 (July 1999)

A description is given of the design, fabrication technology, and characteristics of a sensor for a scanning magnetic microscope using a thin-film dc SQUID with $Nb/Al_2O_3/Nb$ shunted Josephson tunnel junctions. It is shown that at a sample temperature of 4.2 K the spatial resolution of this detector is 10 μm with a field resolution of 70 $pT/Hz^{1/2}$. © *1999 American Institute of Physics.* [S1063-7842(99)02307-7]

INTRODUCTION

Scanning magnetic microscopes using thin-film dc SQUIDS with $Nb/Al_2O_3/Nb$ shunted Josephson tunnel junctions are being increasingly widely used in unique physics experiments (see Refs. 1 and 2, for example). The spatial resolution δx of these devices, currently 1–2 μm, is determined by the larger of two quantities: first, the distance Z between the object and the pick-up coil, and second, the linear dimension D of the pick-up coil. The resolution for the recorded component of the magnetic induction δB_z is determined by the spectral density of the equivalent noise flux $S_{\Phi n}$ in the measuring frequency band and is inversely proportional to the area A of the pick-up coil.

Since the signal and noise characteristics of the sensor of a scanning magnetic microscope depend on the values and relationships between the pick-up coil inductance L_p, the inductance of the main part of the SQUID quantization circuit containing the Josephson junctions L_s, and the inductance of the connecting microstripline L_{str}, and also on various other parameters, the problem of designing and optimizing the multilayer thin-film structure of a SQUID sensor to give selected values of δx and δB_x is of separate interest.

Here we describe the design, fabrication technology, and characteristics of the sensor for a scanning SQUID microscope designed to operate at 4.2 K with a spatial resolution of the order of 10 μm and a field resolution of 70 $pT/Hz^{1/2}$.

DESCRIPTION OF THE SENSOR OF THE SCANNING SQUID MICROSCOPE

A schematic of the sensor, which is similar to that used in Refs. 1 and 3, is shown in Fig. 1a. In this system a pick-up coil *LP* of the required size is positioned some distance from the main SQUID circuit containing the Josephson junctions *JJ* and is connected to it by a low-inductance strip structure. This design can minimize the distortion of the magnetic field of the sample by the superconducting parts of the SQUID integrated microcircuit.

In addition to Josephson junctions *JJ*, the main part of the SQUID contains shunting resistors *RS*, a feedback coil, a damping resistor *RD*, and symmetric power input resistors *RB*. Figure 1b shows a diagram of the integrated SQUID microscope sensor. The sensor is fabricated by multilayer technology and has two layers of main superconducting metallization with a minimum reproducible size of 2.5 μm, and 3.5×3.5 μm planar $Nb/Al_2O_3/Nb$ Josephson junctions. The sensor is mounted on a 3×7 mm rectangular silicon substrate.

A particular feature of the system is the two resistors *RB* designed for a symmetric input current supply. This method of coupling in the supply current allows the pick-up coil to operate with almost zero dc current and can substantially reduce the back-effect influence of the SQUID on the object under study.

In order to achieve a spatial resolution for the SQUID microscope close to the effective size of the pick-up coil, the distance between the coil and the surface of the sample must be of the order of or smaller than the size of this coil. For this purpose the silicon substrate of the chip with the SQUID sensor was tapered near the pick-up coil (Fig. 1b). In the scanning microscope the chip with the sensor is attached to a spring-loaded cantilever at a small angle to the surface of the sample and the pointed corner of the chip glides directly over the surface (Fig. 1c).

During operation of the scanning microscope a magnetic flux considerably greater than the flux being measured may be trapped (or frozen during cooling) in the inductive circuit of the SQUID sensor. A local heater *RTS* is provided to remove it from the SQUID circuit without extracting the entire device from the liquid helium. The heater is a resistive ''snake''positioned above part of the SQUID circuit. When current flows through the heater, this part of the circuit is locally overheated above the critical temperature and loses its superconducting properties, thereby releasing the trapped magnetic flux from the circuit.

FABRICATION TECHNOLOGY

The sensor microcircuit is fabricated on a {100}-oriented single-crystal unoxidized silicon substrate and includes the following layers: 1) an Al_2O_3 insulating sublayer between the substrate and the auxiliary superconducting metallization, 2) auxiliary superconducting Nb metallization, 3) the lower Nb electrode of the $Nb/AlO_x/Nb$ triple layer, 4) the upper Nb electrode of the triple layer, 5) two layers of Al_2O_3 insula-

 © 1999 American Institute of Physics

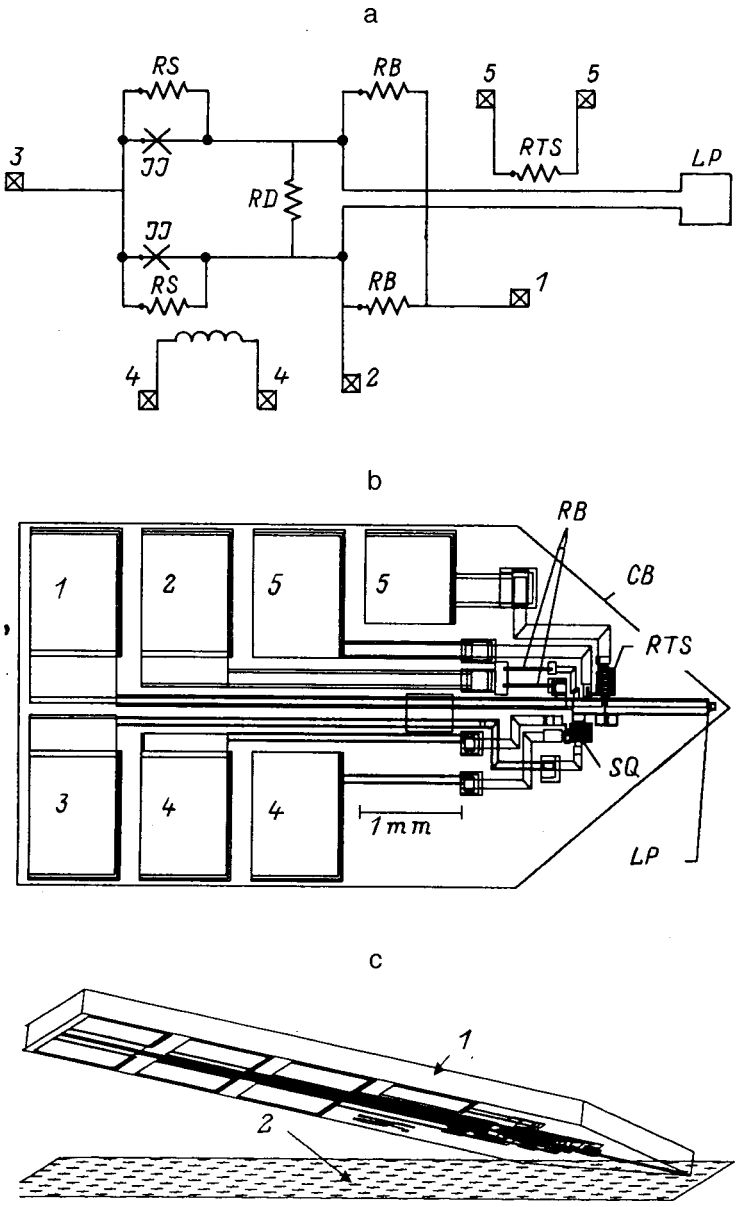

FIG. 1. Sensor for SQUID microscope: a — equivalent circuit diagram of sensor, *1* — SQUID current in, *2* — SQUID voltage out, *3* — common output, *4* — outputs of SQUID feedback coils, *5* — outputs of local heater; b — schematic of sensor: *LP* — pick-up coil (40×40 μm variant), *SQ* — main part of SQUID with Josephson junctions *JJ*, resistors *RS* and *RD*, and feedback coil, *RTS* — local heater resistor, *RD* — symmetric input power resistors, *CB* — edge of sensor substrate; the contact areas are numbered as in Fig. 1a; c — schematic configuration of sensor and sample during scanning: *1* — sensor, *2* — surface of sample (view from surface of sample).

tion between the lower electrode and the main superconducting metallization, 6) a layer of Ti resistors, 7) the main superconducting Nb metallization, and 8) a layer of Ti/Au contact pads.

The x-ray amorphous Al_2O_3 sublayer plays the role of a barrier layer which protects the substrate during the subsequent etching of the layer of auxiliary metallization. The aluminum oxide was deposited by rf magnetron sputtering in a mixture of Ar and O_2 using a Leybold AG Z-400 system. The Ar pressure was 1 Pa and the oxygen pressure 0.25 Pa. The thickness of the deposited insulator was 150 nm. Lift-off photolithography was then used to form windows in the sublayer to provide electrical contact between the layer of auxiliary metallization and the silicon substrate, which is required for the subsequent anodization of the lower electrode.

The auxiliary superconducting metallization, also used for the electrical connections between the lower electrode and other elements of the circuit, was a 200 nm thick layer of Nb. The Nb was deposited by dc magnetron sputtering at an Ar pressure of 1×10^{-2} mbar and 2.5 A discharge current using a Leybold AG L-560 system. The rate of deposition was 140 nm per minute. During pump-down prior to the deposition of the Nb, a Meissner trap was cooled with liquid nitrogen. This resulted in a residual pressure of the order of 2×10^{-7} mbar in the chamber. The substrates were attached to a water-cooled holder, and to improve the thermal contact they were stuck on with vacuum grease.

After the Nb film had been deposited, a mask was formed on it by photolithography for pattern etching. This was carried out by reactive ion etching in SF_6 at 20 norm.cm^3/s at 10 Pa pressure and 50 W discharge power. The end of etching was determined from the change in the coefficient of reflection of laser radiation from the surface of the film.

As we know, in order to obtain high-quality tunnel junctions with niobium electrodes, we need to fabricate *in situ* a sequence comprising a lower niobium electrode, a thin aluminum layer which is deposited thermally and then oxidized

to form a tunnel barrier, and finally the upper niobium electrode. We deposited a lower niobium electrode 200 nm thick (using the same deposition parameters as for the layer of auxiliary superconducting metallization). Then, after a 30 min pause to stabilize the substrate temperature, we deposited 6 nm of aluminum by rf magnetron sputtering at an Ar pressure of 1×10^{-2} mbar and 30 W power. The rate of deposition under these conditions was of the order of 9 nm/min. As a result of the specific growth mechanism, the aluminum film smooths out growth roughness on the niobium surface and thereby creates conditions for fabricating a homogeneous layer of thermal oxide which functions as the tunnel barrier. We carried out oxidation in pure oxygen at a pressure of 4 mbar for 25 min and then deposited a 100 nm thick niobium upper electrode.

The pattern in the lower electrode was produced by reactive ion etching carried out in three stages. First in an SF_6 plasma (20 norm.cm^3/s, 10 Pa, 50 W) we etched the upper niobium layer as far as the barrier layer, at which the etching naturally terminated. The pressure was then reduced to 1 Pa and the power increased to 100 W, which allowed us to etch through the barrier layer and the layer of residual aluminum (the point at which the etching of these layers ended was determined using a laser interferometer). The pressure and power were then returned to their initial levels and the remaining lower electrode was etched.

In the upper electrode of the triple layer we then formed a pattern for Josephson junctions in a ''column'' 0.1 μm high and 3.5×3.5 μm in cross section. This operation was performed by reactive ion etching in an SF_6 plasma (20 norm.cm^3/s, 10 Pa, 50 W) until a natural stop was reached at the barrier layer.

We then fabricated the insulation layers between the lower electrode and the main superconducting metallization. The first layer of insulation in fact consists of two layers: a layer of anodic niobium oxide and a deposited Al_2O_3 layer. This was formed using the same photoresist mask as that used to etch the upper electrode (self-matched process). The side surfaces of the Josephson junctions and the open surface of the lower triple-layer electrode were liquid anodized through the photomask left from the previous etching operation at a voltage up to 20 V with a rate of rise of 1 V/s. The same photoresist mask was then used to deposit the x-ray amorphous Al_2O_3. The thickness of the deposited insulator was 200 nm. A negative mask was then prepared for the second layer of insulation, which differed from the first to reduce the probability of shorting through the two layers of insulation, and a second Al_2O_3 layer 200 nm thick was deposited through this negative mask.

The shunting, load, and damping resistors for the circuit were fabricated by rf magnetron sputtering of titanium through a negative photoresist mask. The titanium was deposited using a Leybold AG Z-400 system at an Ar pressure of 1 Pa. The thickness of the deposited metal was 130 nm and the resultant resistance was 1 $\Omega/\square$.

The main superconducting metallization connecting all the components of the sensor circuit electrically was a 200-nm-thick layer of Nb deposited by the lift-off technique using a negative photomask.

TABLE I.

Sensor No.	D, μm	I_C, μA	R_N, Ω	V_C, μV	R_S/R_D,	L_S, pH	ΔV, μV	$\Delta V/V_C$
1	10×10	45	3.3	150	1	60	66	0.43
2	20×20	40	2.2	70	1	80	31	0.43

The contact areas required for the connections to the terminals of the measuring system were formed by depositing a 50-nm-thick titanium sublayer via a negative photomask to achieve the required adhesion of a 100-nm-thick gold layer which was deposited next.

RESULTS OF MEASUREMENTS AND ELECTRODYNAMIC MODEL

Table I gives the main parameters of two sensors studied with different sizes of pick-up coils. Here D is the size of the pick-up coil, I_C, R_N, V_C, and L_S are the maximum critical current, the normal resistance, the characteristic voltage, and the inductance of the SQUID circuit, ΔV is the percent modulation of the field–voltage characteristic of the SQUID, and R_S/R_D is the ratio of the resistances of the shunting and damping resistors. The experimental current–voltage and field–voltage characteristics of these sensors are plotted in Fig. 2. These clearly show characteristic features in the form of steps at different voltages, denoted by *1–3* in Fig. 2a.

We know that these steps are associated with the excitation of Josephson oscillations at a frequency corresponding to a particular voltage, in resonance with oscillations in the SQUID microstrip structures which comprise high-Q cavities. The appearance of these oscillations leads to various undesirable consequences and particularly an appreciable increase in the SQUID noise level near these steps.

In practical devices these resonances are partially or completely eliminated by introducing and selecting damping resistors. A comparison of the sensor characteristics shows that a reduction in the resistance of the shunting and damping resistors in sensor No. 2 as compared with sensor No. 1 smoothed the field–voltage characteristics of the sensor, although it did not completely suppress the resonances.

The flux noise spectrum of sensor No. 1 measured using a SQUID preamplifier[4] is shown in Fig. 3 and corresponds to 3.5 $\mu\Phi_0/\mathrm{Hz}^{1/2}$. This noise level ensures good resolution (70 pT/Hz$^{1/2}$) for the component of the magnetic field recorded using a 10×10 μm pick-up coil, although it is considerably higher than the known estimate[5] of the white noise level.

Since one possible cause of the excess sensor noise may be that the resonances are not completely suppressed, we made a more detailed study of these resonance properties. Enpuku *et al.*[6,7] developed an electrodynamic model to study the resonance properties of SQUIDS. This model describes the frequency dependence of the circuit impedance for SQUIDS of conventional configuration with well-coupled multiturn planar input coils. The relative simplicity of the electrodynamic system of this sensor as compared with conventional dc SQUIDS means that a simple model can be

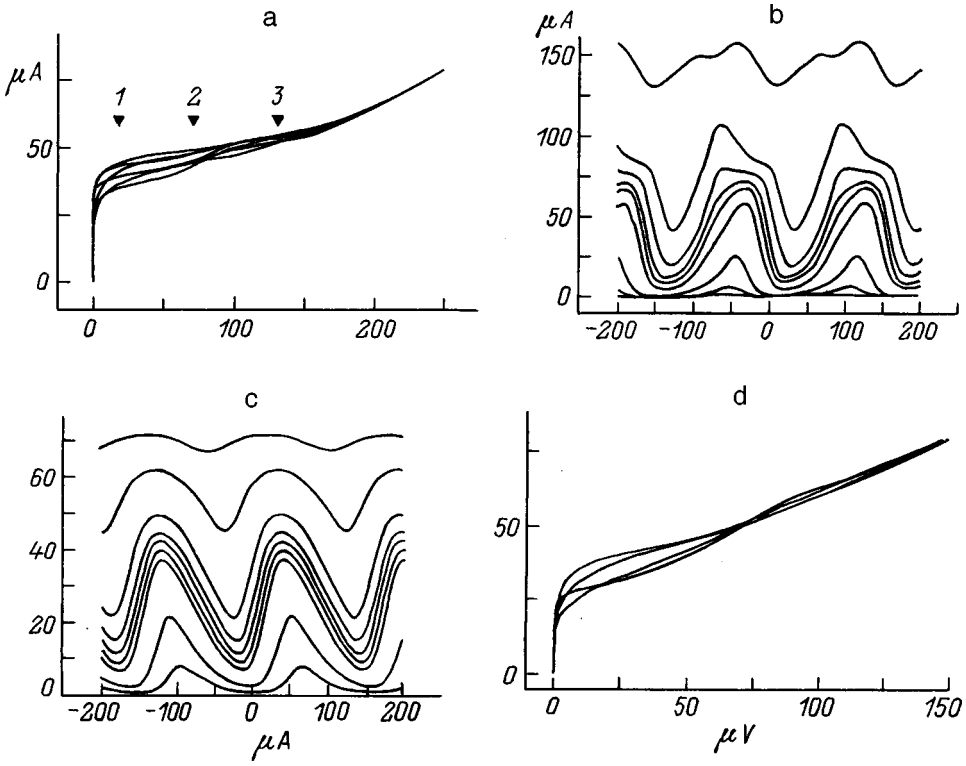

FIG. 2. Current–voltage (*1*) and field–voltage (*2*, *3*) characteristics of sensors *1* (a,b) and *2* (c,d).

used to estimate the resonance properties of its inductive circuit, including the strip structure connecting the pick-up coil and the main part of the SQUID.

