Nonequilibrium phase transitions and a measure of the ordering of motion in a relaxational auto-oscillatory system

A. S. Rudyĭ

P. G. Demidov Yaroslavl State University, 150000 Yaroslavl, Russia
(Submitted April 22, 1996)
Zh. Tekh. Fiz. **68**, 3–10 (January 1998)

An open thermodynamic system is considered. Its state is determined by a one-dimensional temperature field $T(x,t)$ and a heat flux $I(x,t)$, on which nonlocal nonlinear external feedback is imposed. The change in the Clausius entropy and its production in response to the excitation of auto-oscillations in the system is calculated on the basis of the results of a dynamical analysis. The use of the relative increment of the total entropy of the system, normalized to the total entropy production, as a measure of the ordering of motion is proposed. The analogy between the formalism of the Andronov–Hopf bifurcation theorem and the Landau–Ginzburg theory of phase transitions is traced in the second part of the paper. It is shown that in the initial stage of auto-oscillations the phase matching condition, which determines the amplitude of the oscillations within the Andronov–Hopf formalism, becomes meaningless because of fluctuations. In this case the amplitude should be regarded as an order parameter, and the actual state of the system should be determined from the requirement of a minimum for the nonlinear part of the increment of entropy production. The proposed approach permits a description of transient regimes and qualitatively accounts for ''soft'' and ''hard'' bifurcations as being due to nonequilibrium first- and second-order phase transitions. © *1998 American Institute of Physics.* [S1063-7842(98)00101-9]

1. One of the problems in the theory of the self-organization of nonlinear processes in open dynamical systems is selecting fundamental quantitative characteristics of the ordering of motion.[1] A comparative analysis of different ordering criteria is given in Ref. 2. In particular, the measure of order for Thomson systems (oscillatory systems with dissipative nonlinearity) is the Boltzmann entropy

$$S(t)=-k_B\int f(X,t)\ln f(X,t)dX+S_0,$$

where $f(X,t)$ is the single-particle distribution function and X denotes the set of variables which determine the state of the system.[3]

As was shown in Ref. 4, the energy-normalized entropy of an auto-oscillatory system decreases upon passage through the generation threshold and can serve as a criterion of self-organization (Klimontovich's S-theorem).

In open dynamical systems self-organization can take the form of auto-oscillations. In studies of self-organization auto-oscillations have been treated predominantly as ordered states of Thomson systems. Considerably less attention has been focused on the investigation of auto-oscillations in distributed relaxational systems, and open relaxational systems with nondissipative nonlinearity have scarcely been considered. At the same time, the self-organization processes in such systems call for a detailed treatment because of specific features of the methods used to describe them.

In the overwhelming majority of cases the macroscopic level is sufficient for describing the motion of an open relaxational system, i.e., its mathematical model has the form of a deterministic system of parabolic equations. Let us consider the simplest case, in which the motion in a dynamical system is determined only by the diffusion of heat. It would be natural to calculate the entropy of such a deterministic system as the Clausius entropy $S=\int(\delta Q/T)+S_0$. This raises the question of whether the Clausius entropy for a relaxational system is of the same order of magnitude as the Boltzmann entropy for a Thomson system. The evolution of the entropy and its production in a distributed auto-oscillatory relaxational system in response to variation of a control parameter is investigated below. The calculations presented are based on the results of an analysis[5] of a mathematical model of the system.

2. A convenient object for studying self-organization processes in relaxational systems is the mathematical model of the temperature-stabilization system shown schematically in Fig. 1. In such devices all the elements, except the controllable object (dielectric medium *1* in the present case) can be considered localized, and the mathematical model of the system can be reduced to a boundary-value nonstationary heat-conduction problem. For simplicity we assume that the nonuniform temperature field of the system is one-dimensional, and we write the boundary-value heat-conduction problem in the form

$$\dot{T}(x,t)=aT''(x,t), \tag{1}$$

$$T(0,t)=0, \tag{2}$$

$$T'(x,t)|_{x=1}=f[\beta-T(x_0,t)]\cdot\sigma[\beta-T(x_0,t)]. \tag{3}$$

It must be stressed that in system (1)–(3) self-organization is possible only as a result of the self-interaction of a single degree of freedom involving external feedback. It is simulated by the nonlocal boundary condition (3), which assigns the heat flux at the boundary $x=1$ as a function of the temperature at the internal point x_0. Here $f[\beta-T(x_0,t)]$ is a certain smooth function of the temperature having a concave course over its entire domain, which is a necessary condition for the existence of stationary states; σ is the

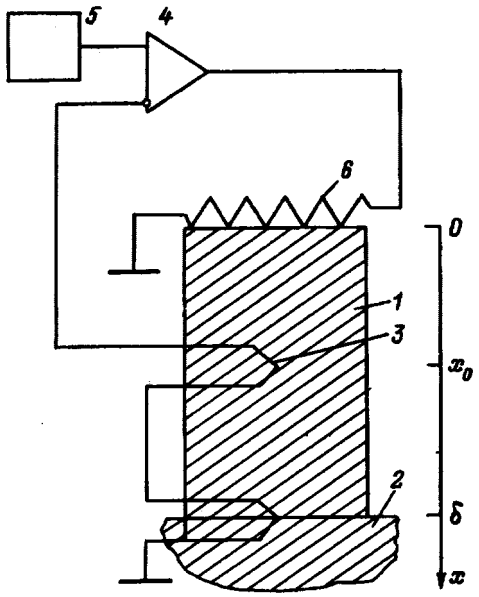

FIG. 1. Temperature-stabilization system: *1*—controlled object, *2*—thermostat, *3*—differential thermocouple, *4*—regulator, *5*—standard voltage source, *6*—heater.

Heaviside step function, which cuts off from f the positive feedback branch corresponding to positive values of the argument of f; and β is a constant. Such a feedback model enables us to cover practically all cases of proportional temperature regulation.

In Ref. 5, where the problem defined by (1)–(3) was investigated for the case of quadratic nonlinearity of f, it was shown that all the parameters which determine the level of negative feedback can be grouped into a single dimensionless parameter A. There is a certain (critical) value $A \equiv A_c < 0$, which is such that for any $A > A_c$ the system is locally exponentially stable and exists in a stationary state:

$$\overline{T}(x) = Cx, \tag{4}$$

where C is the root of the equation

$$C = F(1 - Cx) \tag{5}$$

corresponding to the condition $A < 0$.

When $A < A_c$, the equilibrium state loses stability, and the stable periodic solution

$$T(x,t) = C(\varepsilon)x + \frac{2}{f''}\left\{ \sqrt{\frac{\varepsilon}{b_2}}[V_1(x)\exp(i\omega(\varepsilon)t) + V_1^*(x)\exp(-i\omega(\varepsilon)t)] + \frac{\varepsilon}{b_2}\frac{2|V_1(x_0)|^2}{1 - A_c}x + \frac{\varepsilon}{b_2}[V_2(x)\exp(i2\omega(\varepsilon)t) + V_2^*(x) \times \exp(-i2\omega(\varepsilon)t)]\right\}; \tag{6}$$

$$\omega(\varepsilon) = \frac{\omega_c}{1 + (c_2/b_2)\varepsilon}$$

appears in its vicinity. Here $f'' = [\partial^2 f/\partial T^2]|_{T = Cx_0}$; $\varepsilon = (A - A_c)/A_c$ is the supercriticality of the system; $V_1(x) = [\sinh\sqrt{\omega_c/2a}(1+i)x]$ is the spatial part of the solution of the linearized boundary-value problem (1)–(3); $\omega(\varepsilon)$ and ω_c are the frequencies of the auto-oscillations of the nonlinear and linear systems, respectively; and c_2 and b_2 are Lyapunov coefficients, which can be calculated by the method described in Ref. 5.

The results in Ref. 5 can easily be generalized to the case of nonlinearity of any type, provided that f is a globally concave smooth function which satisfies the condition $f'(0) = 0$. Then f can be expanded in a Taylor series in powers of the oscillating component of the temperature,

$$f[\beta - \overline{T}(x_0) - \widetilde{T}(x_0,t)] = C(\varepsilon) + \frac{A}{x_0}\widetilde{T}(x_0,t) + f''\frac{\widetilde{T}^2(x_0,t)}{2!} + f'''\frac{\widetilde{T}^3(x_0,t)}{3!}\ldots +, \tag{7}$$

and a nonlinear analysis can subsequently be performed using the algorithm presented in Ref. 5. If, in addition to the quadratic nonlinearity, the cubic terms in the expansion of f are taken into account in condition (7), the general form of the solution of the problem defined by (1)–(3) remains unchanged, but the dynamics of the system can undergo significant changes. The derivative f''' can be positive or negative; therefore, the coefficient b_2 can take either sign. If $b_2 > 0$, the soft cycle-creation bifurcation (6) can take place in system (1)–(3). When $b_2 < 0$, the onset of the auto-oscillations is ''hard,'' and the asymptotic form (6) of the periodic solution has meaning only in the subcritical region, where it corresponds to an unstable limit cycle.

The expressions presented above for the temperature field enable us to calculate the entropy of the system S and the entropy production P and to trace their evolution when auto-oscillations are excited. Since the temperature field and, consequently, S and P are determined by the supercriticality of the control parameter A, the dependence of S and P on ε is henceforth investigated as the latter parameter is varied from -1 to 0 for hard bifurcation and from $\varepsilon = -1$ to $-0 < \varepsilon \ll 1$ for soft cycle-creation bifurcation.

3. As was noted in constructing the mathematical model (1)–(3), the only distributed (accumulating) element in the system shown in Fig. 1 is the dielectric medium *1*. Let all the energy supplied to the system in the form of Joule heat be released in an infinitely thin layer on the surface at $x = \delta$, and let the lower surface have the temperature of the thermostat T_0. Then the entropy of the system S in excess of the entropy of the thermostat S_0 will be confined and produced only within the distributed element *1* and on its surface. We introduce the notation $s(x,t) = dS/dV$ for the density of the deviation of the entropy of the medium from S_0 for a certain temperature distribution $T(x,t) + T_0$. Being an additive function of state, the entropy density can be calculated as the result of a reversible process performed over an elementary volume dV, which in the present case is an isothermal layer of thickness dx and area V (we recall that the height of element *1* is equal to unity)

$$s(x,t)=\int_{T_0}^{T_0+T(x,t)} c_v \frac{dT}{T}=c_v\ \ln[1+\Theta(x,t)]. \tag{8}$$

Here c_v is the volumetric specific heat, and $\Theta(x,t)=T(x,t)/T_0$ is the normalized temperature. When Eq. (8) is taken into account, the entropy balance equation transforms into the diffusion equation

$$\dot{s}(x,t)=as''(x,t)+\frac{a}{c_v}[s'(x,t)]^2, \tag{9}$$

where the nonlinear term is the source density or the local entropy production.

Relation (8) enables us to find the exact value of the total entropy of the stationary state,

$$\overline{S}=V\int_0^1 s(x)dx=C_p\left\{\left[1+\frac{1}{\overline{\Theta}(1)}\right]\ln[1+\overline{\Theta}(1)]-1\right\}, \tag{10}$$

and to evaluate its variation in response to the excitation of auto-oscillations. For this purpose, we substitute $\Theta(x,t)=\overline{\Theta}(x)+\widetilde{\Theta}(x,t)$ into (8) and represent the resulting expression in the form

$$s(x,t)=c_v\ \ln[1+\overline{\Theta}(x)]+c_v\ \ln\left[1+\frac{\widetilde{\Theta}(x,t)}{1+\overline{\Theta}(x)}\right]. \tag{11}$$

The entropy density increment corresponding to the excitation of auto-oscillations

$$\Delta s(x,t)=c_v\left[\frac{\widetilde{\Theta}(x,t)}{1+\overline{\Theta}(x)}-\frac{1}{2}\frac{\widetilde{\Theta}^2(x,t)}{[1+\overline{\Theta}(x)]^2}+\cdots\right] \tag{12}$$

will be characterized using the average value of the entropy density $\langle\Delta s(x,t)\rangle_{\Pi(\varepsilon)}$ over the period $\Pi(\varepsilon)$. Utilizing solution (6) and calculating the average values of $\widetilde{\Theta}$ and $\widetilde{\Theta}^2$ to within terms of order ε

$$\langle\widetilde{\Theta}(x,t)\rangle_{\Pi(\varepsilon)}=\frac{2}{f''T_0}\frac{\varepsilon}{b_2}\frac{2|V_1(x_0)|^2}{1-A_c}x,$$

$$\langle\widetilde{\Theta}^2(x,t)\rangle_{\Pi(\varepsilon)}=\frac{4}{(T_0f'')^2}\frac{\varepsilon}{b_2}2|V_1(x)|^2 \tag{13}$$

and then substituting them into (12), we obtain

$$\langle\Delta s\rangle_{\Pi(\varepsilon)}=\frac{4\varepsilon}{b_2f''}\frac{c_v}{[1+\overline{\Theta}(x)]T_0}\times\left[\frac{|V_1(x_0)|^2}{1+|A_c|}x-\frac{1}{f''}\frac{|V_1(x)|^2}{[1+\overline{\Theta}(x)]T_0}\right]. \tag{14}$$

It follows from (14) that the entropy density increment corresponding to the excitation of auto-oscillations can be either positive or negative. The sign of the increment of the total entropy calculated as the volume integral of (14) depends significantly only on the coordinate of the temperature sensor x_0 and the quadratic nonlinearity of the system, which is characterized by f''. Clearly, in a weakly nonlinear system ($f''\ll 1$) the excitation of auto-oscillations will always be accompanied by lowering of the total entropy of the system.

In an open thermodynamic system the entropy balance is determined by the fluxes of entropy at the boundaries and its production within the system. We shall show that the lowering of the entropy as $f''\to 0$ is not associated with a decrease in its production and that, conversely, it occurs against a background of an increase in energy production. Using (9), we express the total entropy production by heat flow in terms of the rate of variation of the total entropy and the difference between its fluxes at the boundaries. After adding the entropy production by a heat source to the expression obtained, we find the total entropy production in the system

$$P=I_s(0)+\frac{\partial\widetilde{S}}{\partial t}, \tag{15}$$

where $I_s(0)=V\lambda\Theta'(x,t)|_{x=0}$ is the flux of entropy into the thermostat and $\widetilde{S}$ is the oscillating component of the total entropy.