Figure 4a shows the equivalent electric circuit of the model. Here the lumped elements are components of the SQUID circuit whose dimensions are much smaller that the wavelength, at least for frequencies up to 500 GHz, which correspond to a voltage of up to 1 mV at the Josephson junctions (much higher than the working voltage). These elements are the inductances of the pick-up coil *LP*, the slit in the main part of the SQUID *LS*, and the tapering in the strip structure for efficient operation of the local heater *LTS*, and also the damping resistor *RD*. Two sections of the strip structure are represented by the transmission line sections $TL1$ and $TL2$ having different lengths and wave impedances. On the basis of the sensor geometry, the values of the circuit components were estimated as follows: $LS=30$ pH, $LTS=2$ pH, $LP=16$ pH (for a 10×10 μm coil), and $RD=4\ \Omega$.

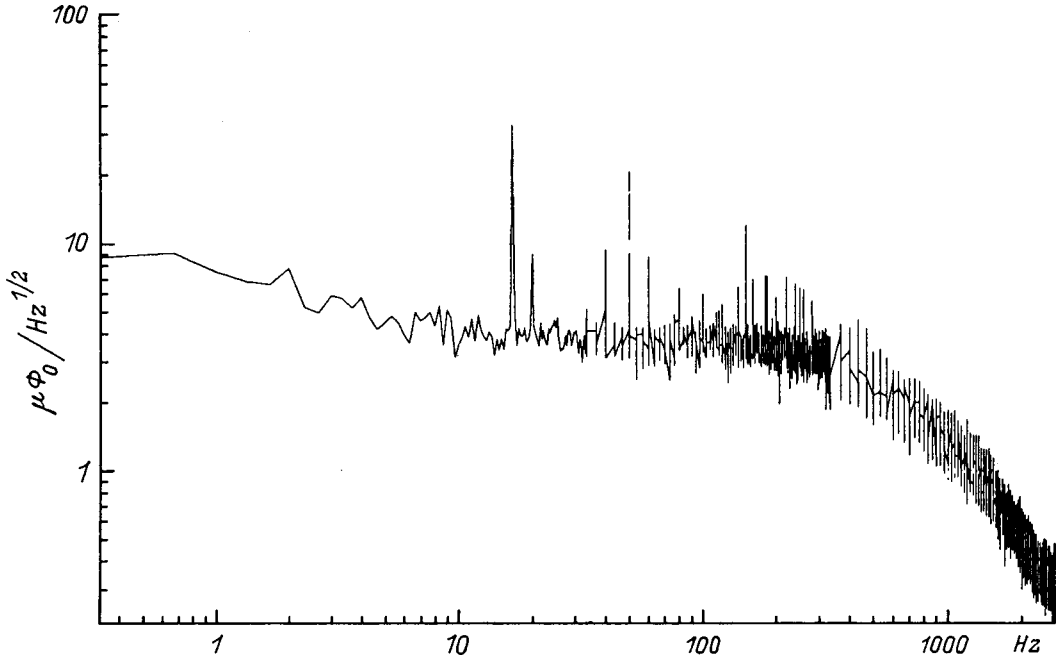

FIG. 3. Spectrum of equivalent noise flux of SQUID sensor (*1*).

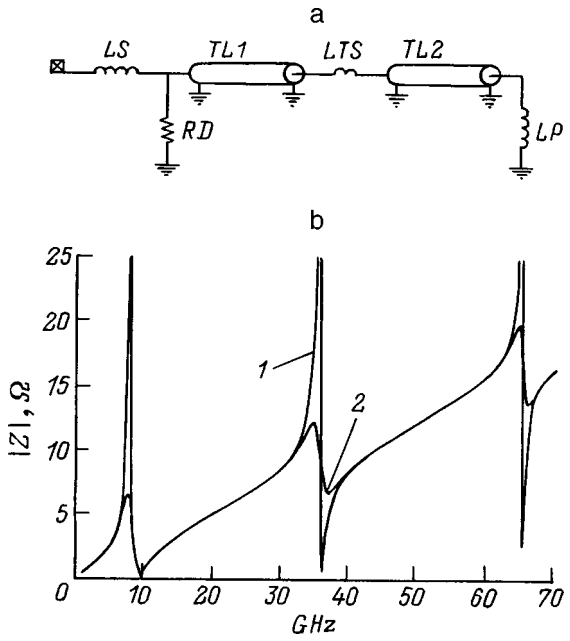

FIG. 4. Calculation of resonances of SQUID microscope sensor: a — equivalent circuit of model used to calculate impedance of SQUID circuit; b — frequency dependence of SQUID circuit impedance (RD = 100 (*1*) and 6 Ω (*2*)).

The characteristics of the transmission lines were determined in accordance with conventional formulas for a niobium superconducting stripline in the absence of losses,

$$Z_0 = 120\,pi\,\frac{d}{w}\sqrt{\frac{1+2\lambda/d}{\varepsilon_r}}, \quad c = \frac{c_0}{\sqrt{\varepsilon_r(1+2\lambda/d)}},$$

where Z_0 is the wave impedance, c is the speed of light in the line, c_0 is the speed of light in vacuum, $\lambda = 85$ nm is the depth of penetration of the magnetic field in the superconductor, $\varepsilon_r = 8$ is the permittivity of the insulator between the strips, $d = 400$ nm is the insulator thickness, and $w = 110\,\mu$m is the strip width.

Figure 4b gives results of calculating the impedance of the SQUID circuit for two different resistances of the damping resistor *RD*. The curves clearly show the resonant structure of the frequency dependence of the circuit impedance. The frequencies of the first three resonances were estimated as 9, 36, and 66 GHz, which are equivalent to Josephson junction voltages of 18, 72, and 132 μV. Their position corresponds to the numbers *1–3* in Fig. 2a. It can be seen that the resonances near *2* and *3* do in fact occur. Resonance *1* near 18 μV is possible but is insignificant because of its closeness to the superconducting branch and the appreciable broadening of the current–voltage characteristic. Thus, these calculations allowed us to identify resonances and study the influence of the shunting resistors on their Q-factor.

CONCLUSIONS

A sensor for a magnetic SQUID microscope has been developed and investigated experimentally. A $10\times10\,\mu$m pick-up coil and a sensor noise level 3.5 $\mu\Phi_0/\mathrm{Hz}^{1/2}$ were obtained, giving a resolution of 70 pT/Hz$^{1/2}$ for the recorded magnetic field component. An electrodynamic model of the sensor quantization circuit yielded an estimate of the resonance behavior of the sensor in satisfactory agreement with the measured results.

This work was supported by the Russian Fund for Fundamental Research (Project No. 96-02-19250) and by the ''InterSQUID'' Project of the Ministry of Science and Technology of the Russian Federation.

We are grateful to S. N. Polyakov, S. A. Gudoshnikov, and Yu. V. Maslennikov for assistance at various stages of this study.

[1] M. Jeffery, T. Van Duzer, J. R. Kirtley, and M. B. Ketchen, Appl. Phys. Lett. **67**, 1769 (1995).
[2] C. C. Tsuei, J. R. Kirtley, M. Rupp *et al.*, Science **271**, 329 (1996).
[3] M. B. Ketchen, J. R. Kirtley, and M. Bhushan, IEEE Trans. Appl. Supercond. **AS-7**, 3139 (1997).
[4] D. E. Kirichenko, A. B. Pavolotskij, I. G. Prokhorova *et al.*, Inst. Phys. Conf. Ser. **158**, 727 (1997).
[5] C. D. Tesche and J. Clarke, J. Low Temp. Phys. **29**, 301 (1977).
[6] K. Enpuku and H. Koch, Jpn. J. Appl. Phys. **32**, 3811 (1993).
[7] K. Enpuku, R. Cantor, and H. Koch, J. Appl. Phys. **71**, 2338 (1992).

Translated by R. M. Durham

BRIEF COMMUNICATIONS

Efficiency of successive compression of radio pulses in a chain of coupled resonators

S. N. Artemenko

Research Institute of Nuclear Physics, Tomsk Polytechnic University, 634050 Tomsk, Russia
(Submitted February 20, 1998)
Zh. Tekh. Fiz. **69**, 118–119 (July 1999)

Calculations are made of the efficiency and gain for the series compression of radio pulses in a chain of coupled resonators. © *1999 American Institute of Physics.* [S1063-7842(99)02407-1]

Didenko *et al.*[1] reported results of experiments to produce high-power ultrawide-band radio pulses by successive resonant compression in a chain of resonators connected by waveguide sections of arbitrary length. In this type of system, additional energy losses always occur during the buildup period in the connecting waveguide sections as a result of reflection losses from the resonators in each stage. It is easily shown that these losses may be ~10% for a two-stage system and ~15–20% for a three-stage system. Hence, the efficiency of similar three-stage ''hot'' systems cannot exceed 40%. The gain of the system will also decrease as a result of these losses. In view of this, the successive compression of radio pulses in a chain of coupled resonators with no connecting waveguide sections, in which the resonator of the previous stage is connected directly to the next-stage resonator is undoubtedly of interest.

An example of this type of system is shown in Fig. 1, where *1* is an rf generator, *2* is a ferrite isolator, *3* is a phase shifter, *4* are the storage resonators of each stage, *5* are switches that switch the resonators from storage mode to output mode, *6* is the load, and h_1 and h_2 are the intraresonator coupling elements. The system uses an interference method of rapidly regulating the resonator coupling with an external load, which is most frequently used in practice. The idea of the system involves using the vibrational transfer of energy from one resonator to another for the successive compression of radio pulses. In order to minimize the losses in the system, the coupling between the stages of the series should be selected so that over the time taken for the wave to make a round trip in the previous-stage resonator, the energy has been completely transferred to the next-stage resonator. In this case, energy will only be lost to the resonator walls during the transfer of energy from one stage to another (losses in the switches and the reflection losses in the first stage are not considered since these exist in any compression system).

In order to determine the efficiency and gain, we shall use the results of Ref. 2, in which the present author showed that the process of energy transfer from one resonator to another with the instantaneous switching of strong coupling between them obeys a law described by the function

$$b_i^2(t)=b_{i-1}^2\exp(-t/\tau_i)T_{i-1}\sin^2(\Omega_i t)/T_i, \qquad (1)$$

where b_i is the amplitude of the traveling wave in the ith-stage resonator, $\tau_i=T_i/2\alpha_i$ is the time constant of the ith resonator, α_i is the attenuation constant for a round trip in this resonator, $\Omega_i\approx h_i/\sqrt{T_iT_{i-1}}$ is the frequency of the relative transfer of energy from one resonator to another and back, and T_i is the round-trip time for a wave in the ith resonator.

It follows from Eq. (1) that in order to ensure that energy is transferred from the previous stage to the next stage within the round-trip time for a wave in the previous resonator, the coupling parameter between the resonators must satisfy the equality

$$h_i=(\pi\sqrt{T_i/T_{i-1}})/2. \qquad (2)$$

In this case, using the well-known expression for the storage efficiency of the first stage[3]

$$\eta_1=4\beta\tau_1(1-\exp(-t(1+\beta)/2\tau_1))^2/t_0(1+\beta)^2 \qquad (3)$$

we can easily derive the following expressions for the efficiency η and the gain M_N^2 of a system of N resonators:

$$\eta=\eta_1\prod_{i=2}^{N}\exp(-T_{i-1}/\tau_i)\approx\eta_1\prod_{i=2}^{N}(1-T_{i-1}/\tau_i)$$
$$=\eta_1\prod_{i=2}^{N}(1-2\alpha_i-1), \qquad (4)$$

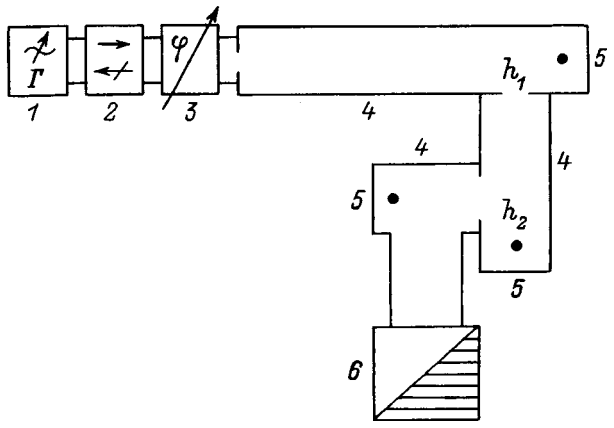

FIG. 1. System for successive compression of radio pulses in a chain of coupled resonators.

 © 1999 American Institute of Physics

$$M_N^2 = \eta t_0 / T_N, \tag{5}$$

where t_0 is the input pulse duration.

These expressions indicate that for the successive compression in a system of coupled resonators the efficiency and the gain may be significantly higher than those for a series of resonators connected by waveguide sections of arbitrary length and are almost exclusively determined by the storage efficiency of the first stage.

We note that successive compression in a chain of coupled resonators may also be effective with oversize resonators whose use in a series coupled via waveguide sections is usually only limited by the first stage because of the exponential decay of the output pulses of these resonators. This opens up possibilities for shaping nanosecond radio pulses with relatively high stored energy in compression systems where the energy is coupled out via interference switches based on rectangular waveguide T-junctions. For example, in accordance with Eqs. (4) and (5), nanosecond radio pulses with a peak power of ≈ 1 GW can be obtained in the 10 cm range with an input pulse duration of 3 μs, an output pulse duration of 3 ns after the second stage, a first-stage storage efficiency of 60%, and a generator power of 2 MW (bearing in mind that in Eq. (4) T_{i-1} should be replaced by the energy transfer time t_n, $T_{i-1} \ll t_n \ll \tau_i$). When the ratio of the energy transfer time to the time constant of the second-stage resonator is $\leqslant 0.1$, the overall efficiency of the system will be close to 55%, whereas for compression with reflection losses the overall efficiency does not exceed 35% because of the exponential decay of the first-stage output signal.

[1] A. N. Didenko, S. A. Novikov, S. V. Razin *et al.*, Dokl. Akad. Nauk SSSR **321**, 518 (1991) [Dokl. Akad. Nauk **36**, 792 (1991)].
[2] S. N. Artemenko, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **30**, 1289 (1987).
[3] S. V. Baraev and O. P. Korovin, Zh. Tekh. Fiz **50**, 2465 (1980) [Sov. Phys. Tech. Phys. **25**, 1444 (1980)].

Translated by R. M. Durham

Calculation of the electrostriction effect in thin-film metal–ferroelectric–metal structures

V. M. Bogomol'nyĭ

State Academy for Domestic Services, 141220 St. Tarasovskaya, Moscow Region, Russia
(Submitted March 4, 1998)
Zh. Tekh. Fiz. **69**, 120–124 (July 1999)

An analysis is made of the influence of electron injection from the cathode on the electrical conductivity and nonuniform thickness distribution of the piezoelectric properties of metal–insulator–metal structures using an electrostrictive, optically transparent FE ceramic such as lanthanum-containing lead zirconate titanate (PLZT). Because thin-film insulators, in contact with metal electrodes, exhibit semiconducting properties, the theory put forward by Mott and Gurney for insulating diodes can be used to calculate the electric fields and currents. The distribution of the polarization over the thickness of the piezoelectric layer was determined by means of an asymptotic solution, and relations were formulated for the electroelasticity of FE ceramic plates and shells. A detailed analysis was made of the electromechanical bending effect in homogeneous piezoelectric plates used as optical radiation modulators. It is shown that these formulas can be used for a computational and experimental determination of the distribution of the polarization over the thickness of thin-film piezoelectric structures.
© *1999 American Institute of Physics.* [S1063-7842(99)02507-6]

INTRODUCTION

Experimental and theoretical data have been used to show that the bending of homogeneous thin piezoelectric plates may be caused by the electronic subsystem, which determines the nonuniform polarization of the ferroelectric (FE). The distribution of the electric field, current, and polarization over the thickness metal–insulator–metal (MIM) structures based on the electrostrictive FE ceramic PLZT has been determined with allowance for electron injection from the cathode. Since the frequency of the electromagnetic waves interacting with the FE crystal is much higher than the frequency of the crystal lattice vibrations, only the electrons interact with the light. Thus, the electronic subsystem may have a significant influence on the optical properties of a FE.[1,2]

Varying the external electric field is the simplest means of controlling the electrooptic and piezoelectric properties of ferroelectric semiconductors and may be used in various piezo- and optoelectronic devices in which piezo- and elastooptic effects take place simultaneously.[3–5] The possibility of using FEs to produce an unconventional controlled reflecting surface for antennas in spacecraft communication systems was examined in Ref. 6.

Having a successful combination of physical properties (high photosensitivity, electrooptic and piezoelectric effects, electrical and mechanical strength, and suitability for fabrication technology), FE ceramic PLZT is used in various devices such as space–time modulators, optical recording of information, sensors, and actuators.[2,7–9]

Studies of the spatial distribution of the electric field, the thermionic emission current, and the polarization can unravel the mechanisms for charge transfer in FEs exposed to electric, thermal, and mechanical fields. In an electrostrictive FE ceramic a strong static field creates polarization and an additional (considerably weaker) alternating field is used to control the functional parameters of piezoelectric transducers. In the ordered polar phase in a FE, nonlinear effects occur which are determined not only by the direction of the external electric fields but also by the direction of the polarization vector. This is accompanied by the appearance of a polar thermopower and ''even'' electrical conductivity with respect to the field (unipolarity effect[8]).

The optical inhomogeneity induced by the external fields (photorefraction) can be attributed to space-charge-limited currents.[9] In the present paper we will obtain an upper limiting estimate for the current–voltage characteristic of an MIM structure. An analysis of various characteristics of FEs such as the metallic type of conductivity and the final polarization can be used to explain the physical principles and for mathematical modeling of the high-temperature superconductivity of lanthanum-doped metal oxide compounds.[10]

The need to calculate the electrical conductivity of thin-film MIM structures arises for various reasons. This conductivity appears in comparatively weak electric fields, and an investigation of the electrical conductivity as a function of the external electric field can provide information on the most important physical properties of FEs, especially the localization of the wave functions and the distribution density of carrier trapping and recombination centers.

A study of the energy spectrum of crystalline FEs is one of the main tasks of solid-state physics. An important factor in the formation of the energy spectrum is the mechanical deformation of the crystal lattice, which makes its own contribution (up to 40%) to changing the band gap.[11] The three-dimensional electroelasticity relations formulated here determine the mechanical stresses and strains of a FE as a function of the electric field and the electronic subsystem.

As a result of a mathematical modeling, we can solve the inverse problem by comparing the calculated results with the

 © 1999 American Institute of Physics

experiment. The experimental determination of the polarization distribution in thin-film MIM structures presents considerable technical difficulties. In the present study we put forward a substantiation for a semi-inverse method of determining the distribution of the polarization over the thickness of thin-film piezoelectric structures (comparable with the electron mean free path).

Experimental investigations of the photoelectric properties of the electric charge in CdS crystals showed that emission currents may form in MIM structures.[1,2] Krumins *et al.*[2] and Khalilov and Dimza[8] reported experimental studies of the influence of the emission currents on the photoelectric processes in optically transparent PLZT ceramic, which is used in optical switches, holographic memory elements, and other devices for recording and processing optical information. In the present study the electric fields, currents, and distribution of the polarization over the thickness of MIM structures are determined explicitly for thermionic emission from the cathode.

THEORY OF EMISSION CURRENTS IN METAL–INSULATOR–METAL STRUCTURES

In thin-film MIM structures in comparatively weak electric fields electrons are injected from the cathode, their concentration decreasing abruptly with increasing distance from it. In this case, the diffusion component of the current may be comparable with the drift component. Known analytical solutions only allow for the drift component of the current, which in many cases is taken to be constant over the thickness of the insulator.[12,13]

The electric fields and monopolar injection currents in MIM structures are determined by solving the Bessel equation, whose integration constants depend on the current and thickness of the dielectric. Calculation of the integration constants yields complex transcendental equations.[13] Thus, despite the simplifications noted earlier, the most comprehensive results of calculating the emission currents and electric fields were only obtained by computer. Here an asymptotic solution of this problem is obtained allowing for electron drift and diffusion.

The density of the thermionic emission current from the cathode $j(x)$ (x is the coordinate measured from the cathode), the electric field $E(x)$, and the concentration of electrons injected into the insulator $n(x)$ are determined from the following system of dimensionless equations:[12]

$$\frac{dE}{dx}=-n, \quad \frac{dn}{dx}+nE=j(x). \tag{1}$$

The units of the instantaneous coordinate $\bar{x}$ (the distance from the cathode), the current density $\bar{j}$, the electric field $\bar{E}$, and the concentration of injected electrons $\bar{n}$ are calculated using the formulas

$$\bar{x}=\left(\frac{\varepsilon_{33}^{T}\varepsilon_0 k^{T}}{4\pi e^2 n}\right)^{1/2}, \quad \bar{j}=\frac{e d n}{\bar{x}}, \quad \bar{E}=\frac{kT}{e\bar{x}},$$

$$\bar{n}=N_c e^{\psi}, \quad \psi=\frac{eU}{kT}, \tag{2}$$

where k is the Boltzmann constant, ε_{33}^{T} is the dielectric constant of the insulator, ε_0 is the permittivity of free space, T is the absolute temperature, U is the potential difference at the electrodes of the MIM structure, $d=\mu kT$ (μ is the mobility of the free charges), $\bar{x}$ is the Debye screening length, N_c is the effective density of quantum states in the conduction band, which is given by:

$$N_c=2\left(\frac{2\pi m k T}{h^2}\right)^{3/2},$$

where h is Planck's constant and m is the effective electron mass.

The main system of equations is reduced to a single equation

$$\frac{d^2E}{dx^2}+E\frac{dE}{dx}+j(x)=0. \tag{3}$$

The following boundary condition is an analytical expression for the properties of the injected cathode:

$$E(0)=0 \quad (\text{for} \quad x=0). \tag{4}$$

We also use the following condition to solve Eq. (3):

$$\int_0^h E(x)dx=-U, \tag{5}$$

where h is the thickness of the insulator and U is the voltage between the electrodes located on the surface of an insulating layer having the coordinates $z=\pm h/2$.

In accordance with the system of equations (1) and the boundary condition (4), we take the distributions $E(x)$ and $j(x)$ over the thickness of the insulating layer in the following form:

$$E(x)=B_1 x^{1/2}, \tag{6}$$

$$j(x)=A_1+A_2 x^{-3/2}, \tag{7}$$

where A_i ($i=1,2$) are unknown constants.

Using Eqs. (5) and (6), we calculate B_1 and then the function $E(x)$

$$B_1=-\frac{3}{2}Uh^{-3/2}, \quad E(x)=-\frac{3}{2}Ux^{1/2}h^{-3/2}. \tag{8}$$

Substituting Eqs. (6) and (7) into Eq. (3), we obtain

$$-\frac{B_1}{4}x^{-3/2}+\frac{B_1^2}{2}+A_1+A_2x^{-3/2}=0.$$

Equating to zero the sum of the coefficients with the same powers x in this equality, we find A_1 and A_2, and then the function $j(x)$ from Eq. (7)

$$j(x)=-\frac{9}{8}\frac{U^2}{h^3}-\frac{3}{8}\frac{U}{(hx)^{3/2}}. \tag{9}$$

The first term in formula (9) is the known expression for the current–voltage characteristic, which was obtained neglecting the diffusion current of electrons injected into the insulator.[12]

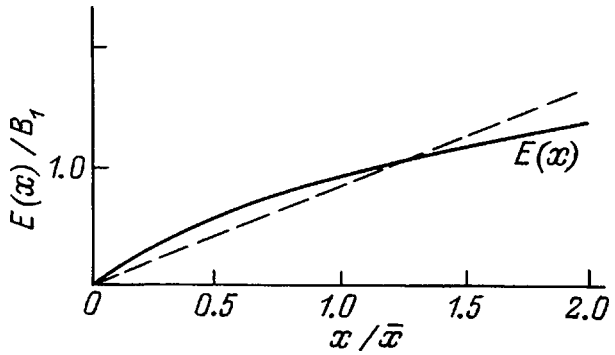

FIG. 1. The function $E(x)$. The coordinate $x=0$ is positioned near the cathode.

For FE ceramic PLZT the function $E(x)$ calculated using formula (8) is plotted in Fig. 1. This can be approximated by a linear dependence on x (dashed curve) with sufficient accuracy for practical applications.

ELECTROELASTICITY RELATIONS FOR PIEZOELECTRIC TRANSDUCERS NONUNIFORMLY POLARIZED OVER THICKNESS

For the most general case of the geometric profile of piezoelectric transducers we shall consider a shell of revolution made of an electrostrictive ceramic in the nonpolar phase (inverse piezoelectric effect), in which a strong static field induces polarization and an additional much weaker alternating electric field creates working strains of the piezoelements. When the polarization lies in the direction of the shell thickness, the equations for the electrostriction effect for temporally constant stresses σ_i, strains ε_i, and electric fields E_3 have the form[14]

$$\varepsilon_1=s_{11}^E\sigma_1+s_{12}^E\sigma_2+Q_{12}E_3^2,$$
$$\varepsilon_2=s_{11}^E\sigma_2+s_{12}^E\sigma_1+Q_{12}E_3^2, \tag{10}$$

where ε_i $(i=1,2)$ are the strains of the shell in the direction of the unit vectors, τ_1 and τ_2 (Fig. 2), σ_i are the mechanical stresses, S_{1i}^E are the elastic compliance constants of the ceramic, Q_{12} is the electrostriction constant, and E_3 is the electric field.

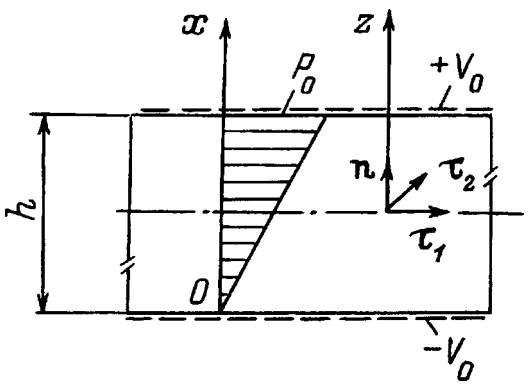

FIG. 2. Distribution of the polarization in an MIM structure. The coordinate $z=0$ is positioned at the middle surface of the piezoelectric layer.

The following assumptions were made to derive the electroelasticity relations.

1) The polarization P_3, the field E_3, and the constant Q_{12} depend linearly on the coordinate z, which is measured from the central surface of the shell $(-h/2\geqslant z\geqslant/2$; Fig. 2),

$$P_3=P_3^0\alpha,\quad E_3=E_3^0\alpha,$$
$$Q_{12}=Q_{12}^0\alpha,\quad \alpha=(1/2+z/h), \tag{11}$$

where P_3^0 and Q_{12}^0 are the values corresponding to the coordinate $z=h/2$ (Fig. 2) and $E_3^0=-2V_0/h$ (h is the shell thickness).

2) In accordance with the Kirchhoff–Love hypotheses[15] and the polarization distribution (11), we take the shell strains in the following form:

$$\varepsilon_1=\varepsilon_i^0\left(1+\frac{2z}{h}\right)+z\varkappa_i, \tag{12}$$

where ε_i^0 and $\varkappa_i$ $(i=1,2)$ are the relative strains and the changes in the principal curvatures of the central surface of the shell.

3) On the basis of the experimental data we take the dielectrci constant of the FE ceramic as a constant averaged over the shell thickness. From the fundamental system of equations (10) we have

$$\sigma_1=\frac{1}{s_{11}^E(1-\mu^2)}(\varepsilon_1+\mu\varepsilon_2-\hat{E}_3),$$
$$\sigma_2=\frac{1}{s_{11}^E(1-\mu^2)}(\varepsilon_2+\mu\varepsilon_1-\hat{E}_3),$$
$$\hat{E}_3=(1+\mu)Q_{12}(E_3^0)^2,\quad \mu=-\frac{s_{12}^E}{s_{11}^E}, \tag{13}$$

where μ is the Poisson ratio.

Expressing the internal mechanical stresses in the shell T_1 and T_2 and the bending moments M_1 and M_2 in terms of integrals of the stresses σ_1 and σ_2 (15), we obtain the following relations for the electroelasticity:

$$T_1=D_T\left[\varepsilon_1^0+\mu\varepsilon_2^0-\frac{(1+\mu)}{4}Q_{12}^0(E_3^0)^2\right],$$
$$T_2=D_T\left[\varepsilon_2^0+\mu\varepsilon_1^0-\frac{(1+\mu)}{4}Q_{12}^0(E_3^0)^2\right],$$
$$M_1=D_M[(2/h)(\varepsilon_2^0+\mu\varepsilon_1^0)+\varkappa_1+\mu\varkappa_2-E_{3M}],$$
$$M_2=D_M[(2/h)(\varepsilon_2^0+\mu\varepsilon_1^0)+\varkappa_2+\mu\varkappa_1-E_{3M}],$$
$$D_T=\frac{h}{s_{11}^E(1-\mu^2)},\quad D_M=\frac{h^3}{12s_{11}^E(1-\mu^2)},$$
$$E_{3M}=\frac{3}{4}\frac{(1+\mu)}{h}Q_{12}^0(E_3^0)^2. \tag{14}$$

We consider a bracket-mounted homogeneous piezoelectric plate nonuniformly polarized over its thickness (Fig.

1). We shall determine the deflection of the free end under the following conditions, which are satisfied over its entire length

$$T_1 = M_1 = 0, \quad \varkappa_1 = \text{const}$$

$$(T_2 = M_2 = \varepsilon_2 = \varkappa_2 = \mu = 0). \tag{15}$$

From the first and third equations in (14) and (15) we have

$$\varepsilon_1^0 = Q_{12}^0 (E_3^0)^2/4, \quad (2/h)\varepsilon_1^0 + \varkappa_1 - E_{3M} = 0. \tag{16}$$

We determine the change in the curvature $\varkappa_1$ from the system (16)

$$\varkappa_1 = Q_{12}^0 (E_3^0)^2/2h.$$

Taking into account this last expression and Eq. (12), we determine the deflection of the free end of the plate using the following formula:

$$f = \frac{l^2 \Delta \varkappa_1}{2} = \frac{l^2 Q_{12}^0 (E_3^0)^2}{2h},$$

$$\Delta \varkappa_1 = [\varepsilon_1(h/2) - \varepsilon_1(-h/2)]/h = Q_{12}^0 (E_3^0)^2/h, \tag{17}$$

where l is the length of the plate, $\varepsilon_1(\pm h/2)$ are the relative strains at the shell surfaces having the coordinates $z = \pm h/2$.

For the field $E_3^0 = 6000$ V/cm and the parameters $l = 30$ mm, $h = 0.15$ mm, $Q_{12}^0 = -1.45 \times 10^{-10}$ mm^2/V, and $\mu = 0.34$, the deflection of the free end of the plate is $f = 0.14$ mm.

Formula (17) can be used to obtain an integral estimate of the change in the polarization over the thickness of thin piezoelectric layers (10–200 μm). Defining the polarization distribution using formula (17), we can calculate the deflection of the free end of a bracket-mounted sample. By comparing the calculated result with the experiment, we can determine the real distribution of the polarization over the thickness of the piezoelectric element by this method of calculation and experiment. The results of the numerical calculations presented here (Fig. 1) are confirmed by the results obtained using numerical[16,17] and analytical[18] methods and also by the experimental data.[19]

CONCLUSIONS

1. The distributions of the electric fields and currents over the thickness of MIM structures have been determined explicitly by means of an asymptotic solution.

2. Formula (9) refines the current–voltage characteristic of an insulating diode obtained by Mott and Gurney. The first term in this formula is the same as the well-known expression[12] and the second term allows for carrier drift.

3. Experiments[19] and theoretical estimates[13,16–18] indicate that electrons are injected from the electrodes in electric fields $E_3 = 250$–500 V/cm. These experimental and theoretical data suggest that the electrooptic and piezoelectric properties of FEs depend strongly on their electronic subsystem. In particular, formulas (6)–(9) can be used for a quantitative estimate of the influence of thermionic emission on photo- and electrooptic effects in ferroelectric semiconductors.