The period-averaged value of the total entropy production

$$\langle P\rangle_{\Pi(\varepsilon)}=V\lambda[\overline{\Theta'}+\langle\widetilde{\Theta}'(0)\rangle_{\Pi(\varepsilon)}]=\frac{V\lambda}{T_0}\left[C(\varepsilon)+\frac{2}{f''}\frac{\varepsilon}{b_2}\frac{2|V_1(x_0)|^2}{1+|A_c|}\right], \tag{16}$$

which depends only on the stationary component of the heat flux, increases upon the excitation of auto-oscillations by an amount $\langle\widetilde{\Theta}'(0)\rangle_{\Pi(\varepsilon)}$, which is evaluated at the steady-state temperature in solution (7). Thus, when f'' decreases and the other parameters remain unchanged, the total entropy production in the system increases.

It follows from a comparison of (14) and (16) that the entropy decreases as the nonlinearity of the system increases. In other words, the ordering of an open system increases as the entropy flux passing through it increases. Obviously, the degree of order can be evaluated correctly from the change in entropy only if its production is constant. In this case it is reasonable to normalize the average value of the total entropy of the system $\langle S\rangle_{\Pi(\varepsilon)}$ to the average value of its production, and to take the relative increment of the normalized mean value

$$\mathrm{So}=\frac{\Delta(\langle S\rangle_{\Pi(\varepsilon)}/\langle P\rangle_{\Pi(\varepsilon)})}{\langle S\rangle_{\Pi(\varepsilon)}/\langle P\rangle_{\Pi(\varepsilon)}}\approx\frac{\Delta\langle S\rangle_{\Pi(\varepsilon)}}{\langle S\rangle_{\Pi(\varepsilon)}}-\frac{\Delta\langle P\rangle_{\Pi(\varepsilon)}}{\langle P\rangle_{\Pi(\varepsilon)}}, \tag{17}$$

where the approximate equality holds for small changes in S and P, as the measure of the ordering of motion.

In particular, for the degree of order of an auto-oscillatory state relative to a stationary state we have

$$\mathrm{So}\approx\frac{\langle S\rangle_{\Pi(\varepsilon)}}{\overline{S_c}}-\frac{\langle P\rangle_{\Pi(\varepsilon)}}{\overline{P_c}}, \tag{18}$$

where $\overline{S_c}$ and $\overline{P_c}$ are the critical values of the entropy and its production in the stationary state. It is clear that So will simultaneously be a criterion of the ordering of motion. For example, if we take two auto-oscillatory states with the same entropy value, the state with greater entropy production, i.e., with a smaller value of So, is more ordered.

4. In accordance with the well-known Prigogine principle,[6] the real stationary state of an open thermodynamic system is characterized by the minimum local entropy production with respect to the possible stationary states. The analytical expression of this principle for a system with N degrees of freedom, in which some of the thermodynamic forces X_j $j=0,1,\ldots,k$ are fixed, has the form

$$\frac{\partial p}{\partial X_i}=0, \tag{19}$$

where X_i denotes the thermodynamic forces that are not fixed.

In a stationary state the system of equations (1)–(3) has only one degree of freedom, i.e., the state is completely defined, if one thermodynamic variable $T'=C(\varepsilon)$ is assigned. According to Ref. 7, two cases are possible for a system with one degree of freedom: 1) $k=1$, where the state of the system is completely defined by an external force; 2) $k=0$, where the system is closed, and its entropy is maximal. They can be classified after de Groot as first- and zero-order stationary states, respectively. A third case, which does not fit this classification and which is considered in the present work, appears in systems with external feedback [for example, like (5)]. In the latter case the actual state will clearly be indistinguishable in any way by the sign of the entropy or its production from the other stationary states, since it is uniquely specified by Eq. (5). Nevertheless, after the system deviates from the stationary state under the action of some external force, it will tend to return to it, i.e., will be stable in the sense of Le Chatelier's principle.

To illustrate this statement, we bring the nonlocal boundary condition (5) into the form

$$I(t)-\varphi[I_0-I(t)]=0, \tag{20}$$

where $I=\lambda VT'$ and $I_0=\beta\lambda V/x_0$.

We assume that the action of an external thermodynamic force $\Delta X^{\rm ext}$ in the system results in the creation of an additional heat flux $\Delta I^{\rm ext}$. The resultant change in the flux ΔI is determined from the coupling condition

$$I+\Delta I=\varphi(I_0-I-\Delta I)+\Delta I^{\rm ext}. \tag{21}$$

Taking into account only the term that is linear with respect to the flux in the expansion of the function of the control signal, we obtain the following relation:

$$\Delta I=\frac{1}{1-A}\Delta I^{\rm ext}, \tag{22}$$

which is an analytical expression of Le Chatelier's principle. Relation (22) can easily be transformed into the expression

$$\frac{dP}{dX^{\rm ext}}=\frac{1}{1-A}\frac{dP^{\rm ext}}{dX^{\rm ext}}, \tag{23}$$

according to which the rate of variation of the entropy production in response to variation of the external thermodynamic force in a system with feedback is $1-A$ times smaller than in a system without feedback and tends to zero as $A\to-\infty$.

5. As we know, there is an analogy between equilibrium phase transitions and self-organization processes in open thermodynamic systems, which have been termed nonequilibrium phase transitions.[8] They include auto-oscillations in Thomson systems, since an influx of energy from without is needed to sustain them, and, therefore, an auto-oscillatory system is a nonequilibrium system.[3] We shall show that the excitation of auto-oscillations in a distributed relaxational system also completely fits the scheme of a nonequilibrium phase transition.

Let us draw some analogies between the phase transitions in condensed media and in the system under consideration. The hydrodynamic mode of a condensed system in a given state corresponds to the unstable mode

$$T(x,t)=\xi V(x)\exp[i\omega(\varepsilon)t], \tag{24}$$

whose amplitude is determined by a complex parameter of order ξ. The choice of the control parameter is also quite obvious: the reduced temperature in Landau's theory $\tau=(T-T_c)/T_c$ parallels the supercriticality $\varepsilon=(A-A_c)/A_c$ of the system.

We note that in the general case the Landau concept of phase transitions does not provide an adequate description of the picture of a nonequilibrium phase transition. As follows from the brief outline of the current theory of nonequilibrium phase transitions presented in the review by Olemskoĭand Koplyk,[9] it differs from the original Landau concept in that it employs at least three additional degrees of freedom, which correspond to the control parameter, the conjugate field, and the order parameter. A transition *per se*, which can be regarded as the spatiotemporal evolution of a hydrodynamic mode, whose amplitude is specified by the order parameter, is a result of the competition between two types of feedback, viz., the positive feedback on the order parameter from the control parameter and the negative feedback on the order parameter from the field that is conjugate to it. The latter implements Le Chatelier's principle. A quasistatic Landau phase transition takes place when the order parameter is proportional to the field, and the control parameter does not depend on the order parameter.[9] In the present case it is assumed for reasons that are discussed below that the control parameter and, therefore, the supercriticality ε do not depend on the order parameter.

Let $\varepsilon>0$, and let the temperature oscillations $\tilde{T}(x,t)$ be established in the system. We write the periodic solution of system of equations (1)–(3) in the general form

$$\begin{aligned}\tilde{T}(x,t)=&\sum_{n=1}^{\infty}|\xi|^{2n}\overline{T_{2n}}(x)+\frac{2}{f''}[\xi V_1(x)\exp(i\omega t)\\&+\xi^* V_1^*(x)\exp(-i\omega t)]+[\xi^2 V_2(x)\exp(i2\omega t)\\&+\xi^{*2}V_2^*(x)\exp(-i2\omega t)]\\&+|\xi|^2[\xi W_1(x)\exp(i\omega t)+\xi^* W_1^*(x)\\&\times\exp(-i\omega t)]+|\xi|^2[\xi V_3(x)\exp(i3\omega t)\\&+\xi^* V_3^*(x)\exp(-i3\omega t)]+\ldots,\end{aligned} \tag{25}$$

where the sum on the right-hand side of (25) takes into account the nonlinearity-induced additions to the stationary temperature, which can be determined in each even step of the algorithm described in Ref. 5, and $V_1(x)=[\sinh\sqrt{\omega_c/2}(1+i)x]$ is the spatial part of the solution of the linearized boundary-value problem (1)–(3).

Using expansion (7), we substitute (25) into condition (3) and average the latter over an oscillation period

$$\left(|\xi|^2\overline{T_2'}(x)+\sum_{n=2}^{\infty}|\xi|^{2n}\overline{T_{2n}'}(x)\right)_{x=1}=\frac{A_c}{x_0}|\xi|^2\overline{T_2}(x_0)$$
$$+\frac{A_c}{x_0}\sum_{n=2}^{\infty}|\xi|^{2n}\overline{T_{2n}}(x_0)+\frac{A_c}{x_0}\varepsilon\Delta\overline{T}(x_0)$$
$$+\frac{f''}{2}\left\{\Delta\overline{T^2}(x_0)+|\xi|^2\frac{8}{f''^2}|V_1(x_0)|^2\right.$$
$$+|\xi|^4\frac{4}{f''}[W_1(x_0)V_1^*(x_0)+W_1^*(x_0)V_1(x_0)]$$
$$\left.+2|\xi|^4|V_2(x_0)|^2+\cdots\right\}+\frac{f'''}{3\cdot2}\left\{\Delta\overline{T^3}(x_0)\right.$$
$$+\Delta\overline{T}\frac{24}{f''^2}|\xi|^2|V_1(x_0)|^2+|\xi|^4\frac{12}{f''^2}[V_2(x_0)V_1^{*2}(x_0)$$
$$\left.+V_1^2(x_0)V_2^*(x_0)]+\cdots\right\}. \tag{26}$$

In condition (26) we equate the terms accompanying $|\xi|^2$. With consideration of the equality $\overline{T_2}(x_0)/x_0=\overline{T_2'}(x_0)$ this gives

$$\overline{T_2'}=\frac{2}{f''}\frac{2|V_1(x_0)|^2}{1+|A_c|}. \tag{27}$$

This solution coincides with the solution obtained in Ref. 5, as can be seen by comparing (27) with the addition to the stationary temperature in (6).

To go from a dynamical description to a thermodynamic method we turn to the algorithm for constructing the asymptote of a periodic solution in the Andronov–Hopf theorem. In a certain step of this algorithm the nonlinear problem defined by (1)–(3) is transformed into a recurrent sequence of linear inhomogeneous boundary-value problems with nonlocal feedback conditions. These conditions must be satisfied by the amplitude and phase of the solution of each of the problems in the sequence. The oscillation amplitude is found from the condition of solvability, which follows from the condition for the complex amplitude of the third problem in the sequence.

If we henceforth follow the dynamical method, in accordance with the formalism of the Andronov–Hopf theorem we must expand the supercriticality ε of the system in a series in powers of ξ, and after equating the remaining terms of the same order with respect to the small parameter in (26), we must express $T_{2n}(x)$ in terms of the higher moments of the temperature oscillations at x_0. At this point, of course, ξ remains unknown, since it is determined in the odd steps of the algorithm. However, as will be shown below, in the vicinity of the bifurcation point the system is not completely deterministic, and the oscillation amplitude should be sought on the basis of other considerations.

We assume that $0<\varepsilon\ll1$ and that the system is so close to the critical point that the oscillation amplitude of the second harmonic at x_0 is comparable to the temperature fluctuations. Then the condition for the complex amplitude of the third problem in the sequence becomes meaningless. At the same time, it is no longer necessary to take into account the feedback on the order parameter from the control parameter, since this feedback is realized through fundamental-frequency oscillations which enter into the nonlinearity and inhomogeneity of the third problem in the sequence. In this case the oscillation amplitude should be regarded as an order parameter, i.e., as an additional degree of freedom, relative to which the system can undergo virtual displacements. The actual state is then determined by the minimum of a functional of a certain thermodynamic potential, whose nature follows from condition (26). To make this clearer, we subject this condition to several transformations.

We perform the following substitutions in (26):

$$\sum_{n=2}^{\infty}|\xi|^{2n}\frac{\overline{T_{2n}}(x_0)}{x_0}\to\sum_{n=2}^{\infty}|\xi|^{2n}\overline{T_{2n}'}(x_0)\to\Delta\overline{T'}(x_0)$$

and after utilizing the equality $\Delta\overline{T'}(x_0)=\Delta\overline{T'}(1)$, we move $-|A_c|\Delta\overline{T'}(1)$ over to the left-hand side of the condition. Taking into account that $\Delta\overline{T'}(1)=\Delta\overline{T'}(0)$ and $\Delta\overline{T'}(0)=(T_0/V\lambda)\Delta\overline{I_s}(0)=(T_0/V\lambda)\langle\Delta P\rangle_{\Pi(\varepsilon)}$, we obtain

$$\langle\Delta P\rangle_{\Pi(\varepsilon)}=\frac{V\lambda}{T_0(1+|A_c|)}\left(\frac{A_c}{x_0}\varepsilon\sum_{n=2}^{\infty}|\xi|^{2n}\overline{T_{2n}}(x_0)\right.$$
$$+\frac{f''}{2}\left\{\left[\sum_{n=2}^{\infty}|\xi|^{2n}\overline{T_{2n}}(x_0)\right]^2\right.$$
$$+|\varepsilon|^4\frac{4}{f''}[W_1(x_0)V_1^*(x_0)+W_1^*(x_0)V_1(x_0)]$$
$$\left.+2|\xi|^4|V_2(x_0)|^2+\cdots\right\}$$
$$+\frac{f'''}{3\cdot2}\left\{|\xi|^2\frac{24}{f''^2}|V_1(x_0)|^2\sum_{n=2}^{\infty}|\xi|^{2n}\overline{T_{2n}}(x_0)\right.$$
$$+|\xi|^4\frac{12}{f''2}[V_2(x_0)V_1^{*2}(x_0)$$
$$\left.\left.+V_1^2(x_0)V_2^*(x_0)]+\cdots\right\}\right). \tag{28}$$

Thus, the potential determining the actual state of the system is the entropy production or, more precisely, the addition to the entropy production caused by the higher moments of the temperature oscillations. It is noteworthy that the principle of minimum entropy production is satisfied si-

multaneously with Le Chatelier's principle. The actual entropy production increment is $1+|A_c|$ times smaller than the entropy production caused by the perturbing heat flux [the right-hand side of (28)].