[1] H. Y. Fan, in *Semiconductors and Semimetals*, Vol. 3 (*Optical Properties of III–IV Compounds*, edited by R. K. Willardson and A. C. Beer (Academic Press, New York, 1967), pp. 405–419.
[2] A. E. Krumins and V. Y. Fritzberg, Ferroelectrics **34**(1), 149 (1981).
[3] K. Kern, J. Phys. Chem. Solids **23**, 249 (1962).
[4] G. Harbeke, J. Phys. Chem. Solids **24**, 957 (1963).
[5] L. S. Kamzina and N. N. Kraĭnik, Fiz. Tverd. Tela (Leningrad) **29**, 1868 (1987) [Sov. Phys. Solid State **29**, 1074 (1987)].
[6] I. S. Bedareva and S. V. Koshevaya, Radiotekh. Elektron. **40**, 1118 (1995).
[7] A. Wegner, S. R. Bruck, and A. Y. Wu, Ferroelectrics **116**, 195 (1995).
[8] A. Khalilov and A. Dimza, Ferroelectrics **69**, 59 (1986).
[9] B. V. Krylov and V. E. Leparskiĭ, Zh. Tekh. Fiz **67**(10), 51 (1997) [Tech. Phys. **42**, 1158 (1997)].
[10] A. S. Shcherbakov, M. N. Katsnel'son, A. V. Trefilov *et al.*, *Ferroelectric Anomalies and Superconductivity in Metal-Oxide Compounds*, No. 6 [in Russian], Syktyvkar (1988), 22 pp.
[11] A. Yu. Zakharov, Fiz. Tverd. Tela (Leningrad) **17**, 1274 (1975) [Sov. Phys. Solid State **17**, 823 (1975)].
[12] M. A. Lampert and P. Mark, *Current Injection in Solids* (Academic Press, New York, 1970), 351 pp.
[13] É. I. Adirovich, Fiz. Tverd. Tela (Leningrad) **2**, 1410 (1960) [Sov. Phys. Solid State **2**, 1282 (1960)].
[14] W. P. Mason, *Piezoelectric Crystals and Their Applications in Ultrasonics* (Van Nostrand, Princeton, 1950).
[15] S. P. Timoshenko, J. M. Lessels, and A. M. Mech, *Applied Elasticity* (Westinghouse Technical Night School Press, East Pittsburgh, 1925), 544 pp.
[16] J. Lindmayer, J. Reynolds, and C. Wrigley, J. Appl. Phys. **34**, 809 (1963).
[17] Ch. Schnitler, Phys. Status Solidi A **45**(2), 179 (1978).
[18] S. Esener and S. H. Lee, J. Appl. Phys. **58**, 1380 (1985).
[19] A. S. Bogdanovich, M. M. Nekrasov, Yu. A. Sikorskiĭ *et al.*, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **6**, 72 (1962).

Translated by R. M. Durham

Voltage rectification by hard superconductors

A. N. Ul'yanov

Donetsk Physical-Engineering Institute, National Academy of Sciences of Ukraine, 340114 Donetsk, Ukraine

(Submitted April 30, 1998)
Zh. Tekh. Fiz. **69**, 125–127 (July 1999)

The Bean critical state model is used to calculate voltages and voltage waveforms in hard superconductors carrying ac and dc transport currents. A comparison with the experiment shows that these characteristics are accurately described by this model. The voltage rectification effect by current-carrying superconductors is explained. © *1999 American Institute of Physics.* [S1063-7842(99)02607-0]

Superconducting composites based on silver-sheathed bismuth superconducting ceramic having a high critical current density ($I_c \sim 2\times 10^4$ A/cm^2) have now been fabricated.[1] These materials can be used to carry ac and dc transport current and as conductors for superconducting magnets. Thin films of high-temperature superconductors can be used in microelectronics. Bulk materials based on $Bi_2Sr_2Ca_2Cu_3O_y$ ceramic can be used to transfer current between liquid-nitrogen and liquid-helium temperature. Transport of ac currents by superconductors is now the subject of very intensive research (see Refs. 2–6, for instance). Both experimental[4–6] and theoretical studies[3,7] are focusing particular attention on the hysteresis losses. The voltage wave form, and the current and frequency dependences of the voltage and losses have been investigated experimentally[2–6] and theoretically.[2,3,5,7] The generation of harmonics by superconducting bismuth ceramic[8] and yttrium films[9] carrying ac transport currents was studied by an inductive method. The behavior of a superconducting plate carrying an ac current in a perpendicular magnetic field was studied in Ref. 10. Grishin *et al.*[2] were most likely the first to investigate the transport of an ac current $I(t)=I_{\rm dc}+I_{\rm ac}\cos\omega t$ containing a dc component. They reported the experimental observation of a voltage ''rectification'' effect by the dc current and used the Kim critical-state model[11] for the case $|I(t)|<I_c$ to show that the magnitude of the rectified voltage depends linearly on the frequency of the ac current.

In the present paper we use the Bean critical state model[12] to calculate the voltage and its wave form for a superconductor carrying an ac transport current $I(t)=I_{\rm dc}+I_{\rm ac}\cos\omega t$ and we explain the origin of the frequency-independent component of the rectified voltage observed in Ref. 2.

1. When a transport current $I_{\rm tr}(t)$ flows through a superconductor, magnetic flux vortices form at its edges for a certain value of the current. Under the action of the Lorentz force these vortices move toward the center of the sample. This vortex motion leads to energy dissipation and causes a drop in the voltage $u_R(t)$ at the sample. Moreover, if the transport current is ac, it induces an alternating self-magnetic field with the flux $\Phi(t)$, which leads to the appearance of hysteresis losses caused by magnetic reversal of the sample in the alternating self-field. A change in the magnetic flux gives rise to a self-induction emf in the sample $e_L(t)=-d\Phi(t)/dt$ and causes the voltage drop $u_L(t)=-e_L(t)$. Hence, the voltage $u(t)$ across a superconductor carrying an ac transport current is determined by the sum of two contributions

$$u(t)=u_R(t)+u_L(t). \tag{1}$$

The ''active'' voltage component $u_R(t)$ has the same phase as the current and is determined by the current–voltage characteristic of the superconductor. The ''inductive'' voltage component $u_L(t)$ contains a cophasal component and a (quadrature) component whose phase differs by 90° from that of the current and may be determined using the critical-state model.

2. We shall determine the active component of the sample voltage. To be specific, we shall assume that the current–voltage characteristic of the superconductor is a power law

$$u_R(I)=u_c(I/I_c)^{\alpha}, \tag{2}$$

where I_c is the critical current of the sample determined as usual from the criterion $u_R=u_c=1\ \mu$V/cm.

Figure 1 gives the voltage for a superconductor carrying an ac current. In this case, if the transport current contains both ac and dc components, the working point on the superconductor current–voltage characteristic is shifted (Fig. 2). If $I_{\rm dc}>0$ the positive half-wave of the current is amplified and the negative half-wave suppressed. This leads to the appearance of a constant, frequency-independent voltage component at the sample

$$u_{Rc}=\frac{1}{T}\int_0^T u_R(I(t))\,dt. \tag{3}$$

This explains the effect of voltage rectification by a superconductor, which was observed experimentally in Ref. 2. The magnitude of the rectified voltage u_{Rc} is determined by the type of current–voltage characteristic (in case (2) by the exponent α), and by the amplitude of the ac and dc transport currents. For instance, for a power function with the exponent $\alpha=3$ the rectified voltage is

 © 1999 American Institute of Physics

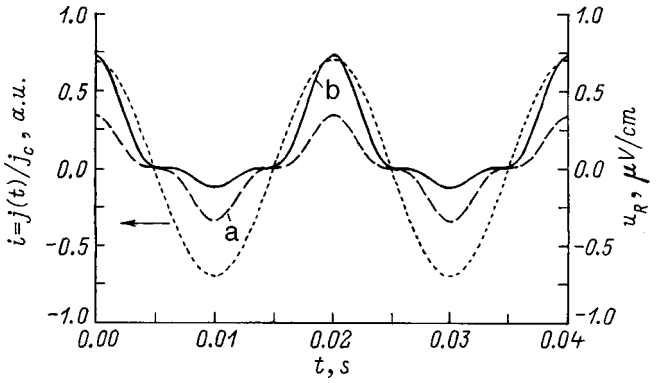

FIG. 1. Time dependences of the current $j(t)=j_{dc}+j_{ac}\cos\omega t$ ($j_{ac}=0.7j_c$, $j_{dc}=0$) and the active sample voltage u_R for $j_{dc}=0$ (a) and $0.2j_c$ (b).

$$u_{Rc}=u_c\frac{1}{2I_c^3}(3I_{ac}^2I_{dc}+2I_{dc}^3). \tag{4}$$

3. We shall determine the inductive voltage component $u_L(t)$ for an infinite plate using the Bean critical-state model.[12] According to this model, the current flowing through a superconductor is only nonzero in that part of the sample where the induction is $B\neq\text{const}$, the current depends only on temperature, and is equal to its critical value. As the external field varies, the distribution of the magnetic induction initially varies in the outer part of the sample and then in the inner part. This is caused by magnetic flux trapping at pinning centers. As a result, the change in induction lags behind the change in the external field in terms of phase, which in this case is determined by the current.

The magnetic field in the sample is described by Maxwell equations

$$\operatorname{curl}\mathbf{B}=\frac{4\pi}{c}\mathbf{j}_c, \tag{5}$$

where c is the velocity of light and j_c is the critical current density.

Solving Eq. (5) and bearing in mind that the field at the sample boundary is equal to the sum of the fields initiated by the dc and ac transport currents $H_{tr}=\pm(2\pi d/c)(j_{dc}+j_{ac}\cos\omega t)$, ($j_{dc}$ and $j_{ac}\cos\omega t$ are the densities of the dc and ac transport currents, respectively, d is the thickness of the plate, and the plus and minus signs correspond to opposite sides of the plate), we can obtain the distribution of the magnetic induction in the sample. The following cases are possible: (a) $I_{dc}+I_{ac}<I_c$, when the amplitude of the ac current plus the dc current is less than the critical current; (b) the opposite case, when $I_{dc}+I_{ac}<I_c$. Here we only consider case (a) since this has practical applications. In this case, the superconductor is in the critical state over the entire period. The induction in the sample is given by

$$B(x)=B_c\begin{cases}0, & 0\leqslant|x|\leqslant x_0,\\ \left(\dfrac{j_{ac}+j_{dc}}{j_c}-1+\dfrac{2|x|}{d}\right)\operatorname{sgn}x, & x_0\leqslant|x|\leqslant x_m,\\ \left[-\left(\dfrac{j_{ac}}{j_c}-1+\dfrac{2|x|}{d}\right)\right.\\ \left.\times\operatorname{sgn}\dot I+\dfrac{j_{dc}}{j_c}\right]\operatorname{sgn}x, & x_m\leqslant|x|\leqslant x_t,\\ \left[\dfrac{j_{ac}}{j_c}\cos\omega t-\left(1-\dfrac{2|x|}{d}\right)\right.\\ \left.\times\operatorname{sgn}\dot I+\dfrac{j_{dc}}{j_c}\right]\operatorname{sgn}x, & x_t\leqslant|x|\leqslant\dfrac{d}{2},\end{cases} \tag{6}$$

the **x** is directed perpendicular to the plane of the plate, $x=0$ corresponds to the middle of the plate

$$B_c=\frac{2\pi d}{c}j_c,\quad x_0=\frac{d}{2}\left(1-\frac{j_{ac}+j_{dc}}{j_c}\right),$$

$$x_n=\frac{d}{2}\left(1-\frac{j_{ac}}{j_c}\right),$$

$$x_t=\frac{d}{2}\left(1-\frac{j_{ac}}{2j_c}(1+\cos\omega t\operatorname{sgn}\dot I)\right),$$

$\operatorname{sgn}\dot I$ is the sign of $dI(t)/dt$.

By integrating expression (6) over the half-thickness of the plate, we can determine the magnetic field flux $\Phi(t)$ per unit length (across half the plate). By then differentiating this expression with respect to time, we obtain the self-induction emf

$$e_L\equiv-\frac{1}{c}\frac{\partial\Phi}{\partial t}=\frac{\pi d^2\omega}{2c^2}\frac{j_{ac}^2}{j_c}(\sin\omega t-\cos\omega t|\sin\omega t|) \tag{7}$$

and the voltage drop per unit length $u_L(t)=-e_L(t)$ (Fig. 3). The total sample voltage $u(t)=u_R(t)+u_L(t)$ is plotted in Fig. 4 and its profile is broadly consistent with the results obtained in Ref. 2. The difference can evidently be attributed to the experiments since the experimental curves obtained in Refs. 2 and 4 differ (without the dc component of the transport current). It follows from Eq. (7) that $u_L(t)$ is directly proportional to the frequency of the ac current and does not depend on the dc component of the transport current. Using the values of the sample thickness $d\sim100\,\mu$m, the width $w\sim4$ mm, and the critical current $I_c=8.3$ A from Ref. 5, we can find that

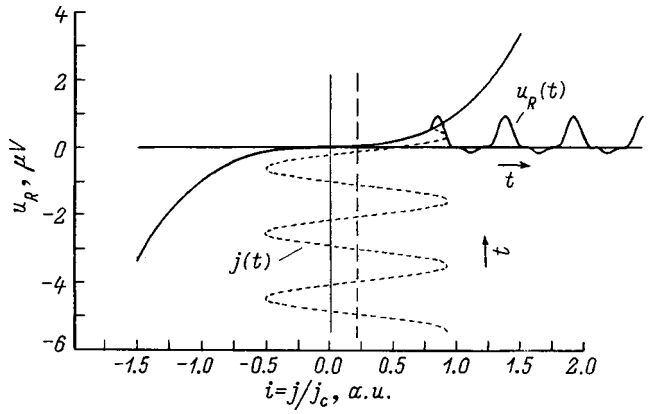

FIG. 2. Current–voltage characteristic of superconductor ($\alpha=3$) and voltage u_R during passage of the transport current $j(t)=j_{dc}+j_{ac}\cos\omega t$, $j_{ac}=0.7j_c$, $j_{dc}=0.2j_c$.

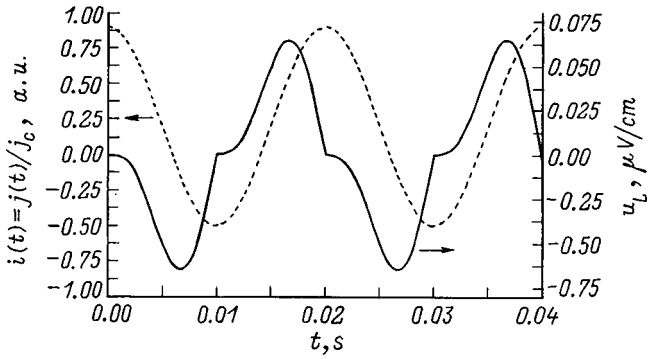

FIG. 3. Time dependences of the current $j(t)=j_{dc}+j_{ac}\cos\omega t$ ($j_{ac}=0.7j_c$, $j_{dc}=0.2j_c$) and the inductive voltage u_L.

$$A_1=\frac{\pi d^2\omega}{2c^2}\frac{j_{ac}^2}{j_c}$$

in Eq. (7) is of the same order of magnitude as $\sim 10^{-2}-10^{-1}\,\mu$V/cm at frequency $\omega=50$ Hz and transport current density $j_{ac}\sim j_c$. This value of A_1 agrees with the experimental data from Refs. 4 and 5.

Grishin *et al.*[2] calculated the sample voltage using the Kim critical-state model, in which the critical current depends on temperature, as in the Bean model, and on the local field

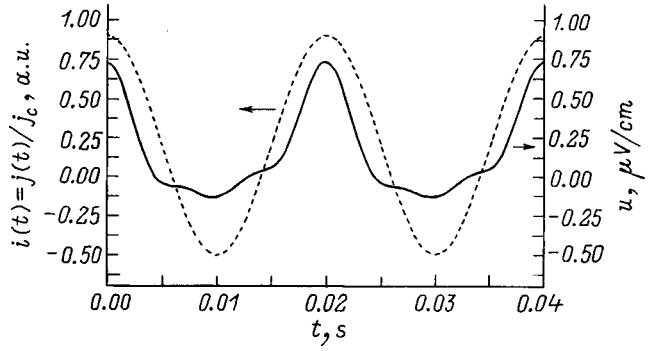

FIG. 4. Time dependences of the current $j(t)=j_{dc}+j_{ac}\cos\omega t$ ($j_{ac}=0.7j_c$, $j_{dc}=0.2j_c$) and the total voltage $u(t)$.

$$j_c(T,H)=\frac{j_c(T,0)}{1+H/H_0},$$

where $H_0\sim 0.1$ T.

The sample voltage for the case in Ref. 2, i.e., $|I(t)|<I_c$, is given by

$$u_{LK}=-\frac{\pi d^2\omega}{2c^2}\frac{j_{ac}^2}{j_c}\left(\sin\omega t-\cos\omega t|\sin\omega t|+\frac{\pi d j_{ac}}{cH_0}e(t)\right). \tag{8}$$

The expression $e(t)$ contains both constant and harmonic components whose magnitude and amplitude depend on the ratio I_{dc}/I_{ac} and are of the order of magnitude of one.[2] Thus, compared with Eq. (7), expression (8) has an additional term which contains the dc rectified component of the sample voltage. However, this component is negligible because of the smallness of the ratio $dj_{ac}/cH_0\sim 10^{-3}-10^{-2}$.

To sum up, the Bean critical-state model has been used to calculate the voltage for a current-carrying superconductor. The results show good agreement with the experimental ones. An explanation has been put forward for the origin of the frequency-independent rectified voltage component for superconductors carrying an ac transport current. The results can be used to design systems for both high-current and low-current electronics.

The author would like to thank Yu. A. Genenko for helpful discussions.