Grouping the terms with identical powers of ξ on the right-hand side of (28) and taking into account that $\Delta\overline{T_2}(x_0)/x_0 = f''|V_1(x_0)|^2/(1+|A_c|)$, we arrive at the following relation between the entropy production increment and the order parameter

$$\langle \Delta P\rangle_{\Pi(\varepsilon)} = -a\varepsilon|\xi|^2 + B|\xi|^4, \tag{29}$$

where

$$a = \frac{V\lambda}{T_0}\frac{|A_c|}{(1+|A_c|)^2}\frac{2}{f''}2|V_1(x_0)|^2,$$

$$B = \frac{V\lambda}{T_0(1+|A_c|)}\left(\frac{f''}{2}\overline{T_2^2}(x_0) + f''|V_2(x_0)|^2 + 2[W_1(x_0)V_1^*(x_0) + W_1^*(x_0)V_1(x_0)] + \frac{4f'''}{f''^3}\frac{4|V_1(x_0)|^4}{1+|A_c|} + \frac{f'''}{f''^2}2[V_2(x_0)V_1^{*2}(x_0) + V_1^2(x_0)V_2^*(x_0)]\right).$$

We note that the value of the coefficient B is not defined here, and only the question of the correctness of estimating it using the solutions obtained within the dynamical approach can be posed. The value of the order parameter is determined by the requirement that the entropy production increment be extremal:

$$|\xi|^2 = \frac{a\varepsilon}{2B}. \tag{31}$$

It is clear that among the periodic solutions which satisfy the extremum condition (31), the solution which corresponds to the entropy production minimum will be orbitally stable.

A transition can be described in general terms as follows. The oscillations of the fundamental frequency which satisfy the feedback condition generate oscillations at higher harmonics, whose second and higher moments, as well as the higher moments at the fundamental frequency, create additional fluxes that produce additional entropy. The amplitude of the oscillations is established such that the entropy production $\langle \Delta P\rangle_{\Pi(\varepsilon)}$ is minimal. Of course, the picture changes as soon as the supercriticality reaches a value at which the phase-matching condition is satisfied for the third harmonic. Despite the restriction of this scheme by the requirement of a small supercriticality, it leads to the same results as the dynamical method at the qualitative level of the description.

6. The relaxation of unconserved order parameters, which include the temperature, to the equilibrium value is described by the Ginzburg–Landau evolution equation

$$\frac{\partial\xi}{\partial\tau} = -\frac{\delta\Phi}{\delta\xi}, \tag{32}$$

where τ is the ''slow'' time and Φ is a functional of the thermodynamic potential which has a minimum in the equilibrium state.[8,9]

Since in the case under consideration the total entropy production (29) is such a functional, the right-hand side of Eq. (32) takes the form

$$\frac{\partial}{\partial\xi^*}\langle \Delta P\rangle_{\Pi(\varepsilon)} = -a\varepsilon\xi + 2B|\xi|^2\xi, \tag{33}$$

where the complex amplitude ξ is a function of the ''slow'' time.

Substituting (33) into (32) and performing the normalization to $\xi = \psi/\sqrt{2B}$, we arrive at the following form of the Ginzburg–Landau equation

$$\frac{\partial\psi}{\partial\tau} = a\varepsilon\psi - |\psi|^2\psi. \tag{34}$$

Its solution

$$|\psi(\tau)|^2 = \frac{a\varepsilon}{1+(a\varepsilon/|\psi_0|^2 - 1)\exp(-2a\varepsilon\tau)}, \tag{35}$$

where $\psi_0 = \psi(0)$, describes the transient processes in the auto-oscillatory system under consideration.

Expressions (31) and (35) enable us to describe soft and hard cycle bifurcation on the qualitative level. In fact, the coefficient B can be positive or negative, depending on the relationship between f'' and f'''. Plots of the dependence of the entropy production on the order parameter are shown in Fig. 2. If $B>0$, then, according to (31), a stable limit cycle exists in the transcritical region, i.e., for $\varepsilon>0$. The process of establishing a limit cycle is described by Eq. (35), in which we should set $\psi_0 = \delta$, where δ is the magnitude of the fluctuations of the order parameter. In this case the limit cycle is clearly an attractor, i.e., phase trajectories from different initial states tend to the equilibrium value of the order parameter $|\overline{\psi|^2} = a\varepsilon$ (Fig. 3a).

If $B<0$, Eq. (31) holds only for $\varepsilon<0$, and the corresponding limit cycle is unstable. As $\varepsilon\to 0$, the unstable cycle shrinks from a region of negative values to an equilibrium point, which then loses its stability. In this case, when the system passes through the bifurcation point, the onset of the auto-oscillations is hard, and the amplitude is infinite within the model under consideration. To characterize qualitatively the time dependence of the order parameter for $B<0$ and $\varepsilon<0$, only the sign of the exponent in (35) must be changed. Let $B<0$ and $\varepsilon<0$, and let the system be displaced from the equilibrium state as a result of a fluctuation of the order parameter. After setting $|\psi_0|^2 = a|\varepsilon| \pm \delta$, where $\delta \ll a|\varepsilon|$, and performing some obvious transformations, we obtain

$$|\psi(\tau)|^2 \approx \frac{a|\varepsilon|}{q \mp (\delta/a|\varepsilon|\exp(2a|\varepsilon|\tau))}. \tag{36}$$

As expected, this cycle is unstable (Fig. 3b), and the auto-oscillations either damp or their amplitude increases to infinity in response to even an infinitesimal deviation from the equilibrium value of the order parameter.

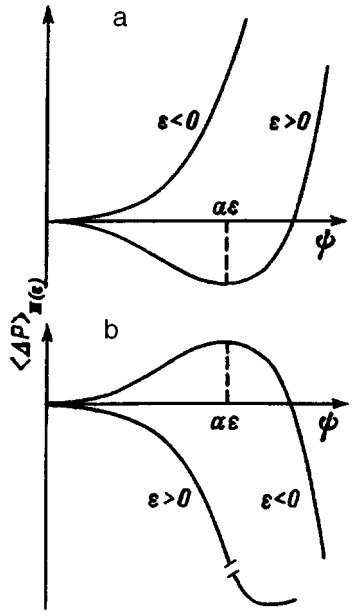

FIG. 2. Dependence of the entropy production increment on the order parameter ψ. When $B>0$, the equilibrium value of the order parameter $|\overline{\psi}|^2=a\varepsilon$ is determined by the minimum on the $\langle \Delta P(\psi)\rangle_{\Pi(\varepsilon)}$ curve, and the absolute value of the order parameter is a smooth function of the supercriticality ε of the system. When $B<0$ and $\varepsilon<0$ (b), the stationary state ($\psi=0$) is metastable, i.e., when the perturbation is sufficiently strong, the system can pass into a more stable oscillatory state. If $B<0$, the order parameter changes abruptly upon passage through the bifurcation point, which is interpreted as a first-order nonequilibrium phase transition.

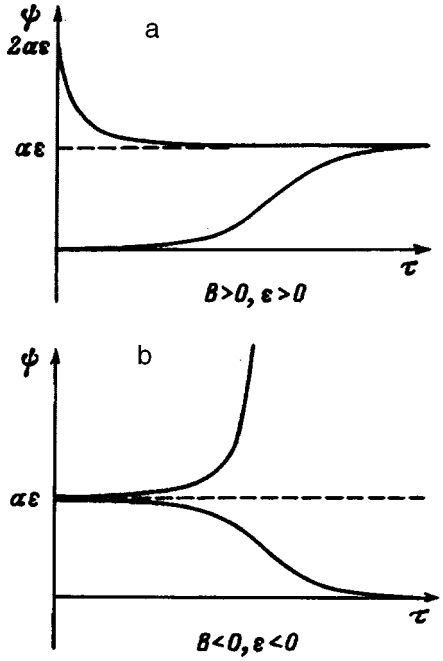

FIG. 3. Approximate form of the time dependence of the order parameter for various initial conditions: a—stable cycle, b—unstable cycle.

In a real system the oscillation amplitude is always restricted by a strong nonlinearity, which corresponds to the global minimum of the entropy production in Fig. 2b corresponds. For example, in the system under consideration it is the supply voltage, whose restricted character is not reflected in the model described by (1)–(3). Therefore, after passage through the bifurcation point, the system finds itself in a new stable state. However, whereas under soft auto-oscillation excitation conditions the order parameter is a smooth function of ε, under hard conditions it changes abruptly. Thus, hard and soft bifurcation can be interpreted as first- and second-order nonequilibrium phase transitions.

7. The results of an investigation of the evolution of the entropy of a relaxational system with nonlinear external feedback in response to variation of a control parameter in the vicinity of a bifurcation point allow us to draw the following conclusions. The entropy increment on excitation of auto-oscillations will be smaller, the larger is the entropy flux caused by the external source. The relative increment of the Clausius entropy normalized to its total production can serve as a measure of the ordering of motion in such a system. In the subcritical region the system considered is stable in the Le Chatelier sense, i.e., when an infinitesimal deviation from the equilibrium state arises under the influence of an external force, fluxes which weaken the influence of that force arise in the system. In the transcritical region Le Chatelier's principle is supplemented by the principle of minimum entropy production, according to which the amplitude of the temperature oscillations is established such that the addition to the entropy production created by the higher moments of the temperature oscillations is minimal. This principle, which was formulated under the assumption that the control parameter does not depend on the order parameter, allows us to treat the ordering of motion as a quasistatic phase transition. In particular, hard and soft bifurcations can be interpreted within this hypothesis as first- and second-order nonequilibrium phase transitions, and the Ginzburg–Landau evolution equation can be used to describe the transient processes.

[1] V. S. Anishchenko, *Complex Oscillations in Simple Systems* [in Russian], Nauka, Moscow (1990).
[2] Yu. L. Klimontovich, *Turbulent Motion and the Structure of Chaos*, Kluwer, Dordrecht (1991).
[3] Yu. L. Klimontovich, *Statistical Physics*, Harwood, Chur, Switzerland –New York (1986).
[4] Yu. L. Klimontovich, Pis'ma Zh. Tekh. Fiz. **9**, 1412 (1983) [Sov. Tech. Phys. Lett. **9**, 606 (1983)].
[5] A. S. Rudyi, Int. J. Thermophys. **14**, 159 (1993).
[6] G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations* [Wiley, Chichester (1977); Mir, Moscow (1979)].
[7] I. P. Bazarov, É. V. Gevorkyan, and P. N. Nikolaev, *Nonequilibrium Thermodynamics and Physical Kinetics* [in Russian], Izd. Mosk. Univ., Moscow (1989).
[8] H. Haken, *Synergetics: An Introduction: Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry, and Biology*, 2nd enl. ed., Springer-Verlag, Berlin–New York (1978); Mir, Moscow (1980).
[9] A. I. Olemskoĭ and I. V. Koplyk, Usp. Fiz. Nauk **165**, 1106 (1995) [Phys. Usp. **38**, 1061 (1995)].

Translated by P. Shelnitz

Propagation of a normal–superconductor boundary along a high-T_c superconducting film heated by microwave radiation

N. A. Buznikov

Scientific Research Center for Problems in Applied Electrodynamics, Russian Academy of Sciences, 127412 Moscow, Russia

A. A. Pukhov

Institute of High Temperatures, Russian Academy of Sciences, 127412 Moscow, Russia
(Submitted June 14, 1996)
Zh. Tekh. Fiz. **68**, 111–116 (January 1998)

The dynamics of the superconducting–normal (S–N) transition in a thin high-temperature superconducting film heated by microwave radiation is investigated theoretically. The dependence of the rate of propagation of the normal–superconductor (NS) interphase boundary on the intensity of the radiation is obtained by solving the two-dimensional nonstationary heat conduction equation. It is shown that in calculating this dependence it is important to take into account two-dimensional effects connected with nonlinearity of heating over the substrate thickness, the reverse side of which is stabilized with respect to temperature. The results obtained may be important in investigating S–N transitions in superconducting devices used in the microwave range. © *1998 American Institute of Physics.* [S1063-7842(98)02101-1]

INTRODUCTION

Incident electromagnetic radiation can cause a film of high-temperature superconducting material on a dielectric substrate to go from its superconducting to its normal state (an S–N transition). This phenomenon has excited considerable interest in recent times, since such films are used in various superconducting devices that operate in the microwave and infrared regions.[1–3] The abrupt change in electrodynamic characteristics of a film undergoing this switching process has suggested potential applications for these films as antennas and resonators,[4] screens and filters,[1,2] switches and power limiters,[5,6] etc. The thermal mechanism for the S–N transition discussed in detail in Refs. 7 and 8 leads to a qualitatively correct description of the nonlinear effects connected with S–N switching of resonators[9,10] observed experimentally.[11] The abrupt increase in resistance of films when the critical temperature T_c is exceeded leads to the appearance of thermal bistability. For certain values of the incident microwave power P the film can be in two stable uniform states: a superconducting state (with temperature below critical) and a normal state (with temperature above critical).[7–10]

This S–N transition occurs uniformly over the entire length of the high-T_c film only when the length is relatively small. When this is not true, the thermal disruption of the superconducting state will, as a rule, be local in character, in which case the S–N transition takes place by propagation of a temperature switching autowave along the film. This autowave consists of a moving NS interphase boundary that transforms the sample ahead of it from the superconducting to the normal state. The asymptotic behavior of the NS boundary is characterized by a constant velocity of propagation v. The nucleation, propagation, and stability of these autowaves, and also the dependence of v on the film parameters, were discussed in considerable detail in Refs. 12 and 13 for low-temperature composite superconductors carrying a transport current.

The propagation of an NS boundary in thin high-T_c films heated by microwave radiation or by a transport current has been investigated both experimentally[14–17] and theoretically.[10–18] The S–N transition in a film–substrate system exhibits a large number of peculiarities that cannot be described within the framework of the standard one-dimensional theory of propagation of temperature autowaves.[12,13] These features are connected with the presence of nonuniform heating over the cross section (heat is released only in the thin high-T_c film[18]) and with the nonlinear distribution of temperature over the thickness of the substrate, which has a high thermal conductivity and plays the role of a ''heat bath'' for the high-T_c film. In Refs. 7–10, the authors analyzed the S–N transition in a high-T_c film under the action of microwave radiation by assuming that the temperature varies linearly over the substrate thickness (from the film temperature T to the temperature T_0 of the heat sink that stabilizes the back side temperature of the substrate). Using this assumption, they were able to solve the problem of propagation of an NS boundary in the one-dimensional approximation; however, their solution is valid only when the S–N transition takes place uniformly over the length of the film. A correct description of propagation of an NS boundary in this system will, generally speaking, require the solution of a two-dimensional nonstationary heat conduction equation.[18]

ONE-DIMENSIONAL APPROXIMATION

The heating of a high-T_c film with thickness D_f placed on a dielectric substrate with thickness D_s (Fig. 1) by microwave radiation is described in the one-dimensional approximation by the heat conduction equation:[7–9]

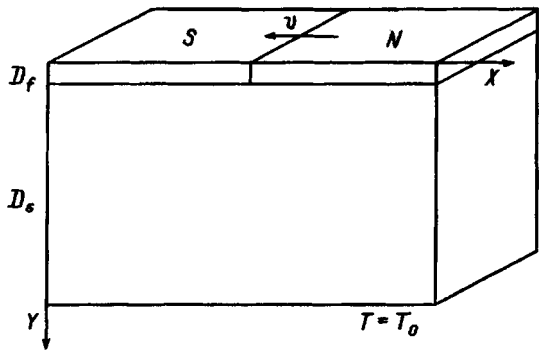

FIG. 1. A sketch of the propagation of an NS interphase boundary along a superconducting film on a substrate. The reverse side of the substrate is stabilized at a temperature T_0.

$$\frac{C_s D_s}{2}\frac{\partial T}{\partial t}=k_s D_s\frac{\partial^2 T}{\partial X^2}-\frac{k_s}{D_s}(T-T_0)=Q(T)D_f, \tag{1}$$

where T is the film temperature, C_s and k_s are the heat capacity and thermal conductivity of the substrate, whose back side is stabilized at a temperature T_0, $Q(T)=\varkappa(T)\cdot P$ is the specific heating power supplied to the film, P is the intensity of the incident radiation, and $\varkappa(T)$ is the microwave absorption coefficient of the film.

Equation (1) incorporates the fact that for typical ratios of the parameters $k_f D_s \gg k_s D_f$, $C_f D_f \ll C_s D_s$, $k_f \ll k_s$ (where C_f and k_f are the heat capacity and thermal conductivity of the film) the film temperature is uniform over its cross section, while the effective heat capacity and thermal conductivity of the film–substrate system are determined only by the properties of the substrate. The justification for this assertion comes from the fact that this is in fact the temperature distribution established in the substrate for the case of an S–N transition in a high-T_c film that is uniform with respect to length ($\partial T/\partial X=0$).

The temperature dependence of $\varkappa(T)$ is connected with the abrupt change in the electrodynamic properties of the high-T_c film during an S–N transition, and within the framework of the two-fluid model of a superconductor this quantity can be described by the relation[7]

$$\varkappa(T)=\varkappa_n\frac{(T_c-T_0)^2}{(T_c-T_0)^2+\beta(T_c-T)^2\,\eta(T_c-T)}. \tag{2}$$

Here $\varkappa_n=4r/(2r+1)^2$ is the absorption coefficient of the film in the normal state, $r=(\varepsilon_0/\mu_0)^{1/2}(\rho/D_f)$, ρ is the resistivity of the film in the normal state, $\beta=[(f/f_0)\times(2r+1)]^{-2}$, f is the frequency of the incident radiation, $f_0=2\rho(T_c-T_0)/\pi\mu_0\lambda^2 T_c$, λ is the London penetration depth at $T=T_0$, and $\eta(x)$ is the Heaviside step function. From Eq. (2) it follows that the temperature nonlinearity of $\varkappa(T)$ is strong when $f\ll f_0$, which leads to microwave bistability of the film. For typical values of the parameters of a Y–Ba–Cu–O film in the range of liquid nitrogen temperatures $T_0\cong 77$ K, $T_c\cong 90$ K, $\rho\cong 5\times 10^{-7}\,\Omega\cdot$m, $\lambda\cong 10^{-7}$m we find that $f_0\approx 3\times 10^{12}$ Hz. Thus, in the microwave region ($f\ll f_0$) the temperature dependence of $\varkappa(T)$ can be approximated by the step function

$$\varkappa(T)=\varkappa_n\,\eta(T-T_c). \tag{3}$$

The temperatures at which uniform steady states of a high-T_c film on a thermally stabilized substrate can exist are determined from the condition that the heat liberated in the film equals the removal of heat to the substrate. From Eq. (1) we obtain the equation for heat balance when $\partial T/\partial t=\partial^2 T/\partial X^2=0$:

$$\varkappa(T)\cdot P=\frac{k_s}{D_f D_s}(T-T_0), \tag{4}$$

whose solutions are the temperatures of the steady-state superconducting ($T_1=T_0$) and normal ($T_2=T_0+\varkappa_n P D_f D_s/k_s$) states. The NS boundary consists of a switching autowave between these two stable states of the high-T_c film.

Assuming for simplicity that the temperature dependences of the heat capacity and thermal conductivity of the substrate can be neglected, and introducing the dimensionless parameters

$$x=\frac{X}{D_s},\quad \tau=\frac{t}{C_s D_s^2/k_s},$$

$$\Theta=\frac{T-T_0}{T_c-T_0},\quad p=P\frac{\varkappa_n D_f D_s}{k_s(T_c-T_0)}, \tag{5}$$

let us write Eq. (1) in the form

$$\frac{1}{2}\frac{\partial\Theta}{\partial\tau}=\frac{\partial^2\Theta}{\partial x^2}-\Theta+p\,\eta(\Theta-1). \tag{6}$$

The propagating NS boundary is described by a self-similar solution to Eq. (6) of the form $\Theta(x,\tau)=\Theta(x+u\tau)$ which satisfies the boundary conditions $\Theta(-\infty)=0$ and $\Theta(+\infty)=p$. Here $u=v/v_h$, $v_h=k_s/C_s D_s$ is a characteristic ''thermal'' velocity of the NS boundary. Note that v_h is determined only by the properties of the substrate, and for characteristic values of the parameters of a MgO substrate ($C_s\cong 5\times 10^5$ J$\cdot$m$^{-3}\cdot$K^{-1}, $k_s\cong 350$ W$\cdot$m$^{-1}\cdot$K^{-1}) or a Al_2O_3 substrate ($C_s\cong 4\times 10^5$ J$\cdot$m$^{-3}\cdot$K^{-1}, $k_s\cong 650$ W$\cdot$m$^{-1}\cdot$K^{-1}) of thickness $D_s\cong 10^{-3}$ m we obtain $v_h\approx 1$ m/s. The dimensionless propagation velocity of the NS boundary u depends on the value of the control parameter p, which is the ratio of film heating in the normal state to the characteristic removal of heat to the substrate.

Equation (6) is a piecewise-linear equation and can be solved analytically,[12,13] which allows us to obtain the following expression for the propagation velocity of the NS boundary:

$$u=2\frac{p-2}{\sqrt{p-1}}. \tag{7}$$

Equation (7) determines the function $u(p)$ over the entire range of bistability values $1<p<\infty$. The normal phase emerges from the superconducting phase ($u>0$) when the intensity of the radiation exceeds a threshold value $p_p=2$ (the ''propagation intensity''). From Eq. (7) we find that $u\cong 2\times(p-2)$ for $p\cong p_p$ and $u\cong 2p^{1/2}$ for $p\gg p_p$. In the

range $1<p<p_p$ the normal phase is suppressed ($u<0$) and superconductivity is re-established in the film.

TWO-DIMENSIONAL APPROXIMATION

We mentioned above that the one-dimensional approximation is valid when the temperature variation over the thickness of the substrate is linear. However, this situation is encountered only far from the front of the NS boundary. A more correct description of propagation of the NS boundary requires the inclusion of the two-dimensionality of the problem. For this, we must consider the film and substrate separately as two interacting thermal subsystems. For $k_f D_s \gg k_s D_f$ the temperature distribution along the film satisfies the one-dimensional heat conduction equation

$$C_f \frac{\partial T}{\partial t} = k_f \frac{\partial^2 T}{\partial X^2} + \theta(T) + \frac{k_s}{D_f}\frac{\partial T}{\partial Y}, \quad 0<Y<D_f. \tag{8}$$

The last term in Eq. (8) corresponds to the removal of heat from the film to the substrate. The temperature distribution in the substrate is described by the two-dimensional heat conduction equation

$$C_f \frac{\partial T}{\partial t} = k_s \frac{\partial^2 T}{\partial X^2} + k_s \frac{\partial^2 T}{\partial Y^2}, \quad D_f<Y<D_f+D_s. \tag{9}$$

Taking into account that $D_f \ll D_s$ and $Q(T)=P\varkappa_n \eta(T-T_c)$, along with Eq. (5), let us write Eqs. (8) and (9) in dimensionless form:

$$C\frac{\partial \Theta}{\partial \tau} = K\frac{\partial^2 \Theta}{\partial x^2} + p\,\eta(\Theta-1) + \frac{\partial \Theta}{\partial y}, \quad y=0, \tag{10}$$

$$\frac{\partial \Theta}{\partial \tau} = \frac{\partial^2 \Theta}{\partial x^2} + \frac{\partial^2 \Theta}{\partial y^2}, \quad 0<y<1, \tag{11}$$

where $C=C_f D_f/C_s D_s$, $K=k_f D_f/k_s D_s$, $y=Y/D_s$.

It is necessary to supplement Eq. (11) by a boundary condition at $y=1$. Taking into account that from the side opposite the film the substrate is temperature stabilized, so that $T=T_0$, we have

$$\Theta=0, \quad y=1. \tag{12}$$

Thus, the dynamic temperature variation in the film–substrate system is described by the two-dimensional nonstationary heat conduction Eq. (11) with the boundary conditions (10) and (12). Simple estimates show that for Y–Ba–Cu–O films ($C_f \cong 9\times 10^5$ J·m^{-3}·K^{-1}, $k_f \cong 5$ W·m^{-1}·K^{-1}) with thicknesses $D_f \cong 10^{-7}$–10^{-6} m on MgO or Al_2O_3 substrates with thicknesses $D_s \cong 10^{-4}$–10^{-3} m the parameters C and K are small: $C\approx 10^{-3}$, $K\approx 10^{-5}$; therefore, we may neglect the first two terms in Eq. (10).

The substrate temperature distribution in a system of coordinates moving with the NS boundary ($z=x+u\tau$) satisfies the equation

$$\frac{\partial^2 \Theta}{\partial z^2} + \frac{\partial^2 \Theta}{\partial y^2} - u\frac{\partial \Theta}{\partial z} = 0. \tag{13}$$

Far from the front of the NS boundary, the high-T_c film is in the uniform states $\Theta(z,0)=0$ as $z\to-\infty$ and $\Theta(z,0)=p$ as $z\to\infty$, and the temperature profile in the substrate is linear. Let us pick the coordinate origin such that the condition $\Theta(0,0)=1$ is fulfilled [this is always possible due to the translation invariance of Eq. (13)]. Then the boundary conditions on Eq. (13), which are determined by Eqs. (10) and (12) taking $C\ll 1$, $K\ll 1$ into account, can be written in the form

$$\left.\frac{\partial \Theta}{\partial y}\right|_{y=0} = -p\cdot\eta(z), \qquad \left.\Theta\right|_{y=1} = 0. \tag{14}$$

The temperature distribution in the film–substrate system is determined by solving a Dirichlet-Neumann problem in the strip $0<y<1$ for the two-dimensional equation (13) with boundary conditions (14), which can be done by the method of separation of variables (see Appendix). This allows us to obtain the following expression for the velocity of propagation of the NS boundary:

$$\frac{p-2}{2p} = u\sum_{k=0}^{\infty}\frac{1}{\lambda_k^2\sqrt{u^2+4\lambda_k^2}}, \tag{15}$$

where $\lambda_k=\pi(2k+1)/2$.

It follows from Eq. (15) that the velocity of the NS interphase boundary $u=0$ at the ''propagation intensity'' $p_p=2$. From Eq. (15) it is convenient to obtain approximate expressions for the function $u(p)$ for $p\cong p_p$ and for $p\gg p_p$. For $p\cong p_p$ ($u\ll 1$) we have

$$u \cong \frac{\pi}{14\cdot\zeta(3)}(p-2) \approx 1.84(p-2), \tag{16}$$

where $\zeta(x)$ is the Riemann ζ function.

Using the Euler–Maclauren summation formula when $p\gg p_p$ ($u\gg 1$), we find from Eq. (15) that

$$u \cong (2/\pi)p. \tag{17}$$

In the range of small propagation velocities of the NS interphase boundary ($u\leqslant 1$), Eqs. (7) and (15) are almost identical, but when $u\geqslant 1$ the function $u(p)$ is qualitatively different (Fig. 2). This difference comes from the fact that for high propagation velocities $u\geqslant 1$ of the interphase boundary the substrate cannot be heated within the time of passage of the NS boundary front.

The solution to Eq. (13) with the boundary conditions (14) also allows us to obtain an estimate for the width of the NS boundary front L (see the Appendix):

$$L = \frac{4D_s}{\pi^2}\sqrt{u^2+\pi^2}. \tag{18}$$

It follows from Eq. (18) that $L\cong 4D_s/\pi^2 \sim D_s$ for $p\cong p_p$. Thus, for a high-T_c film on a thermally stabilized substrate the characteristic scale of the NS boundary front is determined by the substrate thickness $L\sim D_s$.