[1] M. J. Minot, Adv. Cryogenic Eng. A **40**, 131 (1994).
[2] A. M. Grishin, J. Niska, B. Loberg, and H. Weber, J. Appl. Phys. **76**, 6947 (1994).
[3] K. -H. Müller and K. E. Leslie, IEEE Trans. Appl. Supercond. **AS-7**, 306 (1997).
[4] S. P. Ashworth, Physica C **229**, 355 (1994).
[5] M. Ciszek, A. M. Campbell, and B. A. Glowacki, Physica C **233**, 203 (1994).
[6] H. Eckelmann, M. Daumling, M. Quilitz, and W. Goldacker, Physica C **295**, 198 (1998).
[7] W. T. Norris, J. Phys. D **3**, 489 (1970).
[8] A. M. Grishin, V. N. Koreniski, K. V. Rao, and A. N. Ulyanov, Appl. Phys. Lett. **65**, 487 (1994).
[9] A. M. Grishin, V. F. Drobot'ko, A. A. Mazaev, V. D. Stasovskiĭ, and V. A. Khokhlov, Fiz. Nizk. Temp. **19**, 635 (1993) [Low Temp. Phys. **19**, 453 (1993)].
[10] E. H. Brandt and M. Indenbom, Phys. Rev. B **48**, 12 (1993).
[11] Y. B. Kim, C. F. Hempstead, and A. R. Strand, Phys. Rev. **129**, 528 (1963).
[12] C. P. Bean, Phys. Lett. **8**, 250 (1962).

Translated by R. M. Durham

Generalized conductivity and optimum energy release

S. A. Baranov

T. G. Shevchenko Dnestr State University, 278000 Tiraspol, Moldavia
(Submitted May 19, 1998)
Zh. Tekh. Fiz. **69**, 128–129 (July 1999)

A formula for the average generalized conductivity of composites, which specifically allows for an inductive coupling channel and the shape of the microconductor sections. is used to analyze the optimum energy release in the poorly conducting phase. © *1999 American Institute of Physics.* [S1063-7842(99)02707-5]

Composites having two components with different conductivities are now widely used. We note that a material based on a cast microconductor may prove a potentially useful composite. In the present paper, an analysis is made of the problem of optimizing the concentration of the conducting component to maximize the energy release at the nonconducting component. This problem is undoubtedly of enormous practical value and in the limit of infinitely high resistance of the nonconducting component, it can give an indication of the ''percolation level'' of the system.

1. The generalized conductivity of a system may depend on the geometry of the phases.[1] The Odelevskiĭ formula[2] is well-known for a stochastic bulk mixture of powders. Baranov and Shcheglov[3] and Baranov[4] derived an expression which can be used to analyze a mixture of infinitely thin cylinders on a plane and can also allow for an inductive particle coupling channel in the composite. We give the formula for the average generalized conductivity in the form[3,4]

$$\Sigma_m = A(x_i, \Sigma_i, a) + \sqrt{A^2(x_i, \Sigma_i, a) + a\Sigma_1\Sigma_2}, \tag{1}$$

where

$$A = \frac{\Sigma_1(x_1 - ax_2) + \Sigma_2(x_2 - ax_1)}{2},$$

$x_{1,2}$ are the volume concentrations of components having the conductivity $\Sigma_{1,2}$ (the subscript ''2'' will refer to the microconductor), where $x_1 + x_2 = 1$.

We then set

$$a = \frac{1}{l-1}, \tag{2}$$

where l is a certain spatial dimension.

If $l=3$, we obtain the Odelevskiĭ formula. If $l=2$, we obtain a formula which for low concentrations of the conducting phase is ''matched'' with a similar formula (28) from Ref. 5 which was obtained for ordered thin cylinders on a plane (perpendicular to the direction of measurement). The level of strongly varying conductivity Σ_m obtained from the graph for this case (see Fig. 1, curve *1*) $x_2^c \sim 0.5-0.57$ corresponds to the percolation level on a planar lattice (for the Odelevskiĭ formula this level is $x_2^c \sim 0.3-0.38$). In this case, we arbitrarily take the percolation level to mean the region of abruptly varying conductivity.

We also note that in the case $l=2$, formula (1) is matched with the accurate result when $x_{1,2}=1/2$ (Ref. 6)

$$\Sigma_m = \sqrt{\Sigma_1\Sigma_2}.$$

We shall postulate that for cases of inductive coupling we need to replace l by

$$l_1 = l(1 + Y/R), \tag{3}$$

where Y/R is the ratio of the reactance to the resistance for a single element in a composite.

For a microconductor[4] we have

$$Y/R \sim (r/\delta)^2, \tag{4}$$

where r is the radius of the microconductor core, δ is the skin layer depth, and $\delta \sim (\omega\mu\mu_0\sigma)^{1/2}$.

For the real cores of a cast amorphous microconductor ($r \sim 1-20\,\mu$m), the region where $r-\delta$ belongs to the ultrahigh-frequency and microwave ranges. If $Y/R \sim 1$, then $l_1 \sim 4$ and the level of strongly varying conductivity (see Fig. 1, curve *3*) is $x_4^c \sim 0.25-0.27$.

2. When a highly conducting phase is added to a poorly conducting one (such as a microconductor in rubber), the energy (heat) release in the poorly conducting phase has an extremum at the percolation threshold. We determine this from[7]

$$\frac{d\Sigma_m}{dx_1} - k\Sigma_2 = 0, \tag{5}$$

where k is a parameter which determines the fraction of the current passing through the poorly conducting phase.

This equation was obtained using an approximation in which the term dk/dx_1 is neglected. As we shall see, in the asymptotic limit $\Sigma_1 \to 0$ the solution does not depend on the parameter k (physically this approximation can be justified by the fact that k should be a continuous function of the concentration as far as the percolation threshold being studied; allowance for dk/dx_1 merely complicates the solution and does not produce any significant result).

We note the region (fork) of variation of k (Ref. 7)

$$1 \geqslant k \geqslant \left[\frac{(1-x_1)\Sigma_2}{x_1'\Sigma_1 + (1-x_1')\Sigma_2}\right]^2,$$

 © 1999 American Institute of Physics

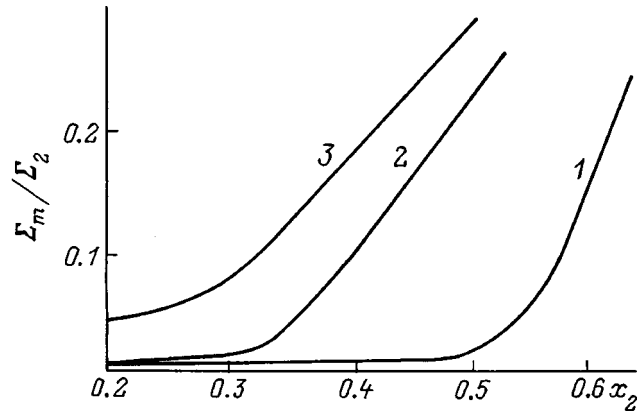

FIG. 1. Generalized conductivity calculated using formula (1) for $\Sigma_2/\Sigma_1 = 10^3$ for $l=2$ (*1*), 3 (*2*), and 4 (*3*).

where x_1' is the concentration corresponding to this extremum (which lies near the percolation threshold).

The value of the parameter k at the percolation threshold for a planar ($l=2$) or two-dimensional problem can be obtained from the results given in Ref. 6, which we shall use. Thus, for the planar problem ($l=2$) we have

$$x_1' \cong \frac{1}{2} + \frac{\sqrt{\Sigma_1 \Sigma_2}}{\Sigma_2 - \Sigma_1} \quad (\Sigma_2 > \Sigma_1), \tag{6}$$

which agrees with the exact results (see, for example, Ref. 6) ($x_1' \to 1/2,\ \Sigma_1 \to 0$).

A formula for the bulk problem was derived and put forward in Ref. 7, where the asymptotic value

$$x_1' \to 2/3 \quad (\Sigma_1 \to 0)$$

was obtained (a similar but very cumbersome formula is obtained[7] for arbitrary l). We shall give the principal asymptotic result of this theory

$$x_2' \to 1/l (\Sigma_1 \to 0)$$

(generalizing this result we postulate that the percolation threshold and the parameter l are inversely proportional).

To sum up, we have put forward simple approximate formulas for the conductivity in the theory of an effective medium. We have solved an extremum problem which can be used to estimate the percolation threshold and relate this to the coupling parameter l. The parameter l can characterize the geometry of the mixture particles or the inductive (capacitive) coupling. Note that a qualitative experimental confirmation has been obtained for this theory.[4]

The author is grateful to P. I. Khadzhi and É. P. Sinyavskiĭ for discussions of this work.

[1] T. N. Dul'nev and V. V. Novikov, Inzh.-Fiz. Zh. **41**(1), 172 (1981).
[2] V. I. Odelevskiĭ, Zh. Tekh. Fiz. **21**, 678 (1951).
[3] S. A. Baranov and Yu. A. Shcheglov, Élektron. Obrab. Mater., No. 6, 73 (1983).
[4] S. A. Baranov, Vest. Prednest. Univ., No. 1, 126 (1994).
[5] V. I. Odelevskiĭ, Zh. Tekh. Fiz. **21**, 667 (1951).
[6] A. M. Dykhne, Zh. Éksp. Teor. Fiz. **59**, 110 (1970) [Sov. Phys. JETP **32**, 63 (1970)].
[7] S. A. Baranov and Yu. A. Shcheglov, Élektron. Obrab. Mater., No. 1, 71 (1985).

Translated by R. M. Durham

Unheated magnetron gun as an electron source for a resonant linear accelerator

G. M. Ivanov, L. A. Makhnenko, and S. A. Cherenshchikov

National Science Center, Institute of Physics and Technology, 310108 Kharkov, Ukraine
(Submitted February 23, 1998)
Zh. Tekh. Fiz. **69**, 130–134 (July 1999)

Investigations of the operation of a cold-cathode magnetron gun as part of a traveling-wave linear accelerator are described. Two operating regimes of the gun are observed: with and without the microwave field of the prebuncher near the gun. In the regime with no microwave field, short (around 2 ns) pulses of accelerated electrons of up to 0.5 A were obtained with a gun current up to 20 A. The presence of a microwave field near the gun makes it possible to obtain a beam of longer duration (up to 1.0 μs), but with a current of up to 20 mA at the accelerator exit and 1 A from the gun. The operating mechanism of the gun is attributed to a secondary-electron current rise and the establishment of self-sustained secondary emission. A comparison between the conditions of beam production from a thermionic gun and that studied for the same injection energies indicates that the characteristics of this gun are acceptable for injection into an accelerator. © *1999 American Institute of Physics.* [S1063-7842(99)02807-X]

INTRODUCTION

The development of accelerator technology and the related technology for powerful microwave sources is imposing new requirements on sources of high-power electron beams. Alternatives to the conventional thermionic cathode sources are needing to be considered increasingly frequently. One such alternative is a secondary-emission magnetron gun in cases where a high current density and long service life are both required.[1] The long service life of the cathode creates new possibilities in accelerator technology such as the production of commercial, sealed accelerator models instead of the existing pumped systems. In addition, as will be shown subsequently, a magnetron gun with an unheated cathode can produce short (nanosecond) beam current pulses in the accelerator from comparatively long voltage pulses at the gun. This has important practical applications for the development of high-current, short-pulse accelerators since it facilitates the formation of the high-voltage pulses required to supply the gun. It is also postulated that a magnetron gun with an unheated cathode will be stable to bombardment by return electrons reflected from the accelerating structure with appreciable energy, which may limit the pulse frequency and the cathode service life, as experience of operating these accelerators has shown.[2] A magnetron gun has various characteristics that can influence its operation as part of a linear electron accelerator. Of particular note are that the cathode is not screened from the magnetic field and the beam is tubular. The aim of the present study is to make a preliminary assessment of their influence and the operating characteristics of this gun as part of a resonant linear electron accelerator.

DESCRIPTION OF APPARATUS

For the tests we used the universal injector complex of the LUÉ-300 accelerator.[3] The injector complex was designed to extend the scope of the LUÉ-300 accelerator in terms of increasing the beam current and obtaining short (nanosecond) durations in a stored energy regime. The injector complex consisted of two accelerating sections and one injector section, which were fitted with a regeneration system to enhance the stored energy; this comprised bunching cavities inserted inside focusing solenoids and a universal triode gun with beam compression based on a porous metal thermionic cathode. For the experiments the triode gun was dismantled and the magnetron gun was mounted on its insulator.

The apparatus is shown schematically in Fig. 1. The magnetron gun *MG* was formed by cylindrical coaxial electrodes inserted inside a prebuncher solenoid *SG*. The cathode *C* consisted of a thin metal rod around 3 mm in diameter mounted in a high-voltage insulator *I* in the drift channel, whose walls were formed by the anode *A*. Such a small cathode diameter was selected to reduce the influence of the magnetic field at the cathode on the passage of the beam. A system for supplying short voltage pulses to the cathode, which served to control the temporarily dismantled Pierce triode gun, was used to measure the current from the cathode. It consisted of a coaxial cable which served as the secondary winding for a high-voltage pulse transformer *PT*. This cable supplied the gun current signal from the high-voltage circuits to a measuring oscilloscope *IO*. The rig forming the universal injector complex was described in greater detail in our previous study.[3]

GUN TESTS

When high-voltage pulses were supplied to the cathode using the pulse transformer, current pulses of up to 20 A were obtained from the cathode after switching on the *SG* solenoid current, which created a magnetic field near the gun. The duration of these pulses varied between 20 ns and 0.5 μs depending on the regime. The current–voltage characteristic of the gun plotted in Fig. 2 is nearly quadratic. By using the same cathode material as in the gun studied

 © 1999 American Institute of Physics

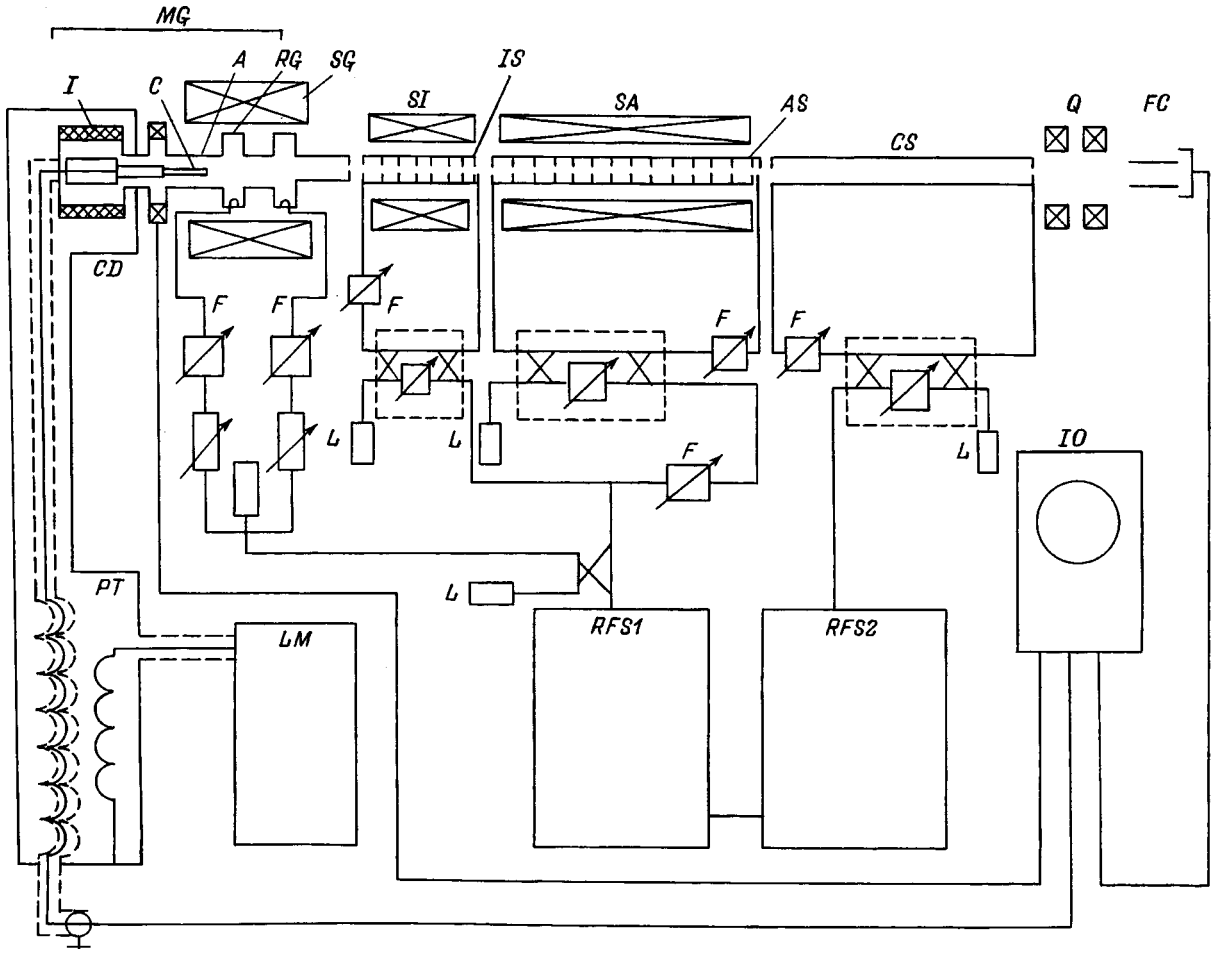

FIG. 1. Schematic of experiment: *MG* — magnetron gun with unheated cathode, *C* — cathode, *I* — high-voltage insulator, *A* — anode (drift tube), *PT* — pulse high-voltage transformer, *LM* — linear modulator, *CD* — induction current detector, *RG* — prebuncher cavities, *SG*, *SI*, *SA* — focusing solenoids of prebuncher, injector, and accelerating sections, *IS* — injector section, *AS* — accelerating section, *CS* — collimating section, its power supply system was not switched on during these experiments, this section was used as a collimator to estimate the emittance, *L* — load, *F* — phase shifters, *RFS*1 and *RFS*2 — sources of microwave power (klystron amplifiers), *Q* — quadrupole lenses, *FC* — Faraday cup, and *IO* — oscilloscope.