A HIGH-T_c FILM HEATED BY A TRANSPORT CURRENT

Let us now discuss the propagation of an NS boundary along a high-T_c film carrying a transport current. In this case,

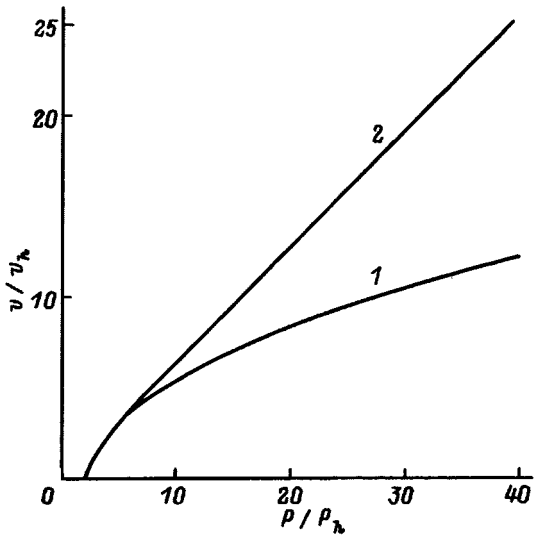

FIG. 2. Dependence of the propagation velocity of the NS boundary on the radiation intensity: *1* — calculation using Eq. (7), *2* — calculation using Eq. (15).

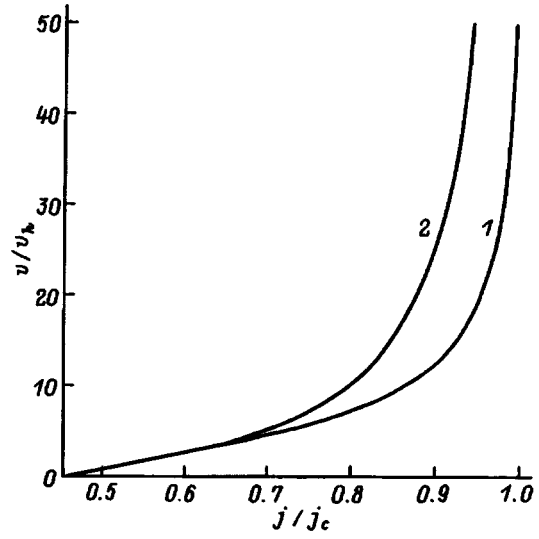

FIG. 3. Dependence of the propagation velocity of an NS boundary on the transport current density: *1* — calculated from Eq. (20), *2* — calculated from Eq. (21), $\alpha=5$.

the thermal bistabililty observed experimentally[14–17] is connected with Joule self-heating of the film. The temperature distribution in the film–substrate system is described by Eq. (1) in the one-dimensional approximation and Eqs. (8), (9) in the two-dimensional approximation. The specific heating power delivered to the film by the transport current is given by the expression[12,13]

$$Q(T)=\rho j^2\cdot\eta(T-T_r), \tag{19}$$

where j is the transport current density, $T_r=T_0+(1-j/j_c)\times(T_c-T_0)$ is the temperature of the resistive transition, and j_c is the critical current density at $T=T_0$.

From Eqs. (1), (2), (9), and (19), it follows that the dimensionless velocity of the NS boundary in a film heated by a current depends only on the single dimensionless parameter $\xi=k_s(T_r-T_0)/\rho j^2 D_f D_s$, while the expression for u can be obtained from Eq. (7) (the one-dimensional approximation) and Eq. (15) (the two-dimensional approximation) by making the replacement $p\to 1/\xi$. Thus, for u we obtain in the one-dimensional approximation

$$u=2\frac{\alpha i^2+2i-2}{\sqrt{(\alpha i^2+i-1)(1-i)}} \tag{20}$$

and in the two-dimensional approximation

$$\frac{\alpha i^2+2i-2}{\alpha i^2}=u\sum_{k=0}^{\infty}\frac{1}{\lambda_k^2\sqrt{u^2+4\lambda_k^2}}. \tag{21}$$

Here $\alpha=\rho j_c^2 D_f D_s/k_s(T_c-T_0)$ is the effective Steckl parameter of the substrate–film system, while $i=j/j_c$. From Eqs. (20) and (21) it follows that the NS boundary is at rest for a ''propagation current'' $j_p=j_c[(1+2\alpha)^{1/2}-1]/\alpha$. For typical values of the film and substrate parameters given above, along with $j_c\cong 10^{10}\ \mathrm{A\cdot m^{-2}}$, we obtain $\alpha\approx 1-10$, i.e., for a high- T_c film the quantity j_p is comparable to j_c. Near the propagation current $j\cong j_p$ ($u\ll 1$), from Eq. (21) we obtain $u\cong[2\pi^3/7\zeta(3)](j-j_p)/j_c\approx 7.36(j-j_p)j_c$, which differs only slightly from the expression $u\cong 8(j-j_p)/j_c$ that follows from Eq. (20). An important distinction arises when $j\cong j_c$ ($u\gg 1$): in the one-dimensional approximation we have $u\cong 2[\alpha i^2/(1-i)]^{1/2}$, whereas from Eq. (21) it follows that $u\cong(2/\pi)\alpha i^2/(1-i)$ (Fig. 3).

CONCLUSION

In this paper we have discussed the dynamics of S–N transitions induced by microwave radiation heating or a transport current in a high-T_c film on a thermally stabilized substrate. The systems we have discussed are characterized by significant nonuniformity of the heating over the cross section and a nonlinear distribution of temperature through the thickness of the substrate, which possesses a high thermal conductivity and plays the role of a ''heat bath'' for the high-T_c film. These facts imply that the dynamics of the thermal transition cannot be correctly described within the framework of the one-dimensional theory of propagation of an NS boundary.[12,13]

By solving the two-dimensional nonstationary heat conduction equation we obtain the function $v(P)$. When the velocity of the NS interphase boundary is small, i.e., $v\leqslant k_s C_s^{-1} D_s^{-1}$ [$P\cong P_p=2k_s(T_c-T_0)/\varkappa_n D_f D_s$], this function almost coincides with the function $v(P)$ obtained within the framework of the one-dimensional approximation. For ''fast'' propagation of the NS interphase boundary $v\geqslant k_s C_s^{-1} D_s^{-1}(P\gg P_p)$ the function $v(P)$ differs qualitatively from the one-dimensional form: in the two-dimensional approximation we have $v\propto P$, whereas the one-dimensional approximation predicts that $v\propto P^{1/2}$. This circumstance arises from the fact that when $v\geqslant k_s C_s^{-1} D_s^{-1}$ the substrate cannot be fully heated within the transit time of the NS boundary front and its temperature distribution deviates significantly from linear. An analogous feature of the propagation of the NS boundary occurs when the film is heated by a transport current.

In conclusion we note that another case is possible where the high-T_c film is placed on a substrate that is completely immersed in liquid nitrogen.[16,17] In this case, simple estimates show that the Biosystem parameter for the substrate–film system is small, i.e., $hD_s/k_s \ll 1$ (the heat-removal coefficient in nitrogen is $h \cong 10^4$ $\mathrm{W \cdot m^{-2} K^{-1}}$, $D_s \cong 10^{-3}$ m, $k_s \cong 10^2$ $\mathrm{W \cdot m^{-1} K^{-1}}$) and the temperature is practically unchanged throughout the substrate thickness. Consequently, the S–N transition can be satisfactorily described within the framework of the one-dimensional theory of propagation of the NS boundary.[16,17]

This work was carried out with the support of the Science Advisory on Problems in High-T_c Superconductivity (Project No. 93027) and the Russian Fund for Fundamental Research (Project No. 96-02-18949).

APPENDIX

Let us seek a solution to the problem (13), (14) by the method of separation of variables, which allows us to write $\Theta(z,y)$ in the form of a Fourier series:

$$\Theta(z,y) = p\,\eta(z)(1-y) + \sum_{k=0}^{\infty} f_k(z)\cos(\lambda_k y), \tag{A1}$$

where $\lambda_k = \pi(2k+1)/2$, $f_k(z)$ is an unknown function.

Expression (A1) is a general solution to Eq. (13) that satisfies the boundary condition (14). Substituting Eq. (A1) into Eq. (13) and solving the resulting ordinary differential equations, we find for the function $f_k(z)$

$$f_k(z) = \begin{cases} A_k \exp\left\{\left(\dfrac{u}{2} + \mu_k\right) z\right\}, & z < 0, \\ B_k \exp\left\{\left(\dfrac{u}{2} - \mu_k\right) z\right\}, & z > 0, \end{cases} \tag{A2}$$

where A_k, B_k are numerical coefficients and $\mu_k = (u^2/4 + \lambda_k^2)^{1/2}$.

Using the expansion

$$1 - y = 2\sum_{k=0}^{\infty} \lambda_k^2 \cos(\lambda_k y),$$

we have from Eqs. (A1) and (A2) that

$$\Theta(z,y) = \begin{cases} \displaystyle\sum_{k=0}^{\infty} A_k \exp\left\{\left(\frac{u}{2} + \mu_k\right) z\right\} \cos(\lambda_k y), \\ z < 0, \\ \displaystyle\sum_{k=0}^{\infty} \left[\frac{2p}{\lambda_k^2} + B_k \exp\left\{\left(\frac{u}{2} - \mu_k\right) z\right\}\right] \cos(\lambda_k y), \\ z > 0. \end{cases} \tag{A3}$$

The coefficients A_k and B_k are determined from the condition of continuity of the temperature and its derivative with respect to z on the line $z = 0$. Omitting intermediate calculations we present the expression for the temperature distribution in the substrate:

$$\Theta(z,y) = p \begin{cases} \displaystyle\sum_{k=0}^{\infty} \frac{1}{\lambda_k^2}\left(1 - \frac{u}{2\mu_k}\right) \\ \quad \times \exp\left\{\left(\dfrac{u}{2} + \mu_k\right) z\right\} \cos(\lambda_k y), \quad z < 0, \\ \displaystyle\sum_{k=0}^{\infty} \frac{1}{\lambda_k^2}\left[2 - \left(1 + \frac{u}{2\mu_k}\right)\right. \\ \quad \left. \times \exp\left\{\left(\dfrac{u}{2} - \mu_k\right) z\right\}\right] \cos(\lambda_k y), \quad z > 0. \end{cases} \tag{A4}$$

Taking into account that the coordinate origin was chosen so that $\Theta(0,0) = 1$, we obtain from Eq. (A4)

$$\frac{p-2}{2p} = u \sum_{k=0}^{\infty} \frac{1}{\lambda_k^2 \sqrt{u^2 + 4\lambda_k^2}}. \tag{A5}$$

From Eq. (A4) we can also obtain an estimate for the width of the NS boundary front, which obviously is determined by the smallest value of μ_k: $\mu_0 = (u^2 + \pi^2)^{1/2}/2$. Finally, for the width of the front $L = D_s/(\mu_0 + u/2) + D_s/(\mu_0 - u/2)$ we have

$$L = \frac{4D_s}{\pi^2}\sqrt{u^2 + \pi^2}. \tag{A6}$$

[1] O. G. Vendik, L. Kovalevich, A. P. Mitrofanov *et al.*, Sverkhprovod. (KIAE) **3**, 2133 (1990) [Superconductivity **3**, 1573 (1990)].
[2] K. K. Likharev, Supercond. Sci. Technol. **3**, 325 (1990).
[3] N. Newman and W. G. Lyons, J. Supercond. **6**, 119 (1993).
[4] R. C. Hansen, IEEE Trans. Aerosp. Electron. Syst. **26**, 345 (1990).
[5] O. G. Vendik, A. Karpyuk, S. G. Kolesov, and A. Yu. Popov, Sverkhprovod. (KIAE) **3**, 2161 (1990) [Superconductivity **3**, 1598 (1990)].
[6] M. M. Gaĭdukov, A. B. Kozyrev, L. Kovalevich *et al.*, Sverkhprovod. (KIAE) **3**, 2170 (1990) [Superconductivity **3**, 1607 (1990)].
[7] A. A. Zharov, A. L. Korotkov, and A. N. Reznik, Sverkhprovod. (KIAE) **5**, 419 (1992) [Superconductivity **5**, 413 (1992)].
[8] A. A. Zharov, A. L. Korotkov, and A. N. Reznik, Supercond. Sci. Technol. **5**, 104 (1992).
[9] A. N. Reznik, A. I. Smirnov, and M. D. Chernobrovtseva, Sverkhprovod. (KIAE) **6**, 242 (1993).
[10] A. N. Reznik, A. A. Zharov, and M. D. Chernobrovtseva, IEEE Trans. Appl. Supercond. **AS-5**, No. 2(III), 2579 (1995).
[11] A. M. Portis, H. Chaloupka, M. Jeck *et al.*, Supercond. Sci. Technol. **4**, 436 (1991).
[12] M. N. Wilson, *Superconducting Magnets* [Oxford University Press, London, (1983); Mir, Moscow, (1985)].
[13] A. Vl. Gurevich, R. G. Mints, and A. L. Rakhmanov, *Physics of Composite Superconductors* [in Russian], Nauka, Moscow, 1987.
[14] V. N. Skokov and V. P. Koverda, Sverkhprovod. (KIAE) **6**, 1646 (1993).
[15] V. N. Skokov and V. P. Koverda, Phys. Status Solidi A **142**, 193 (1994).
[16] M. O. Lutset, Pis'ma Zh. Tekh. Fiz. **20**, 7 (1994) [Sov. Phys. Tech. Phys. **20**, 726 (1994)].
[17] M. O. Lutset and S. V. Klimov, Sverkhprovod. (KIAE) **7**, 1372 (1994).
[18] C. Levillain, P. Manuel, and P. G. Therond, Cryogenics **34**, 69 (1994).

Translated by Frank J. Crowne

A study of the microwave dielectric permittivity of liquid crystals in electric and magnetic fields

B. A. Belyaev, N. A. Drokin, V. F. Shabanov, and V. N. Shepov

L. V. Kirenskiĭ Physics Institute, Siberian Department, Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

(Submitted May 14, 1996)
Zh. Tekh. Fiz. **68**, 117–121 (January 1998)

A microwave detector based on a self-sustained oscillator circuit is proposed as a means to investigate the real and imaginary components of the dielectric permittivity of liquid crystals in external electric and magnetic fields. Results are given for measurements of a 500 MHz oscillator frequency for two types of nematic crystals, 5CBP and MBBA. Fundamental regularities are identified in the behavior of the microwave dielectric permittivity of samples in electric and magnetic fields. It is shown that the minimum of the high-frequency dielectric loss in liquid crystals correspond to a situation in which the long axes of the molecules are oriented parallel to the direction of the microwave electric field. © *1998 American Institute of Physics.* [S1063-7842(98)02201-6]

INTRODUCTION

As a rule, the dielectric constants of liquid crystals are studied in the frequency range $f=0-10^7$ Hz, where the dispersive properties of the materials arising from the orientational mechanisms for polarizing the molecules are most strongly manifest. In contrast, in the higher frequency decimeter wavelength range ($f=10^8-10^9$ Hz) the behavior of the dielectric permittivity of liquid crystals and its interrelation with molecular orientation processes in electric and magnetic fields have been practically unstudied. This is mostly due to experimental difficulties, since traditional detectors based on lumped elements no longer work in this frequency range, while electrodynamic structures with distributed parameters are too cumbersome, and hence possess inadequate sensitivity.

In order to study liquid crystals in the decimeter wavelength range, we developed a new and original miniature microwave detector structure built around microstrip circuits.[1] We showed that microstrip measurement cells can be used to accurately identify changes in both the real and imaginary components of the dielectric permittivity of liquid crystals subjected to an external electric field, even for quite small liquid crystal samples (with volumes $\simeq 10^{-3}$ cm^{-3}). The thickness of the sample layer in the microwave detector is determined by a gap between measurement plates, and ordinarily is $\simeq 100$ μm. This is comparable to the thickness of real liquid crystal cells used in various practical devices. Gaps of this size allow us to obtain important information about the influence of surfaces that bound the liquid crystal sample on the behavior of the dielectric characteristics of the material.

The microstrip detectors described in Ref. 1 were constructed for operation in tandem with standard devices for measuring amplitude-frequency characteristics. In this paper we discuss complete microwave detector structures in which the microstrip measurement resonator is included as part of the frequency-defining loop of a microwave self-sustained oscillator circuit. This arrangement significantly increases the sensitivity of the system and the accuracy of the measurements. The operation of the system is illustrated by the results of experimental studies of the behavior of the dielectric constant of two typical nematogens: 4-n-pentyl-4- cyanobiphenyl (5CBP) and 4-methoxybenzinidene-4- butylaniline (MBBA) under the action of dc electric and magnetic fields. Discussion of the results obtained is based on descriptions of processes whereby electric and magnetic fields orient the dipole molecules of the liquid crystals and change the value of the microwave dielectric loss connected with them.

SELF-SUSTAINED OSCILLATOR DETECTOR CIRCUIT

The frequency-defining loop of the microwave detector self-sustained oscillator circuit (Fig. 1) is a microstrip ''ring'' resonator that includes the vertical measurement plates.[1] The oscillator is placed in a metal package with dimensions $30\times24\times15$ mm. The resonator structure itself, which serves as the top cover of the package, is made on a substrate of polycor ($\varepsilon=9.8$) with thickness $h=1$ mm, dimensions 24×30 mm. In this case the lower metallized side of the substrate, which we referred to as the ground plane, was soldered to the walls of the metal package of the detector along its entire perimeter. The picture shown in Fig. 1 also includes the wires on the upper side of the substrate. Gold-plated measurement plates (a) with dimensions 2.5 $\times$2.5 mm were soldered to the metal pads at the ends of the striplines of the microstrip resonator vertical to the plane of the substrate so that the measurement gap had a value of 100 μm. A liquid crystal placed in the gap is subjected to surface tension forces. In order to decrease the effect of the capacitance between the edges of the detector package and the measurement plates, the latter were located close to the center of the substrate, at the expense of bending the strip lines. To do this, a rather small portion of the metallized screen was removed from the backside of the substrate directly under the detector measurement plates by chemical etching.

The self-sustained oscillator circuit was tuned for exci-

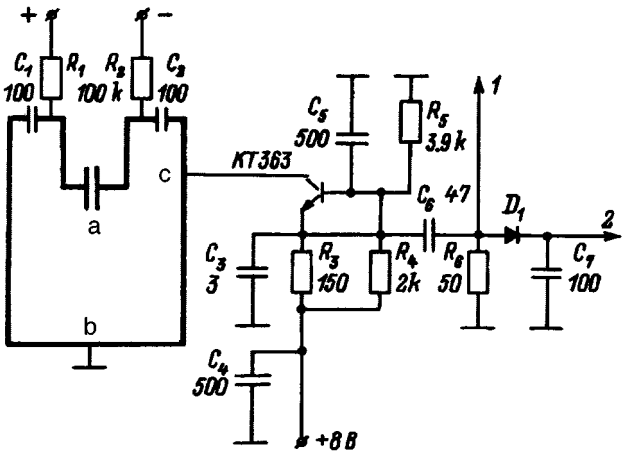

FIG. 1. Basic circuit for the self-sustained oscillator detector. *1*—output (frequency), *2*—output (amplitude).

tation of the fundamental half-wave mode of oscillation of the microstrip resonator. In order to feed the transistor with a constant current, the upper strip of the resonator is short-circuited by a jumper to the package at a point of minimum high-frequency voltage (point b in Fig. 1). In this case the collector of the transistor was connected directly to the strip of the microstrip resonator at point (c) in order to create conductive (autotransformer) coupling with the resonator at high frequencies. An orienting electric field was created across the measurement plates by applying a dc voltage to the latter through the decoupling resistors R_1 and R_2. In order to galvanically decouple the power supply we used divider capacitors C_1 and C_2. The microwave oscillator was designed to have two outputs: one for measuring frequency, and the other for measuring the amplitude of the self-sustained oscillations. The circuit was tuned by setting up optimum conductive coupling between the resonator and the microwave transistor, i.e., we looked for an optimum point to connect the transistor collector to the strip line of the resonator (c), and carefully chose point (b) near the midpoint of the stripline of the microwave resonator as ground. Pouring liquid crystal into the measurement gap significantly reduced the Q-factor of the resonator half- wave mode, and we found that the oscillator spontaneously switched to higher-Q ''parasitic'' resonances. In order to avoid this instability we needed to position the circuit components compactly near the resonator and choose a capacitance C_3 to shunt the higher resonances to ground. The working frequency of the self-sustained oscillator microwave detector, which was measured using a digital frequency meter, was around 500 MHz when the sample of liquid crystal was poured into the gap. In this case the microwave oscillation amplitudes measured by a digital volt meter at the output to the microwave detector were found to be about 0.1 V.

Note that all our experiments on liquid crystals were carried out at a temperature $T=24\pm0.1$ °C. Therefore, the microwave detector was placed in a miniature thermostat, which was placed between the poles of an electromagnet. The magnetic field could be varied up to a value $H=2.3$ kOe. The self-sustained oscillator package could be rotated between the end-face poles of the magnet so that the direction of the field H was either parallel or perpendicular to the direction of the high-frequency and external dc electric fields.

SAMPLES AND METHODS OF INVESTIGATION

As we have already mentioned, as samples for our investigations we used two widely known nematogens—the compound 5CBP, with a relatively high positive dielectric anisotropy,[2] and the weakly anisotropic liquid crystal MBBA with negative anisotropy.[3–5] Our measurement method was the following: the liquid crystal was placed in the gap between the measurement plates of the self-sustained oscillator microwave detector, and was kept there for an hour in order for it to reach its equilibrium state. Within this time, the temperature of the detector stabilized and initial values of the frequency and amplitude of the self-oscillations were established. For each sample of liquid crystal a magnetic field H was applied, and the dependence of the relative frequency change $F(H)$ was measured along with the relative change in amplitude $P(H)$ of the detector self-sustained oscillations in the steady-state regime. Analogous functions $F(U)$ and $P(U)$ were measured when an electric potential U was applied to the measurement plates of the detector. We also investigated the dependence of the changes in the microwave parameters of the self-sustained oscillator under the combined action of electric and magnetic fields on the liquid crystal.

For an optimal choice of the structural coupling between the microstrip resonator and the microwave transistor, the level of self-oscillations in the oscillator depended linearly on the Q-factor of the frequency-determining resonator, which in turn was related in this experiment to the dielectric loss tangent of the liquid crystal sample. As a result, the observed change in the microwave oscillation amplitude $P(U,H)$ was inversely proportional to the change in the imaginary component of the dielectric permittivity of the sample $\Delta\varepsilon''$. As is well known, changing the real component of the dielectric permittivity of the liquid crystal sample $\Delta\varepsilon'$ leads to a change in the resonant frequency of the self-sustained oscillator detector: the frequency increases as the dielectric permittivity ε' decreases, and conversely. Consequently, $F(U,H)\sim(\Delta\varepsilon')^{-1}$.

RESULTS OF EXPERIMENTAL INVESTIGATIONS

Figure 2 shows the dependence of the relative change in the level of microwave oscillations of a detector containing a sample of 5CBP. For convenience of comparison the dependences on electric potential $P(U)$ and on magnetic field $P(H)$ are shown on the same figure. Analogous curves plotted for a sample of MBBA are shown in Fig. 3. Curves *1* on these figures were plotted for no magnetic field $(H=0)$, while curves *2* and *3* were plotted for no electric field $(U=0)$. The dependences 2 were plotted for the case where the direction of the dc magnetic field H was parallel to the direction of the high-frequency field e in the measurement gap of the detector, while the curves *3* were opposite, i.e., the direction of the field H was perpendicular to the direction of

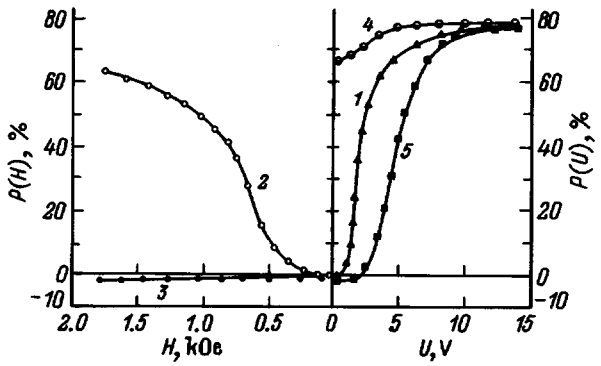

FIG. 2. Dependence of the relative level of microwave oscillations in a detector with a sample of 5CBP on the electric potential $P(U)$ and magnetic field $P(H)$.

e. Moreover, the functions $P(U)$ were plotted while a magnetic field $H=2.3$ kOe acted simultaneously on the sample; this field was directed either parallel to e (curves *4)* or perpendicular to e (curves *5)*.

The features observed in the behavior of the curves $P(U)$ and $P(H)$ are connected with the well-known second-order orientational phase transitions that have been given the name Fredericks transitions.[5] These transitions are a consequence of the competition between two forces that act on molecules of the sample, one exerted by the walls of the cell that bound the liquid crystal and the other by the external field. The primary feature of a Fredericks transformation is the presence of critical fields above which the initial orientational configuration of the molecules created by the walls becomes unstable and the director of the molecules shifts to a new state determined by the magnitude and direction of the external fields. For liquid crystals of nematic type with anisotropies in the dielectric permittivity $\varepsilon_a=\varepsilon_\parallel-\varepsilon_\perp$ and magnetic susceptibility $\chi_a=\chi_\parallel-\chi_\perp$, the relation between the critical electric fields E_c and magnetic fields H_c can be described by the following expression[6]:

$$\frac{1}{2}\chi_a H_c^2=\frac{\varepsilon_a E_c^2}{8\pi}. \tag{1}$$

From Figs. 2 and 3 it is clear that the critical fields are easy to determine from the experimental results. For the 5CBP samples the field $H_c\approx 120$ Oe, and the field $E_c\approx 3$ V/mm. Since the anisotropy of the diamagnetic susceptibility of the crystal χ_a is a quantity of order 10^{-7}, Eq. (1) allows us to infer that the anisotropy of the dielectric permittivity $\varepsilon_a\approx 2$. This value of ε_a is in good agreement with the results of other measurements on liquid crystals of this type.

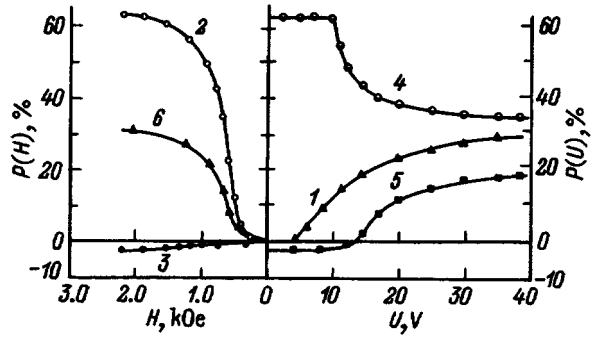

FIG. 3. Dependence of the relative level of microwave oscillations of a detector with a sample of MBBA on the electric potential $P(U)$ and magnetic field $P(H)$.

The measurements on the samples of liquid crystal MBBA (Fig. 3) show that the electric and magnetic critical fields are almost an order of magnitude larger in these samples than in the crystal 5CBP. This is entirely natural, since it is well known that the crystal MBBA is only weakly anisotropic.

By analyzing the behavior characteristics of the electric and magnetic dependences $P(U)$ and $P(H)$ in fields above the critical fields we can obviously determine qualitatively the initial configuration of the molecular states and the processes that orient them under the action of these competing forces. Note that the orienting action of the magnetic field is connected with the diamagnetic component of the susceptibility caused by the presence of benzene rings in the composition of the molecules. The minimum in magnetic energy corresponds to a state where the plane of the benzene rings lies in the direction of the magnetic field. This implies that in this type of liquid crystal the long axis of the molecule is oriented parallel to the magnetic field.