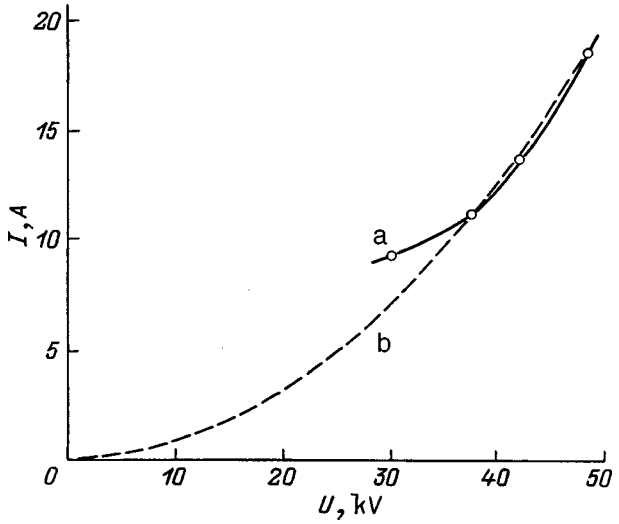

FIG. 2. Current–voltage characteristic of magnetron gun (magnetic field 220 mT): a — experimental data, b — approximation by quadratic parabola.

previously,[4] we can compare their parameters on the basis of similarity theory using the formula from Ref. 1

$$I = C\frac{U^2}{BD_c \ln\frac{D_a}{D_c}}, \tag{1}$$

where I is the beam current of the magnetron gun, D_a and D_c are the diameters of the gun cathode and anode, respectively, C is a parameter which depends on the emission properties of the cathode, U is the gun voltage, and B is the magnetic induction.

A comparison reveals that the current is higher in our case, evidently because of the stronger magnetic field. This departure from the above relationship agrees qualitatively with the dependence of the current on the magnetic field measured later,[5] where the decrease in current with increasing field is weaker than an inversely proportional dependence.

HIGH-CURRENT ACCELERATION REGIME

When rf power was supplied to the accelerating sections and the beam guide system was suitably aligned, an acceler-

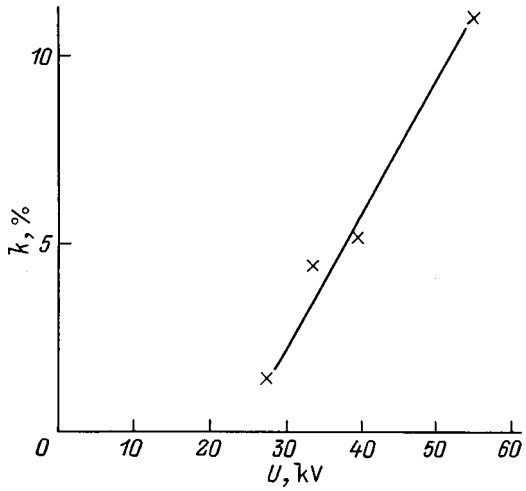

FIG. 3. Capture coefficient as a function of voltage for a conventional gun with a thermionic cathode.

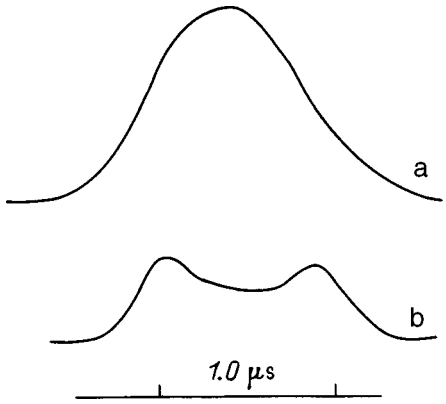

FIG. 4. Oscilloscope traces of pulses in electron cyclotron resonance regime: a — gun voltage, b – accelerated electron current at accelerator exit.

ated electron current of up to 0.5 A was obtained from the injector complex with a gun current of around 12 A and voltage around 45 kV. The low beam capture coefficient in the acceleration regime is probably related to the low injection energy, which corresponds to a low gun voltage. It can be seen from later measurements of the capture coefficient for a conventional Pierce gun with a thermionic cathode as a function of the accelerating voltage, which are plotted in Fig. 3, that the capture coefficients are similar at corresponding voltages. The duration of the relativistic electron pulse at the accelerator exit was around 20 ns. The duration of the gun current and voltage were significantly greater, 50 ns and 1 μs, respectively. The reduction in the pulse duration at the output Faraday cup is mainly attributable to two factors: the short (up to 1 ns) excitation time of the secondary emission in the gun near its peak voltage and the stored energy regime during supply of the gun and the accelerating sections. The pulse leading edge is formed by the rapid excitation of secondary emission and the duration will depend on the stored energy and the beam current, increasing as the stored energy increases and decreasing as the current increases.

Only the bunching and injector accelerating sections were switched on for the tests. The accelerated beam passed through an ''empty'' section, i.e., without a microwave power supply and focusing solenoid. In addition, this section had a low series coupling resistance[3] and for the current achieved in our case the influence of the induced field on the beam motion can be neglected. Thus, the section only plays the role of a collimator in the form of a long tube with a known aperture. In this case, for the length L and the diameter d an upper estimate of the beam emittance[6] can be obtained as the product of the angular spread d/L and the aperture d divided by π, as given by

$$\varepsilon \leqslant \frac{d^2}{\pi L}.$$

Substituting numerical values for the aperture 3.0 cm and the length of the drift tube 200 cm, we obtain the estimate 14 cm·mrad.

EXCITATION OF CURRENT IN A COLD-CATHODE MAGNETRON GUN NEAR THE ELECTRON CYCLOTRON RESONANCE FREQUENCY

During testing of the cold-cathode magnetron gun in the universal injector complex we observed that the supply of rf power to the cavity influenced the excitation of current in the gun. The gun was positioned near the prebuncher cavity in a magnetic field, as shown in Fig. 1. In this case, the microwave field through the short section of drift channel, which comprised a cutoff waveguide, could penetrate to the gun cathode. We initially observed a current from the cathode which was not captured in the acceleration regime. The current flowed for almost the entire duration of the voltage pulse at the cathode (around 1 μs, 60 kV). The current from the cathode reached 1 A and occurred in a narrow range of magnetic fields near 136–140 mT. By tuning the phase of the buncher, it was possible to obtain a current of electrons accelerated to energies higher than a few MeV at the exit from the injector complex with a magnitude of up to 20 mA and a pulse duration corresponding to that of the voltage pulse. Typical oscilloscope traces of the pulses are shown in Fig. 4. The duration of the microwave pulse exceeded that of the gun voltage pulse (around 2 μs) and the locking system was tuned so that during the voltage pulse the microwave power could be considered to be constant. Thus, in order to obtain a beam at the exit from the complex, we need to supply voltage to the gun and apply microwave power, with a quite specific magnetic field in the gun and microwave phase ratio in the prebuncher and in the bunching and accelerating sections.

Bearing in mind that the operating frequency of the complex to which the prebuncher is tuned is 2797 MHz and is close to the electron cyclotron oscillation frequency, we can make the following assumptions as to the nature of the processes. Electrons emitted from the cylindrical side surface of the cathode as a result of field emission when a voltage pulse is applied, move over trajectories close to cycloidal and return to the cathode. Under the action of a relatively weak microwave field, whose oscillation period is the same as the period of their motion, some of the electrons emitted at appropriate times acquire additional energy from the microwave field and bombard the cathode. Under the influence of

the electron bombardment the cathode begins to emit secondary electrons, which also acquire additional energy because the period of their motion matches that of the field. The power supplied to the buncher, a few kilowatts per pulse, was clearly sufficient to give a coefficient of secondary emission greater than unity. This should lead to a rapid rise in the secondary emission current from the side surface of the cathode until this is limited by the space charge field. The nonuniformity of the field near the cathode edge will then cause the incipient electron bunches to be expelled toward the buncher and the accelerating system. Acceleration of the bunches will only be possible when the accelerating phase of the field in the section coincides with the moments of expulsion of the bunches. We know that in a multipactor electron gun, which uses a similar effect, the phase dimension of the clusters is extremely small.[7] On this basis and also from observations of the phase tuning to achieve acceleration, we can postulate that the phase dimension of the cluster is also small in our case. Another argument in support of the hypothesis that the bunch has a small phase dimension may be the relatively low gun current in the cyclotron resonator regime. No direct measurements were made of the phase dimensions of the bunches because of the need to convert the complex to a different operating program.

We need to consider the slight disagreement between the working frequency of the oscillations and the cyclotron frequency. Conventionally,[8] cyclotron resonance in magnetrons is characterized by the product of the magnetic induction B and the wavelength λ. For free electrons in the absence of external electric fields the following numerical relation is obtained for resonance

$$B\lambda = 1.065 \ \ \mathrm{T\cdot cm}. \tag{2}$$

We also know that the electron efficiency of magnetrons and magnetron oscillations with a smooth anode has a dip near 1.2 T cm (Ref. 8). An increase in the product $B\lambda$ is attributable to the nonuniformity of the electric field generated by a thin cathode. The resonant frequency should depend, albeit slightly, on the voltage at the magnetron gun. A result of this dependence may be the double-humped pulse profile at the accelerator exit shown in Fig. 4, if the resonance occurs at a voltage below the peak value. This formula (2) is also valid for electron oscillations between the electrodes of a planar magnetron in the absence of space charge. Since in our case the ratio of the anode and cathode diameters is large, this product may be larger for a cylindrical magnetron. It can be calculated exactly neglecting the influence of space charge. This influence can be neglected since at the beginning of the rise in the secondary emission the space charge will not significantly influence the electron motion. The electron transit time from the cathode and back for a cylindrical magnetron will be determined using formulas given in the Appendix to the lectures in Ref. 9. The equation for the radial electron motion beginning at the cathode is

$$r = -\frac{d\Pi}{dr},$$

$$\Pi(r) = \frac{e}{m} C \ln\frac{r}{a} + \frac{\Omega^2}{8}\left(r - \frac{a^2}{r}\right)^2, \tag{3}$$

where r is the radial coordinate of the electron, e and m are the electron charge and mass, respectively,

$$C = \frac{U}{\ln\dfrac{b}{a}},$$

the constant C characterizes the electric field in terms of the applied voltage U and the radii of the anode b and cathode a, and $\Omega = eB/m$ is the cyclotron frequency.

This equation has the quadrature solution

$$\tau = \int_a^r \frac{dr}{\sqrt{-2\Pi(r)}}. \tag{4}$$

The period of the motion was calculated as twice the time $2\tau_{\max}$ for motion as far as the maximum radius $r_{\max}$, for which $\Pi(r_{\max})=0$. Substituting numerical values for the geometric dimensions of the gun $a = 0.15$ cm and $b = 1.5$ cm and experimental data (the average magnetic field induction from the interval in which emission is excited $B = 138$ mT and the gun voltage $U = 30$ kV), computer calculations give the frequency $F = 2.967$ GHz, which agrees satisfactorily with the working frequency of 2.797 GHz.

CONCLUSIONS

1. The main characteristics and operating stability of a magnetron gun with a cold secondary-emission cathode have been confirmed in the frequency regime up to and including 50 Hz.

2. A relativistic electron beam has been obtained at the exit from a resonant accelerator having a magnetron gun as the electron source with capture no worse than that for conventional guns.

3. It has been shown that an accelerating complex with an unheated magnetron gun can naturally form high-current nanosecond pulses of relativistic electrons with microsecond high-voltage pulses at the gun. This can be achieved because a large secondary-emission current is excited rapidly by the high voltage under conditions where energy is stored in the accelerating structure.

4. It has been established that when the resonator is excited near the magnetron gun and with magnetic fields corresponding to electron cyclotron resonance, injection is observed in the accelerator and relativistic electrons are observed at its exit over the entire time of application of voltage to the gun. Satisfactory agreement is obtained between the calculated electron oscillation frequency in the gun and the working frequency of the microwave field in the accelerator.

5. A distinguishing feature of the electron cyclotron resonance regime is that the current is an order of magnitude lower and the current at the accelerator exit (until this completely disappears) depends strongly on the phase of the rf field exciting the current.

6. The estimated beam emittance (14 cm·mrad) can solve the problem of the practical application of accelerators with this type of electron source.

In conclusion, the authors would like to thank A. N. Opanasenko for assistance with the calculations and stimulating discussions.

[1] S. A. Cherenshchikov, AIP Conf. Proc. **337**, 350 (1994).
[2] G. P. Antropov, N. A. Arkhangelov, B. V. Bekhtev *et al.*, Vopr. At. Nauki Tekh. Ser. Tekh. Fiz. Éksp., No. 2(23), 3–5 (1985).
[3] G. M. Ivanov, V. I. Kurilko, L. A. Makhnenko *et al.*, Zh. Tekh. Fiz **64**(4), 115 (1994) [Tech. Phys. **39**, 411 (1994)].
[4] S. A. Cherenshchikov, in *Proceedings of the 13th Meeting on Charged Particle Accelerators* [in Russian], Dubna (1993), Vol. 2, pp. 142–144.
[5] S. A. Cherenshchikov, Bull. Am. Phys. Soc. **40**, 1166 (1995).
[6] V. A. Moskalev, G. I. Sergeev, and V. G. Shestakov, *Measurement of the Parameters of Charged Particle Beams* [in Russian], Atomizdat, Moscow (1980), 160 pp.
[7] W. J. Gallagher, Proc. IEEE **57**, 94 (1969).
[8] I. V. Lebedev, *Microwave Technology and Devices. Vol. II, Microwave Electric-Vacuum Devices* [in Russian], Vysshaya Shkola, Moscow (1972), 376 pp.
[9] L. A. Vaĭnshteĭn and V. A. Solntsev, *Lectures on Microwave Electronics* [in Russian], Sovet-skoe Radio, Moscow (1973), 400 pp.

Translated by R. M. Durham

Initiation of a low-pressure glow discharge in a plasma electron source with a ribbon beam

V. Ya. Martens

Stavropol' State Technical University, 355038 Stavropol', Russia
(Submitted May 18, 1998)
Zh. Tekh. Fiz. **69**, 135–137 (July 1999)

Results are presented from an experimental investigation of a low-pressure glow discharge with a wedge-shaped hollow cathode in a plasma electron source, where this discharge is initiated by reflex and magnetron discharges. © *1999 American Institute of Physics.* [S1063-7842(99)02907-4]

One of the most effective methods for the heat treatment of semiconductor materials and devices is the electron beam method.[1] In particular, cw ribbon electron beams of width 1–3 mm with power densities up to 1 kW/cm^2 and electron energies less than 10 keV are suitable for recrystallizing polycrystalline silicon on an insulator.[2,3] In order to deliver these beams with a small angle of convergence, the emitters should provide emission current densities up to 100 mA/cm^2. It is advisable to solve this problem using cw plasma emitters in which a plasma of the required density and homogeneity is generated in special gas-discharge structures. In this type of plasma electron source a linear emission channel may be fabricated in the anode or cathode electrodes of the plasma generator. In this case, the comparatively large channel cross section prevents a substantial pressure drop from being established between the working volume of the emitting plasma generator and the accelerating gap for the emitted electrons. Under these conditions the electrode structure of the plasma electron source should contain special initiating systems in the plasma generator to facilitate the ignition of a discharge at low pressure.

Here we present results of an investigation of a low-pressure discharge in an electrode structure with a wedge-shaped hollow cathode, where this discharge is initiated by reflex and magnetron discharges (Figs. 1a and 1b). In the first case, a reflex (auxiliary) discharge is initiated in a discharge cell formed by a planar cathode *1*, an anode *2*, and the planar outer part of a wedge-shaped hollow cathode *3*. Electrons from the reflex discharge penetrate through a 3 mm diameter aperture to the inside of the wedge-shaped hollow cathode and initiate the main discharge. The minimum required auxiliary discharge current is 50 mA with a working gas (argon) flow rate of 4.5 m^{3}mPa/s. A 60×2.5 mm emission slit is provided in the main-discharge anode *4*. After the main discharge has been ignited, the auxiliary discharge is quenched by removing the voltage U_{ad}. For an argon flow rate of 4.5 m^{3}mPa/s, the pressure in the accelerating gap was $\sim 10^{-2}$ Pa and that in the wedge-shaped cathode was $\sim 10^{-1}$ Pa. A wedge-shaped hollow cathode was selected for the following reasons. The primary electrons leaving the cathode surface at a certain initial velocity as a result of ion-electron emission and accelerated in the cathode potential drop reach the cathode potential drop at the opposite part of the hollow cathode and enter the cathode, not along the normal but at a certain angle to it since the surfaces of the wedge-shaped hollow cathode are not parallel. The normal velocity component of these electrons is such that they cannot overcome the cathode potential drop and reach the opposite part of the hollow cathode, and they are reflected by the electric field of the cathode potential drop without energy losses. This electron reflection can be repeated many times and in consequence, the average lifetime of the primary electrons and thus the gas ionization efficiency are increased substantially. This effect may mean that a discharge can exist at low pressure without being sustained by electrons injected from an auxiliary discharge.