Let us first discuss the effect of dc magnetic and electric fields on the value of the microwave dielectric loss for the liquid crystal 5CBP. As is clear from Fig. 2, when the external magnetic field is parallel to the high-frequency field e, the amplitude of the self-sustained oscillations of the detector increase monotonically with increasing H up to saturation (curve *2)*. In this case the long axes of the dipole molecules are rotated normal to the plane of the measurement plates, i.e., parallel to the high-frequency field e, which appreciably decreases the microwave dielectric loss in the liquid crystal sample.

However, in the opposite case, when the direction of the magnetic field $H\perp e$, the amplitude of the self-sustained oscillations, and consequently the dielectric loss in the liquid crystal sample, are almost unchanged (curve *3* in Fig. 2). Obviously this direction of the magnetic field should orient the long axes of the molecules along the measurement plates, that is the long axes of the molecule should rotate perpendicular to the direction of the high-frequency field. As a result, the dielectric loss should increase and the amplitude of the self-sustained oscillations should decrease. The fact that experiments do not reveal any significant change in $P(H)$ in this situation is evidence that all the dipoles are already oriented along the walls of the measurement plates due to the orienting forces exerted on the molecules by the surface.

Since ε_a is positive for the liquid crystal 5CBP, the parallel orientation of the director along the microwave field should be established even when the external field applied to the measurement plates acts on the sample as well. And, in fact, as is clear from Fig. 2, the behavior of the function $P(U)$ (curve *1)* shows practically no difference from that of the corresponding ''magnetic'' curve *(2)*. Based on what we have said above, it is not difficult to explain the behavior of

the ''electric'' curves *4* and *5*, which were plotted with the maximum magnetic field $H=2.3$ kOe acting on the sample at the same time. Obviously the strong magnetic field applied parallel to the high-frequency field in this experiment almost completely orients the liquid crystal, rotating the long axes of the molecules along the direction of e. As a result, as the additional electric field applied in this direction increases we observe only a rather small rotation of the molecular axes; therefore, there is only a slight increase in the oscillation amplitude $P(U)$ (curve *4)*.

However, for a perpendicular applied magnetic field the long axes of the molecules are oriented almost perpendicular to the direction of e. As a result, we require a considerably larger dc electric field to rotate the molecular axis by 90°, because it is necessary to overcome not only the orienting forces exerted by the surfaces of the measurement plates but also the orienting forces exerted by the magnetic field. Therefore, in these experiments we observe a significant increase in the value of the critical field E_0 (Fig. 2, curve *5)*.

The liquid crystal MBBA differs from the crystal 5CBP by the negative anisotropy of its dielectric permittivity ($\varepsilon_a=-0.59$); however, the anisotropy of the diamagnetic susceptibility $\chi_a=1.23\times10^{-7}$ in these materials is positive. As a result, the long axes of the MBBA molecules are oriented along the direction of the magnetic field, just as in the crystal 5CBP, while the ''magnetic'' functions shown in Fig. 3 (curves *2, 3)* are very similar to the analogous curves for 5CBP (Fig. 2). The magnetic field parallel to the high-frequency field orients the axis of the molecules along e; therefore, the dielectric loss decreases to a minimum (curve *2)*. Keeping in mind the fact that increasing the magnetic field $H_\perp$ causes only a slight change in of $P(H)$ in this experiment (Fig. 3, curve *3)*, we may assert that the orienting forces exerted by the walls of the measurement plates are primarily directed in the planes of the plates for MBBA, just as for 5CBP.

However, in contrast to the previous cases, when we simultaneously apply the maximum magnetic field $H_\parallel$ to the sample of MBBA crystal and increase the voltage U, we observe a slight decrease in the value of $P(U)$, i.e., an increase in the dielectric loss in the crystal (curve *4)*. This implies that in this case the electric field causes the direction of the molecular axes to deviate from the direction of the parallel orienting magnetic field. This anomalous behavior of $P(U)$ is explained by the fact that the parallel and perpendicular components of the dielectric permittivity in the MBBA crystal differ only slightly from one another: $\varepsilon_\parallel=5.17$, $\varepsilon_\perp=4.58$.[5] In this case the direction of the resulting polarization vector is determined as the vector sum of these two quantities. Since the polarization vector at large U is oriented along the dc electric field, the long axis of the molecule turns out to be rotated relative to its original state at $U=0$. As a result, the microwave dielectric loss increases accordingly. The angle through which the long axes of the molecules rotate can be determined from the relation

$$\varphi=\arctan(\varepsilon_\perp/\varepsilon_\parallel)=42.34°. \quad (2)$$

As a proof of what was said above, in Fig. 3 we show curve *6*, which was plotted for a magnetic field oriented at the computed angle $\varphi\approx42°$. It is clear that the observed change in the quantity $P(H)$ is in rather good agreement with the maximum change in $P(U)$, which confirms that an oblique orientation of the director forms in the electric field. The rather small values of the dielectric anisotropy in the liquid crystals MBBA, as we already mentioned above, causes the critical field to increase significantly to $E_c=50$ V/mm (curve *1)*, and to increase to $E_c=140$ V/mm (curve *5)* when a perpendicular magnetic field acts simultaneously on the sample.

Note that in all the experiments the relative change in the self-sustained oscillator frequency did not exceed a value of $F\approx0.6\%$. Decreasing the microwave loss in the liquid crystal samples under the action of electric or magnetic dc fields, as a rule, led to a decrease in the frequency of self-sustained oscillations of the detector, i.e., to an increase in the real part of the dielectric permittivity, and conversely. Thus, these experiments show that even in the decimeter wavelength range the processes that orient the molecules are accompanied by rather large changes not only in the imaginary part but also the real part of the components of the microwave dielectric permittivity of these liquid crystal samples.

CONCLUSIONS

The self-sustained oscillator detector scheme described in this paper for studying the properties of liquid crystals at microwave frequencies allows us to identify operationally and with good accuracy the relative changes in both the real and imaginary components of the high-frequency dielectric permittivity of samples subjected to external fields. The high sensitivity of this method reveals even greater potential for performing new and very precise experiments connected with the study of orienting processes in liquid crystals, and also distinctive features of the interaction between liquid crystal molecules and surfaces made from various materials that bound the sample.

In this paper we have shown that the changes in microwave dielectric characteristics of the liquid crystals 5CBP and MBBA are uniquely related to processes that orient the dipole molecules. We have also shown that the case where the microwave electric field is parallel to the longitudinal component of the dielectric permittivity of the liquid crystal corresponds to a minimum dielectric loss. In this situation we established experimentally that the value of the dielectric losses in liquid crystal samples at microwave frequencies are comparatively large, although, as follows from other papers (see Refs. 2–4,6) the dielectric permittivity of liquid crystals has the tendency to decrease rapidly as the frequency increases due to the large inertia of the polarization oscillations of the molecules. From this point of view, in the microwave range a liquid crystal should have only insignificant dielectric losses approaching those of the optical absorption coefficient in magnitude.

It would be logical to assume that strong absorption of microwave power observed in these experiments is caused by charging of ion complexes or impurities that are present in liquid crystals,[7,8] whose concentration can be quite high. However, this assumption does not explain the fact that dielectric loss in the experiments changes significantly when

the molecules of these crystals are oriented in dc electric and magnetic fields. Therefore, it is most likely that the nature of the observed high losses in liquid crystal samples at microwave frequencies are connected with the broad spectrum of relaxation oscillations of individual ring molecular groups connected with the benzene rings, the so-called ''tails.''

[1] B. A. Belyaev, N. A. Drokin, and V. N. Shepov, Zh. Tekh. Fiz. **65**(2), 189 (1995) [Tech. Phys. **40**, 216 (1995)].
[2] P. G. Cummins, D. A. Dunmur, and D. A. Laidler, Mol. Cryst. Liq. Cryst. **30**, 109 (1975).
[3] F. V. Bulygin, Vestn. Mosk. Gos. Univ., No. 5, Ser. 3, 77 (1988).
[4] B. S. Saburov, U. Mamabaev, I. U. Rodzhab, and P. V. Adomenas, Dokl. Akad. Nauk SSSR **306**, 1393 (1989).
[5] L. M. Blinov, *Electro-Optical and Magneto-Optical Properties of Liquid Crystals* [Wiley, New York, 1983; Nauka, Moscow, 1978], p. 121.
[6] P. G. DeGennes, *The Physics of Liquid Crystals*, 1st ed. [Clarendon Press, Oxford, 1993; Mir, Moscow, 1982], p. 115.
[7] N. I. Gritsenko, I. N. Gulenko, and N. V. Moshel', Izv. Vyssh. Uchebn. Zaved. Fiz. **7**, 9–12 (1989).
[8] G. J. Spokel, Mol. Cryst. Liq. Cryst. **22**, 249 (1973).

Translated by Frank J. Crowne

Linear topological defects in electromagnetic vector fields

M. O. Sopin

Yu. Fed'kovich State University, 274012 Chernovtsy, Ukraine
(Submitted June 8, 1996)
Zh. Tekh. Fiz. **68**, 122–124 (January 1998)

An analysis is made of the topological structure of an electromagnetic vector field near the point where the amplitude of the field vanishes. Linear topological defects in the form of dislocations of the wave front and disclinations are studied. It is shown that the polarization of the field near a zero of the amplitude differs from the initial value. The structural stability of the amplitude zeros is studied. © *1998 American Institute of Physics.* [S1063-7842(98)02301-0]

1. In 1931, Dirac drew attention to the fact that the phase of a wave function, by its very meaning, is defined only modulo 2π. Dirac considered the exceptional case encountered when a wave function goes to zero — in that case the phase of the wave function becomes meaningless.[1] A situation where the phase is indeterminate was considered in Ref. 2 with reference to linear scalar waves. The phase singularities which are lines in space or points on a plane about which the phase undergoes a jump of 2π were called wave-front dislocations. It was shown that the dislocations are the fine structure of the wave field in the sense that they reveal unusual phase topology on a scale level determined by the wavelength. Amplitude zeros and associated wave-front dislocations have been studied in statistical fields[3,4] and in fields having a deterministic nature.[5] Note that the problem was analyzed in the scalar approximation. An attempt to allow for the vector character of the speckle field was made in Ref. 6. In the present paper it is shown that allowance for the vector properties of the electromagnetic field yields interesting new results which are not found in scalar waves.

2. Therefore, let us consider a monochromatic electromagnetic wave field possessing spatial inhomogeneity. We shall position the origin at a certain point in space where the amplitude of the field vanishes, directing the x_3 axis along the wave vector **k**. Because of the wave character, the vector potential of the field may be taken in the form of a two-component function

$$\mathbf{A}=\begin{bmatrix} u_1(x_1,x_2,x_3)e^{i\theta_1} \\ u_2(x_1,x_2,x_3)e^{i\theta_2} \end{bmatrix} e^{i(\omega t-kx_3)}. \tag{1}$$

Here θ_1 and θ_2 are certain real constants. As usual, the field obeys the wave equation and the additional gauge condition[7]

$$\Box\mathbf{A}=0, \tag{2}$$

$$\operatorname{div}\mathbf{A}=0. \tag{3}$$

Assuming that the components of the vector potential are smooth and considering for generality the case of an m-fold zero, we express the field amplitude in a certain neighborhood of the zero point in the form

$$u_{1,2}(x_1,x_2,x_3)=\sum_{j_1,j_2,\ldots,j_m=1}^{3} f^{(1,2)}_{j_1j_2\ldots j_m}x_{j_1}x_{j_2}\ldots x_{j_m}. \tag{4}$$

Here $f^{(1,2)}_{j_1j_2\ldots j_m}$ are constant coefficients. Satisfaction of conditions (2) and (3) has the result that $u_{1,2}$ are harmonic functions of two variables x_1 and x_2. Two cases may then be distinguished.

3. Let us assume that $f^{(1,2)}_{j_1j_2\ldots j_m}$ are complex quantities. This allows $u_{1,2}$ to be considered as a function of a complex variable. It is natural to impose the constraint that these functions are analytic (if x_1+ix_2 is considered to be variable) or antianalytic (when x_1-ix_2 is variable). Thus, two types of local solutions are constructed, which have the following form in the cylindrical coordinate system (ρ,φ,x_3):

$$\mathbf{A}^{\pm}=a\rho^m\begin{bmatrix} 1 \\ \pm i \end{bmatrix} e^{i(\omega t-kx_3\pm m\varphi)}. \tag{5}$$

Here a is a complex constant which is unimportant to us. We assume that the vectors **k**, **E**, and **H** are pairwise orthogonal ($\mathbf{E}=-\partial\mathbf{A}/c\partial t$ and $\mathbf{H}=\operatorname{curl}\mathbf{A}$ are the electric and magnetic field strengths). The two types of solution correspond to field states with right and left circular polarization. The phase surfaces of the solutions are helicoids of opposite twist, where the direction of twist and the character of the circular polarization are rigidly interrelated. As a result of the analyticity (antianalyticity) of the components of the complex amplitude, the zero is an isolated point in the $x_3=0$ plane but may describe some curve in three-dimensional space. This linear phase singularity, being a carrier of an amplitude zero, will be a screw dislocation line. Topologically stable solutions correspond to lines having no beginning or end.

The nontriviality of the topological structure of the electromagnetic field near the zero can be seen if we consider the plane field of the phase gradient $\mathbf{w}=(\partial_1\Psi,\partial_2\Psi)$. The integral trajectories of this field are closed and describe planar vortices. The topological invariant is

$$Q\equiv\frac{1}{2\pi}\oint w^{-2}(w_1dw_2-w_2dw_1)=m,$$

which expresses the number of rotations of the vector **w** about the zero point.[8] For physicists, this quantity has been called the topological charge and in this case may have integer values $Q=0$ (background), $Q=+1$ (monopolar vortex),

$Q=+2$ (dipolar vortex), and so on. It can be seen from Ref. 5 that unlike topological charges may be assigned to the two types of solution.