Initiation of the main discharge by a magnetron discharge takes place as follows (Fig. 1b). A switch *S* is set to position *a*, whereupon an auxiliary magnetron discharge is ignited between the cathodes *1* and *3*, its anode being the wedge-shaped hollow cathode *3*. The switch *S* is then switched to position *p* and the main discharge is ignited between the wedge-shaped hollow cathode formed by electrodes *1* and 3 and the anode *4*. During the switching time no magnetron discharge burns since no voltage is taken from the anode *4* and at this time it is the anode for the magnetron discharge. The minimum required magnetron discharge current for which the main discharge can be initiated is higher than that in the first initiation system and is 200 mA for an argon flow rate of 6.8 m^{3}mPa/s.

Figure 2 shows current–voltage characteristics of a discharge with a wedge-shaped hollow cathode initiated by reflex (a) and magnetron (b) discharges. We know[4] that a low-voltage hollow-cathode discharge can only burn stably at currents exceeding a certain critical level. A reduction in the discharge current specifically leads to expansion of the cathode potential drop zone, whose opposite sections can ultimately overlap inside the cavity. This is usually accompanied by an abrupt increase in the discharge burning voltage or the discharge is quenched, as occurs in our case. The points on the extreme left of the experimental curves in Fig. 2 correspond to the critical values of the main discharge current. As the gas flow rate increases and the discharge burning voltage U_d decreases, the critical currents decrease.

 © 1999 American Institute of Physics

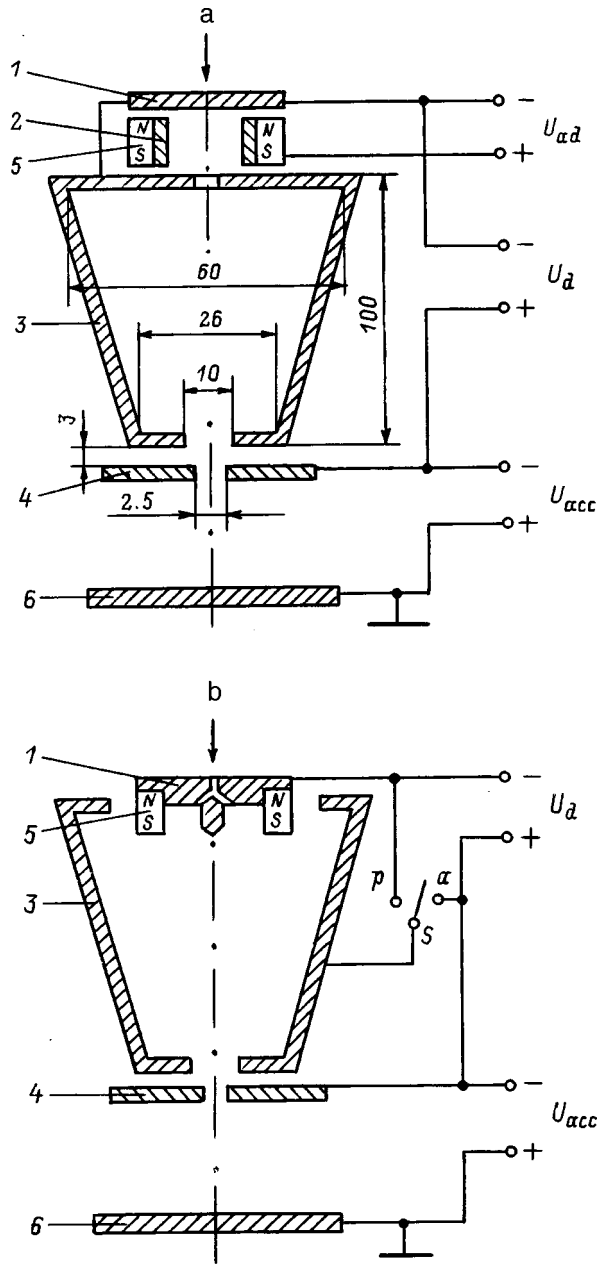

FIG. 1. Schematic diagrams of a plasma electron source with a ribbon beam, where the main discharge is initiated by reflex (a) and magnetron (b) discharges: *1* — planar cathode, *2* — auxiliary discharge anode, *3* — wedge-shaped hollow cathode, *4* — main discharge anode, *5* — permanent ring magnet, and *6* — collector.

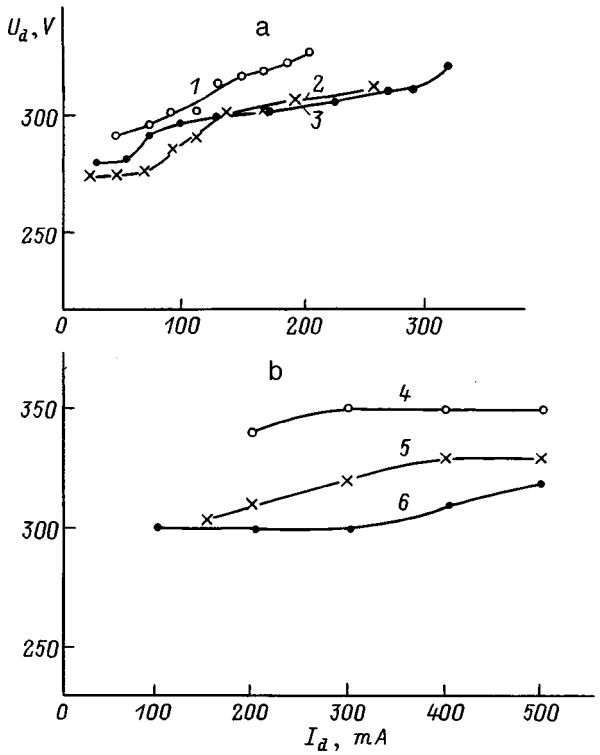

FIG. 2. Current–voltage characteristics of discharge with wedge-shaped hollow cathode initiated by reflex (a) and magnetron (b) discharges. Working gas — argon; Gas flow rate Q, m^3mPa/s: *1* — 3.4, *2* — 4.5, *3* — 5.6, *4* — 5.1, *5* — 6.8, and *6* — 8.4.

If the main discharges with different systems of initiation are compared, attention is drawn to the fact that in the second case (Fig. 1b), even at higher working gas flow rates, the critical currents are higher than in the first case (Fig. 1a). This is possibly because of differences in the magnitude and configuration of the magnetic field in the upper part of the wedge-shaped hollow cathode.

Using these plasma emitters we obtained emission currents I_e up to 200 mA with an emission efficiency $\alpha = 0.5–0.7$ ($\alpha = I_e/I_d$, where I_d is the main discharge current) and economy $H = 1.7–2.3$ mA/W. A comparison between the characteristics of the main discharge initiated by reflex and magnetron discharges showed that in the first case, the critical currents are lower and a lower working gas flow rate is required although the design and power supply system are slightly more complex. The choice of a particular variant to develop a plasma electron source with a ribbon beam depends on the specific requirements for the source.

The author would like to thank Yu. A. Burachevskiĭ for assistance with the experiment.

[1] R. A. McMahon, H. Ahmed, D. J. Godfrey *et al.*, Microelectron. J. **15**(2), 5 (1984).
[2] L. R. Thompson, J. A. Knapp, C. A. Moore *et al.*, Mater. Res. Soc. Symp. Proc. **107**, 195 (1988).
[3] C. A. Moore, J. D. Meyer, J. T. Fukumoto *et al.*, Mater. Res. Soc. Symp. Proc. **107**, 207 (1988).
[4] S. P. Bugaev, Yu. E. Kreĭndel', and P. M. Shchanin, *Large Cross-Section Electron Beams* [in Russian], Energoatomizdat, Moscow (1984), p. 43.

Translated by R. M. Durham

Enhancement of the measurement sensitivity in real-time optical image processing for objects with a periodic structure

A. M. Lyalikov

Ya. Kupala Grodno State University, 230023 Grodno, Belarus
(Submitted July 23, 1998)
Zh. Tekh. Fiz. **69**, 138–139 (July 1999)

A possible method of real-time visualization of large-scale objects giving enhanced measurement sensitivity using incoherent illumination is reported. © *1999 American Institute of Physics.* [S1063-7842(99)03007-X]

Displacements, deformations, macroscopic defects, and also surface relief can be visualized by using coatings in the form of grids or retrodyes applied artificially to the object being studied[1,2] and also by using the natural surface structure such as various types of inclusions and characteristic formations.[3,4] These optical methods of deciphering information on the state and parameters of an object are similar to speckle photography and holographic interferometry but unlike these methods, they can be used to study objects of appreciable dimensions such as multistorey buildings merely by using natural sunlight. When objects with a periodic surface structure are studied, the sensitivity of the measurements can only be regulated at the stage of optical processing of the photographs using an image of these objects.[5]

In the present paper it is shown that the sensitivity of the measurements can be enhanced by real-time incoherent optical processing of images of objects having a periodic structure. The use of incoherent illumination means that it is possible to use a method of enhancing the sensitivity of the measurements to study large-scale objects by illuminating them with sources of white light or simply with sunlight.

We postulate that an optical system produces an image of an object with a periodic surface structure. Examples of these objects are slit-mask amplitude screens[3] and various types of brickwork or tiled structures.[4] For almost all these objects the illuminance distribution in the image of the periodic structure, where the y axis is oriented parallel to the image, may be represented as a Fourier series expansion

$$I(x,y)\sim\sum_{n=1}^{+\infty} a_n \cos\left[\frac{2\pi nx}{T}+n\varphi(x,y)\right], \tag{1}$$

where a_n are coefficients and T is the period of the image of the structure.

Depending on the object the function $\varphi(x,y)$ may determine the deviation of the surface profile from a plane[3] or simply the orientation of the object in space.[4] The task for investigating this type of object is to visualize the behavior of $\varphi(x,y)$.

Unlike earlier methods,[5,6] to enhance the sensitivity of measurements to visualize $\varphi(x,y)$ in real time, we suggest that an image of the periodic structure of an object should be projected onto a standard transparency in the form of an amplitude diffraction grating with a carrier-band period $T_1=T/m$, where $m=2,3,4,\ldots$. This approach was used earlier in a Talbot interferometer to enhance the sensitivity of measurements to visualize phase objects[7] and also to visualize the spatial orientation of photographed buildings, but in this last case only at the optical processing stage of the photographs.[8]

If the spatial orientation of planar surfaces such as buildings needs to be visualized in real time using the technique proposed in Ref. 6, a normal amplitude grating with a corresponding reduction in the period of the carrier bands will be used as a standard transparency to enhance the sensitivity of the measurements. However, if the surface profile of the object differs from planar and needs to be visualized in real time with the measurement sensitivity enhanced m times, for example, its macrodefects compared with a standard object, the standard transparency should be a photograph with the amplitude transmission

$$\tau(x,y)\sim\cos^2\left[\frac{\pi mx}{T}+\frac{m}{2}\varphi_0(x,y)\right], \tag{2}$$

where $\varphi_0(x,y)$ determines the deviation of the surface profile of the standard object from a plane.

For this purpose we record a photograph of a standard object under nonlinear conditions and then copy it in a photograph copying system[9] which isolates the waves diffracted in the $\pm l$th diffraction orders where $l=m/2$. Recording the photograph of the standard object under nonlinear conditions ensures that the highest diffraction orders ($|l|>1$) appear on the image of the periodic structure in the nonlinear photograph. When the photograph is copied, the magnification of the optical system should be exactly one, and the recording conditions should preferably be linear to eliminate higher harmonics which complicate the treatment of the visualized pattern.

When the object is observed through a standard transparency of the type (2), the illuminance distribution will be determined by the product $I(x,y)\tau(x,y)$. It can be shown that in order to describe the pattern of moiré fringes which appear on the combined image of the object with the standard transparency, we can confine ourselves to the product

 © 1999 American Institute of Physics

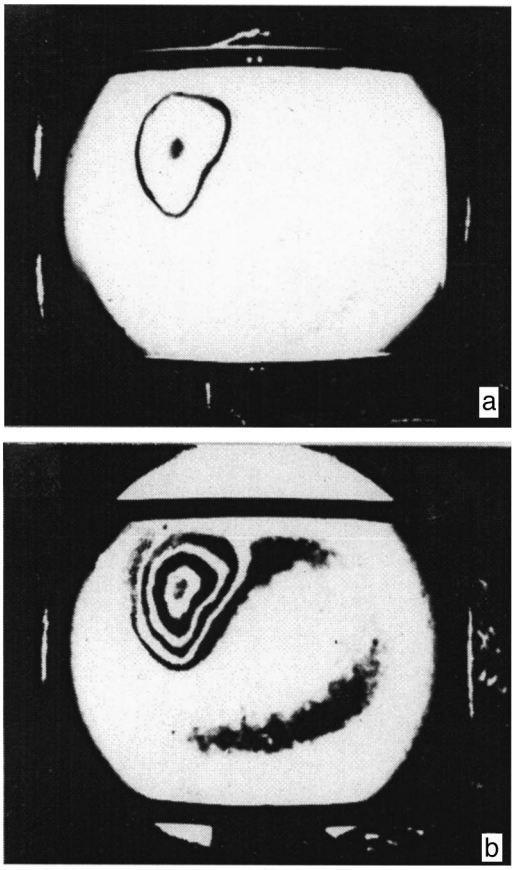

FIG. 1. Moiré patterns visualizing a surface macrodefect in the slit mask of a television picture tube without enhanced measurement sensitivity (a) and with doubled sensitivity (b).

$$\frac{1}{2}a_m \cos\left\{\frac{\pi m x}{T}[\varphi_0(x,y)+\varphi(x,y)]\right\} \times \cos\left\{\frac{\pi m x}{T}[\varphi_0(x,y)-\varphi(x,y)]\right\}. \quad (3)$$

Regions for which the second cosine in Eq. (3) vanishes will correspond to poor visibility of the image of the periodic structure of the object. In this case, the equations for the family of moiré fringes are

$$\frac{\pi x}{T}[\varphi_0(x,y)-\varphi(x,y)]=(2N+1)\frac{\pi}{2m}, \quad (4)$$

where $N=0,1,2,\ldots$.

The factor $1/m$ on the right-hand side of Eq. (4) indicates that for the visualization of a macrodefect determined by the difference $[\varphi_0(x,y)-\varphi(x,y)]$, the sensitivity of the measurements increased m times compared with the usual moiré pattern.[4–6]

Slit masks of television picture tubes were used as the object to confirm this method experimentally. The deviations of the surface macrorelief of the mask from the standard, which characterize the surface macrodefects, were visualized using a system described in the first part of Ref. 5. It should be noted that the surface profile of these slit masks differed from planar.[3] Figure 1 shows a moiré pattern which visualizes the surface macrodefect in real time (a) and a moiré pattern with the measurement sensitivity doubled (b). The standard transparency was a photograph of a standard mask copied in the first positive or negative diffraction orders. Figure 1b is more informative because of the enhanced measurement sensitivity.

This work was supported by the Ministry of Education of the Republic of Belarus.

[1] N. Wadsworth, M. Marchant, and B. Billing, Opt. Laser Technol. **5**(3), 119 (1973).
[2] N. G. Vlasov and A. E. Shtan'ko, *Proceedings of the Ninth All-Union Holography School* [in Russian], LIYaF (1977), pp. 256–266.
[3] A. M. Lyalikov, Opt. Zh. **61**(5), 23 (1994) [J. Opt. Technol. **61**, 376 (1994)] .
[4] A. M. Lyalikov, Opt. Zh. **62**(8), 78 (1995) [J. Opt. Technol. **62**, 563 (1995)].
[5] A. M. Lyalikov, Opt. Zh. **62**(1), 28 (1995) [J. Opt. Technol. **62**, 26 (1995)].
[6] A. M. Lyalikov, Opt. Zh. **64**(1), 82 (1997) [J. Opt. Technol. **64**, 68 (1997)].
[7] K. Patorski, Appl. Opt. **27**, 3875 (1988).
[8] A. M. Lyalikov, Opt. Zh. **65**(3), 54 (1998) [J. Opt. Technol. **65**, 220 (1998)].
[9] A. M. Lyalikov, Opt. Zh. **63**(5), 73 (1996) [J. Opt. Technol. **63**, 391 (1996)].

Translated by R. M. Durham

Increase in the correlation length of nonmonochromatic radiation during propagation in a single-mode optical fiber containing random inhomogeneities and the influence of this increase on the operation of a fiber ring interferometer

G. B. Malykin and V. I. Pozdnyakova

Institute of Applied Physics, Russian Academy of Sciences, 603699 Nizhniĭ Novgorod, Russia
(Submitted July 23, 1998)
Zh. Tekh. Fiz. **69**, 140–143 (July 1999)

An analysis is made of the conditions under which the coupling of orthogonal polarization modes at random inhomogeneities in single-mode optical fibers leads to an increase in the correlation length of a source of nonmonochromatic radiation. It is shown that when long-base fiber ring interferometers with a single-mode fiber ring system possessing weak linear birefringence are used, the correlation length of the nonmonochromatic radiation at the interferometer exit is increased, which means that the interference pattern can have satisfactory visibility even when there is an appreciable difference between the interferometer arms as a result of the Sagnac effect due to the Earth's rotation. The calculations were made by mathematical modeling of random inhomogeneities in the fiber. © *1999 American Institute of Physics.* [S1063-7842(99)03107-4]

The correlation length of nonmonochromatic radiation is one of the most important parameters in optical interferometry, especially fiber interferometry. For some applications it is advantageous to have a short correlation length and for others a long correlation length. For instance, in fiber optical tomography[1] the spatial resolution of adjacent layers of tissue is mainly determined by the correlation length and in order to improve the resolution it is necessary to have the shortest possible correlation length. Conversely, in Michelson and Mach–Zehnder interferometers with different arm lengths, the correlation length must be increased otherwise the visibility of the interference pattern may very poor.

A difference between the interferometer arms is an inevitable consequence of their inadequate balancing. We cite an example where the counterpropagating waves in a fiber ring interferometer (FRI) cover essentially different paths. Scully *et al.*[2] suggested measuring the postulated anisotropy of the velocity of light using an FRI fabricated using a single-mode optical fiber with a large-area ring system. Estimates show that in order to achieve the required accuracy, the area of the FRI ring system should be at least several tens of square kilometers. Since the FRI rotates with the Earth, the Sagnac effect leads to an appreciable difference in the optical lengths for the counterpropagating waves in the ring system, of the order of a few hundred micron, and visibility of the interference pattern at the interferometer exit will be very poor. A deterioration in the visibility of the interference at the exit of an FRI with a wide-band radiation source was observed by Burns and Moeller,[3] who studied an FRI with a comparatively small-area ring system but having a high angular rotation speed.

The aim of the present study is to show that after propagating through a fairly long section of single-mode fiber with weak linear birefringence, the correlation length of nonmonochromatic radiation may increase considerably as a result of the coupling of polarization modes at fiber inhomogeneities. We also study the influence of an increase in the correlation length on the operation of an FRI with a single-mode fiber ring system possessing weak linear birefringence. The physical meaning of this effect can be explained as follows. When the condition $L \gg h^{-1}$ is satisfied, where L is the fiber length in the FRI ring system and h is the polarization conservation parameter for the single-mode fiber[4] (the value of h^{-1} characterizes the fiber length at which effective energy exchange takes place between the polarization modes), radiation propagating along the slow and fast axes of birefringence of the fiber is repeatedly transferred from one polarization mode to another. As a result, the spectral characteristics of the radiation at the FRI exit differ substantially from those at the entrance: if the input radiation spectrum was Gaussian, at the exit it becomes extremely jagged (although conserving its previous width), which increases the correlation length.[5] In other words, the single-mode fiber is a two-channel system and each channel has a single polarization mode. In the absence of any inhomogeneities in the fiber, these channels are independent, i.e., no energy is exchanged between them, and the spectral characteristics of the radiation in each polarization mode will be the same as those at the FRI exit. However, in the presence of random inhomogeneities, energy exchange takes place between the two channels, fast and slow polarization modes, which changes the profile of the spectrum in each polarization mode. In the absence of dichroism the spectrum of the total intensity of both polarization modes for a wave propagating in one particular direction in the FRI ring system remains constant and is the same as the radiation spectrum on entry to the FRI. Hence, this effect is a consequence of the fact that the fiber used to fabricate the FRI ring system is a two-channel system in which coupling takes place between the two channels as a result of random inhomogeneities.

 © 1999 American Institute of Physics

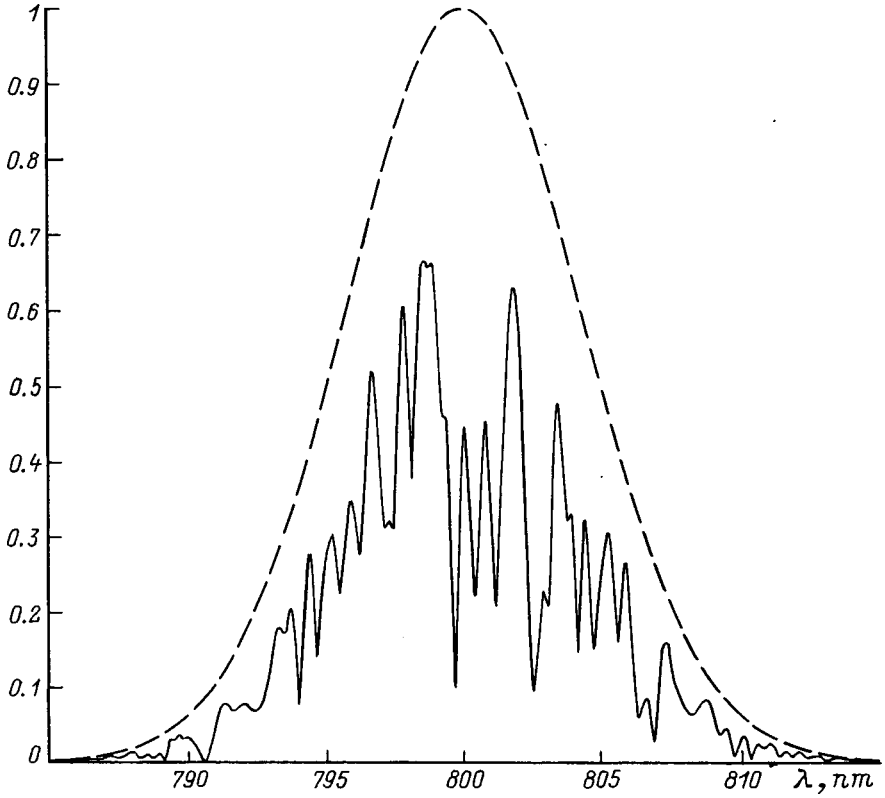

FIG. 1. Radiation spectrum at FRI entrance (dashed curve) and exit (solid curve) for $\varepsilon = 10^{-4}$.

In order to calculate the correlation function of nonmonochromatic radiation after passage through the single-mode fiber, we use a model proposed in our earlier study[6] to describe random inhomogeneities in a single-mode fiber. This model assumes that the axes of linear birefringence of the fiber are randomly twisted and that the entire length of the fiber may be divided into sections of random length whose twist is random but constant. Regardless of the linear birefringence of the fiber, the length of the random sections has a Poisson distribution with an average of 2.5 cm and the random twist is uniformly distributed in the range ± 2 rad/m. As we showed in Ref. 6, calculations of the parameter h as a function of the linear birefringence of a single-mode fiber based on this model of random inhomogeneities give good agreement with experiments carried out by various authors using different types of single-mode fibers. It should be noted that in order to ensure that the envelope of the radiation spectrum at the FRI exit is jagged, the twist of the axes of birefringence over the fiber length need not be random and may vary periodically. In this last case, however, the dependence of the visibility of the interference pattern at the FRI exit on the difference between the optical length of the arms ΔL will decrease periodically to zero. Figure 1 shows the radiation spectrum at the FRI entrance and exit. Note that in the case $\varepsilon \ll 1$, when the polarization nonreciprocity of the FRI is small,[7,8] the spectra for the counterpropagating waves at the FRI exit are almost the same. We also note that in the absence of a polarizer ($\varepsilon = 1$), the spectrum for the orthogonal polarizations in each of the counterpropagating waves is mutually complementary and, as we have already noted, the spectrum of the total intensity of both polarizations of each counterpropagating wave is the same as the radiation spectrum at the FRI entrance. The correlation length of nonmonochromatic radiation at the FRI exit without a polarizer was not considered in the present study.

In order to determine the correlation length, we calculated the visibility of the interference pattern at the FRI exit as a function of the optical path difference of the FRI ring system ΔL for the counterpropagating waves

$$\gamma = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \tag{1}$$

where I is the radiation intensity at the FRI exit.

The value of γ was calculated for two cases: 1) the FRI ring system was made from a fiber without random inhomogeneities, 2) the FRI ring system was made from a real fiber with random inhomogeneities. We write an expression to calculate the radiation intensity at the FRI exit

$$I = |E_x^+ + E_x^-|^2 + |E_y^+ + E_y^-|^2, \tag{2}$$

where E_x^+, E_y^+, E_x^-, and E_y^- are the components of the electric fields of the counterpropagating waves at the FRI exit.

The calculations were made for an FRI ring system 20 km long, made of a single-mode fiber for which the difference between the refractive indices in the slow and fast axes of linear birefringence was $\Delta n = 10^{-6}$, using a 0.8 μm radiation source and linewidths of 10 and 1 nm. The extinction coefficients of the polarizer ε were 10^{-4} and the azimuths of the axes of linear birefringence at the FRI entrances α_1 and α_2 coincided with the direction of transmission of the polarizer. The dependence of I on ΔL constructed using expression (2) resembles damped oscillations with a period equal to the wavelength against the background of a constant

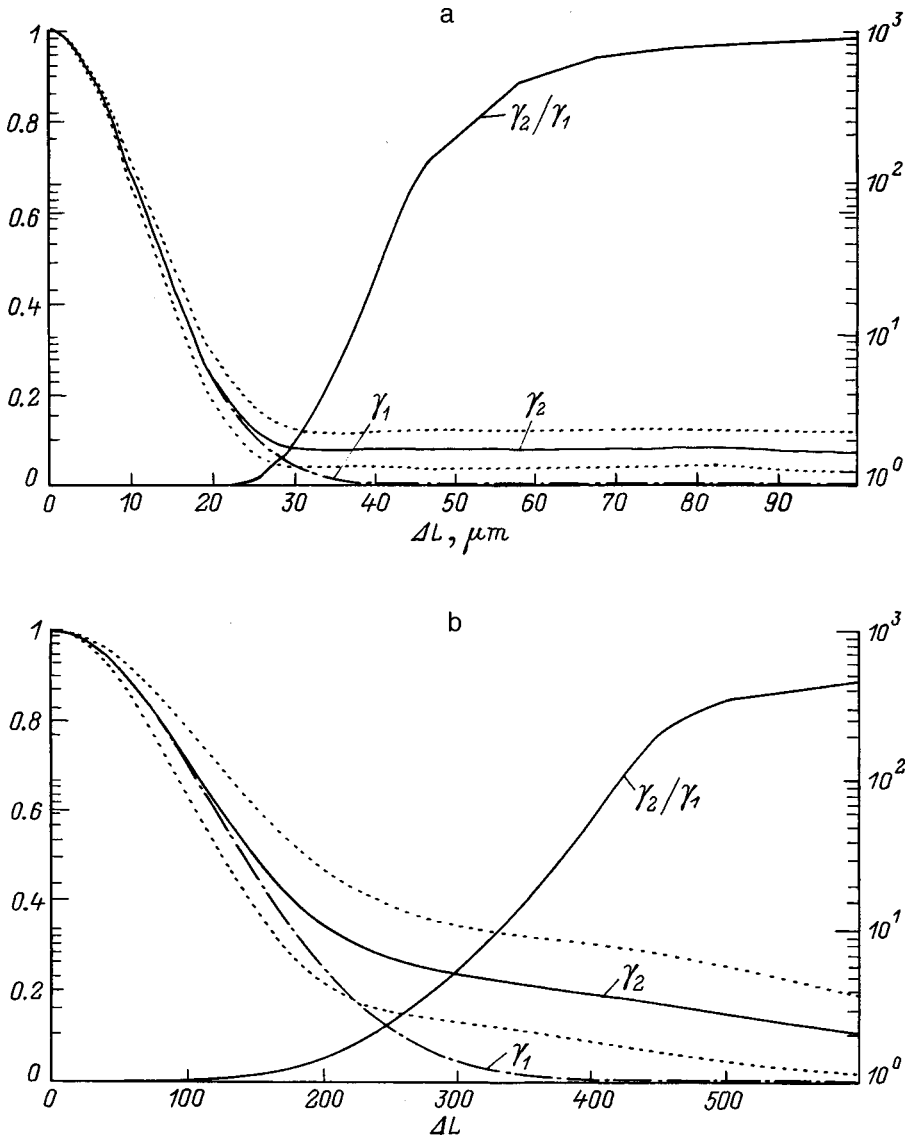

FIG. 2. Visibility of interference pattern at FRI exit as a function of the optical path difference of the counterpropagating waves caused by the Sagnac effect for an ideal single-mode fiber without random inhomogeneities (γ_1 — dot-and-dash line) and for a real fiber with random inhomogeneities (γ_2 — solid curve, dashed curve — range of possible γ_2 values for various realizations of random inhomogeneities in a the fiber ring system). The length of the ring system is 21 km, the average wavelength of the radiation source is $\lambda_0 = 0.8\ \mu$m, and the linewidth is $\Delta\lambda = 10$ (a) and 1 nm (b).

pedestal. The values of I_{max} and I_{min} were defined as the maximum and minimum of the curve per period. The results of the numerical simulation showed that the radiation spectrum at the FRI exit was highly jagged, as was predicted.

Figure 2 gives the dependence $\gamma(\Delta L)$. It can be seen from Fig. 2a that when the bandwidth of the radiation source is 10 nm for optical length differences up to $\Delta L \sim 25\ \mu$m, the visibility of the interference pattern at the FRI exit does not depend on the presence of random inhomogeneities in the fiber. For large ΔL the visibility of the interference pattern at the exit of an FRI with a fiber ring system containing random inhomogeneities is substantially higher (approximately one thousand times higher) than that for a fiber without random inhomogeneities. When the bandwidth is 1 nm (Fig. 2b) as far as $\Delta L \sim 220\ \mu$m the visibility of the interference pattern when random inhomogeneities are present in the fiber is slightly lower than that without inhomogeneities, but then becomes substantially higher (approximately five hundred times higher).

The main result of this study is that we have shown by means of a numerical simulation of random inhomogeneities in a single-mode fiber that the correlation length of radiation after propagating through an FRI ring system made of a single-mode fiber with random inhomogeneities is considerably greater than the correlation length without random inhomogeneities. This means that sources of nonmonochromatic radiation can be used in unequal-arm fiber interferometers fabricated using single-mode fibers with weakly linear birefringence and interference patterns of satisfactory visibility can be obtained. The presence of random inhomogeneities in the fiber ring system of an FRI leads to the appearance of an additional phase drift of the interference pattern at the FRI exit, unrelated to the rotation, whose order of magnitude does not exceed ε, as has been shown by the results of theoretical[7,8] and numerical calculations.[9] This imposes the constraint that the fiber temperature must be kept stable otherwise the phase of the interference of the counterpropagating waves will vary with temperature. Thus, the presence of random inhomogeneities in the single-mode fiber on the one hand causes some deterioration in the FRI characteristics, since it leads to an additional zero shift but on the other hand, it can produce an interference pattern of satisfac-

tory visibility in the presence of nonreciprocal effects which result in a difference between the optical paths for the counterpropagating waves.

In conclusion, the author would like to thank I. A. Shereshevskiĭ for discussions of the results.

This work was partially supported by RFBR Grant No. 96-15-96742.

[1] V. M. Gelikonov, G. V. Gelikonov, N. D. Gladkova *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **61**, 149 (1995) [JETP Lett. **61**, 158 (1995)].
[2] M. O. Scully, M. S. Zubairy, and M. P. Haugan, Phys. Rev. A **24**, 2009 (1981).
[3] W. K. Burns and R. P. Moeller, J. Lightwave Technol. **LT-5**, 1024 (1987).
[4] S. C. Rashleigh, J. Lightwave Technol. **LT-1**, 312 (1983).
[5] M. Born and E. Wolf, *Principles of Optics*, 4th ed. [Pergamon Press, Oxford, 1969; Nauka, Moscow, 1973, 719 pp].
[6] G. B. Malykin, V. I. Pozdnyakova, and I. A. Shereshevskiĭ, Opt. Spektrosk. **83**, 843 (1997) [Opt. Spectrosc. **3**, 780 (1997)].
[7] S. M. Kozel, V. N. Listvin, S. V. Shatalin, and R. V. Yushkaĭtis, Opt. Spektrosk. **61**, 1295 (1986) [Opt. Spectrosc. **61**, 814 (1986)].
[8] G. B. Malykin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **34**, 817 (1991).
[9] G. B. Malykin and V. I. Pozdnyakova, Opt. Spektrosk. **84**, 145 (1998) [Opt. Spectrosc. **84**, 131 (1998)].

Translated by R. M. Durham