4. Let us assume that $f^{(1,2)}_{j_1 j_2 \cdots j_m}$ are real quantities. The vector potential near the zero then has the form

$$\mathbf{A}=\rho^m\begin{bmatrix}\alpha_1 \cos m\varphi+\alpha_2 \sin m\varphi\\ \alpha_2 \cos m\varphi-\alpha_1 \sin n\varphi\end{bmatrix}e^{i(\omega t-kx_3)}. \quad (6)$$

Here, α_1 and α_2 are real constants which are unimportant to us. We again impose the constraint that the set of three vectors **k**, **E**, and **H** should be pairwise orthogonal. An important difference from the previous case is that the polarization is linear near this zero.

The nontriviality of the topological structure of the electromagnetic field near the zero can be seen if we consider the invariant

$$Q\equiv\frac{1}{2\pi}\oint_{x_3,t=\text{const}} A^{-2}(A_1 dA_2-A_2 dA_1)=-m.$$

For $m=1$ the vector field is nondegenerate at the zero point.[9] In this case, the zero, being an isolated point in the plane $x_3=0$, may describe some curve in three-dimensional space. The topologically stable solutions correspond to lines without a beginning or end. These lines, where the direction of the vector is indeterminate, are called disclinations, and they will be carriers of a zero of the amplitude in this particular situation. The topological charge can generally have negative integer values: $Q=0$ (background), $Q=-1$ (monopolar charge), $Q=-2$ (quadrupolar charge), and so on.

5. It is known[9] that a plane smooth vector field defines some continuous map of a circle in a circle. The introduction of the topological charge allows the set of these maps to be divided into nonintersecting classes of equivalence in terms of Q. This separation makes the problem of the creation (annihilation) of topological defects nontrivial. Nevertheless, considering the merging of two dislocations of the same strength but of opposite twist, we obtain

$$\mathbf{A}^+ + \mathbf{A}^- = 2a\rho^m\begin{bmatrix}\cos m\varphi\\ -\sin m\varphi\end{bmatrix}e^{i(\omega t-kx_3)}. \quad (7)$$

It can be seen that the dislocations annihilate to create a disclination having a zero of the corresponding order. A mechanism for the creation of two unlike dislocations from a point zero of the amplitude point was indicated in Ref. 4. Our example implies that there can be conjugation of disclinations with the ends of unlike dislocations, or, in other words, pairs of dislocations may originate from disclinations and terminate in disclinations.

6. Finally, we make some observations on the reason for the structural stability of the amplitude zeros of the electromagnetic vector field under study. We recall that a function is described as structurally stable if its critical points do not change their type for any sufficiently small smooth perturbations.[10] It follows from Eq. (3) that a smooth function $F(x_1,x_2)$ can always be found such that the amplitude components of the vector potential of the field will be expressed in terms of the derivatives of this function $u_1=\partial F/\partial x_2$, $u_2=-\partial F/\partial x_1$, and as before $u_3\equiv 0$. Formally, this follows from the possibility of introducing a symplectic structure defined by the 2-form $dx_1/\!\wedge dx_2$ near the zero point. In this case, the amplitude of the vector potential is expressed in the form of the skew-symmetric gradient $\mathbf{u}=s$ grad F. The zero of the field amplitude is therefore the critical point of the function $F(x_1,x_2)$. It is known that the critical point is structurally stable when and only when it is nondegenerate, and the degeneracy of the zero is then determined by the rank of the Hessian matrix $\|\partial^2 F(0,0)/\partial x_j \partial x_k\|$ (Ref. 10). It is easy to see that only the simple amplitude zeros are structurally stable (the function F has the Morse form) while the zeros of higher multiplicity do not possess this property. In other words, a small perturbation (of the boundary conditions, for example) only shifts the position of a simple zero in the plane $x_3=$ const, without destroying it. A zero of higher multiplicity may be destroyed by such a perturbation, as was noted for a dislocation in Ref. 11.

7. Thus, in spatially inhomogeneous electromagnetic wave fields a distinction should be made between amplitude zeros associated with wave-front dislocations and amplitude zeros associated with disclinations. Although both are topologically stable formations, only the simple amplitude zeros possess properties of structural stability. In this sense, the zeros of higher multiplicity are atypical topological objects. Regardless of the nature of the field, its polarization characteristics are strictly determined near an amplitude zero, the polarization being circular near one type of zero and linear near the other type. The presence of topological objects — dislocations and disclinations — converts a simply connected manifold into a multiply connected one. The quantitative characteristic of these linear defects is a certain invariant — the topological charge. Note that local properties of a general type of field have been considered here. The size of the zero neighborhood in this formulation of the problem will be determined by the characteristic parameter of the field — the wavelength. In our next study we shall show that the two types of zeros considered here by no means exhaust all the fine-structure possibilities of an electromagnetic wave field.

[1] P. A. M. Dirac, *On the Development of Quantum Field Theory: Major Articles 1925–1958* [in Russian], Nauka, Moscow (1990).
[2] M. V. Berry, *Singularities in Waves and Rays*, Lectures at 1980 Les Houches Summer School (North-Holland, Amsterdam, 1981).
[3] N. B. Baranova and B. Ya.Zel'dovich, Zh. Eksp. Tekh. Fiz. **80**, 1789 (1981) [Sov. Phys. JETP **53**, 925 (1981)].
[4] N. B. Baranova, B. Ya.Zel'dovich, A. V. Mamaev *et al.*, Zh. Eksp. Tekh. Fiz. **83**, 1702 (1982) [Sov. Phys. JETP **56**, 983 (1982)].
[5] N. N. Rozanov, Opt. Spektrosk. **75**, 861 (1993) [Opt. Spectrosc. J. **75**, 510 (1993).
[6] O. V. Angelsky, R. N. Besaha, I. I. Mokhun, and M. O. Sopin, Proc. SPIE **2647**, 75 (1995).
[7] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields*, 4th ed. [Pergamon Press, Oxford, 1975; Nauka, Moscow, 1973].
[8] V. I. Arnol'd, *Ordinary Differential Equations* [MIT Press, Cambridge, Mass, 1973; Nauka, Moscow, 1975].
[9] J. W. Milnor and A. Wallace, *Differential Topology* [Russian trans., Mir, Moscow, 1972].
[10] R. Gilmore, *Catastrophe Theory for Scientists and Engineers* [Wiley, New York, 1981; Mir, Moscow, 1984].
[11] M. A. Bolshtyanskiĭ, Opt. Spektrosk. **79**, 512 (1995) [Opt. Spectrosc. J. **79**, 475 (1995).

Translated by R. M. Durham

Liquid-crystal diode generator of low-frequency oscillations

Ya. Barton' and A. A. Kal'nin

St. Petersburg State Electrical Engineering University, 197376 St. Petersburg, Russia
(Submitted December 29, 1995)
Zh. Tekh. Fiz. **68**, 125–127 (January 1998)

The excitation of rhythmic current oscillations in a diode cell containing a nematic liquid crystal is studied. The external electric field in the interelectrode gap is directed parallel to the surfaces which orient the liquid crystal molecules. The current oscillations are accompanied by the formation of an autosoliton at the cathode, which propagates and disappears at the anode. A hypothetical model is proposed to explain this current instability. © *1998 American Institute of Physics.* [S1063-7842(98)02401-5]

The present investigation is concerned with a dynamic effect accompanied by current instability in nematic liquid crystals. It has been established that under nonequilibrium conditions various types of instabilities appear in nematics.[1] A classical example is the formation of Kapustin–Williams domains in an external electric field transverse to the surfaces orienting the liquid crystal molecules.[2] This effect is a manifestation of self-organization in open systems[3] and is characterized by a threshold external electric field and by the establishment of spatial ordering at the macroscopic level. The nonlinearity caused by the anisotropy of the liquid crystal properties creates preconditions for the action of an autocatalytic mechanism (internal positive feedback) which promotes the buildup of instabilities.[4] Studies of nematic liquid crystals have mainly been carried out in a transverse field, and comparatively few have examined the response of these crystals to the action of a longitudinal field, i.e., acting parallel to the molecule-orienting surfaces.

Here we investigate the current instabilities in a planar diode structure and we attempt to develop a simple economical generator of current oscillations in the biorhythmic frequency range. This necessitated finding the conditions for the excitation of current instability in a nematic liquid crystal. As a result of carrying out numerous probe experiments in which the orientation of the liquid crystal molecules relative to the direction of the external electric field, the size, shape, and relative position of the electrodes in the liquid crystal, and the microgap between the surfaces defining the molecular orientation were varied, we succeeded in identifying the conditions required for the buildup of current instability.

The design of a planar diode cell is shown schematically in Fig. 1. Two metal electrodes were placed in the gap between two glass or quartz plates *1* and the space between the electrodes was filled with an MBBA nematic liquid crystal or a mixture of MBBA and EBBA. The gap between the plates was 50–100 μm and that between the electrodes was 0.1–1.0 mm. The width of the electrodes was 0.1–0.5 mm. The surfaces of the dielectric plates wetted by the liquid crystal were preprocessed to produce a microtexture to control the orientation of the liquid crystal molecules. The surfaces were first carefully cleaned to remove any contaminants using a 1M KOH solution, followed by a 50% (mol) HNO_3 solution, and were then washed with distilled water.

The electrodes of the diode cell were connected to a power supply via a series-connected resistance $R_s = 1.2$ MΩ to allow the current instabilities to be observed with an oscilloscope.

The results of the experimental observations are as follows:

1. As the interelectrode voltage is increased, a domain structure forms but it is less clearly defined than that obtained when the field is perpendicular to the planes of the dielectric plates.

2. As the voltage is increased further to a certain critical value U_c, a region exhibiting strong light scattering is formed at specific points on the cathode surface. This region appeared at the cathode and propagated as a traveling wave (autosoliton) toward the anode where it disappeared at the surface. The nucleation, propagation, and disappearance of the autosoliton was a periodic process. As soon as this region of anomalous light scattering disappeared at the surface of the anode, it reappeared again at the cathode.

3. Current oscillations with a large difference between their extremum values appeared, matched with the autosoliton motion. Typically, the maximum current was 3–5 μA and the minimum was 0.2–0.5 μA. Before the oscillations appeared, the direct current was some tens of nanoamperes (prethreshold regime). The threshold external field was $E_c = (1-2)\times 10^3$ V/cm (interelectrode gap $l = 1$ mm).

Oscilloscope traces demonstrating the current oscillations are shown in Fig. 2, where it can be seen that the amplitude of the oscillations and the frequency depend on the voltage. These dependences are plotted in Fig. 3.

Observations using an optical microscope showed that the current has a maximum at the beginning of the autosoliton and a minimum when it reaches the anode. As the external voltage is increased further, the current oscillations become stochastic at the upper level of the average.

4. The oscillations only occur for small gaps h between the plates ($h \leqslant 150$ μm). In this case, the orientation of the liquid crystal director has a strong influence. The investigations showed that no current oscillations occurred in samples for which the director was oriented parallel and perpendicular to the external field. Experiments with homeotropically oriented liquid crystal molecules also yielded no positive result. A traveling autosoliton regime was only observed in a twisted nematic (and was particularly clearly defined when the angle of relative rotation of the plates was close to 180°).

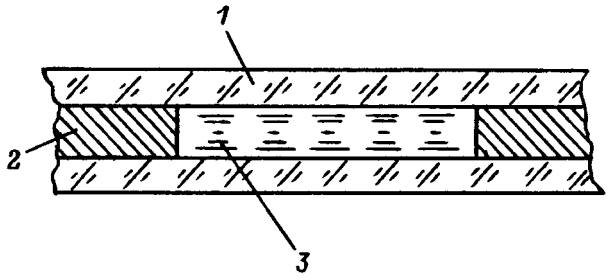

FIG. 1. Schematic of nematic-liquid-crystal diode generator of electrical oscillations: *1* — dielectric plates with surface geometric anisotropy, *2* — metal electrodes, and *3* — nematic liquid crystal.

Thus, the homogeneous orientation of the liquid crystal molecules must be impaired for these oscillations to occur.

5. In the experiments, we observed cases where the autosolitons did not propagate along the shortest path from the cathode to the anode but along a curvilinear ''trajectory'' or

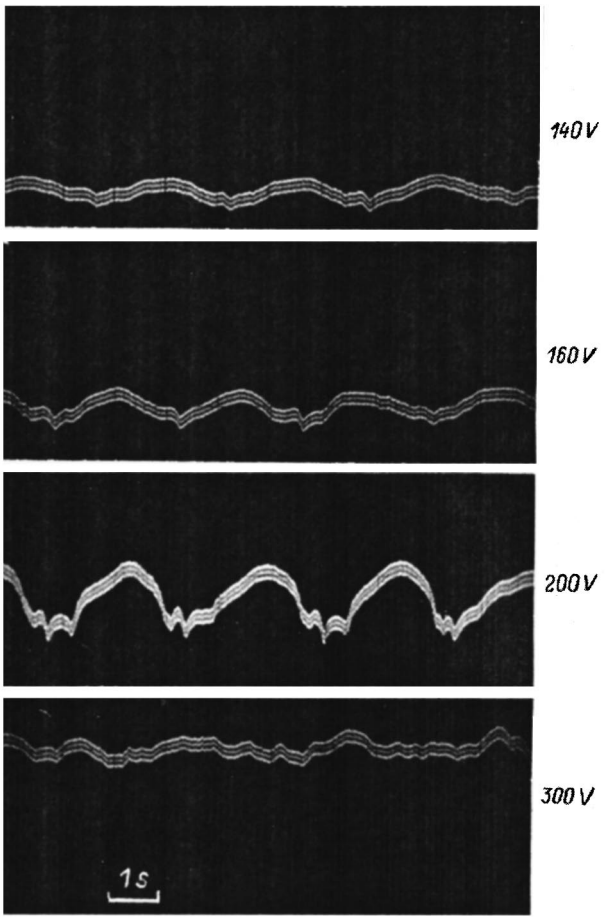

FIG. 2. Oscilloscope traces of current obtained for different voltages on electrodes of diode cell. The voltages are given on the appropriate traces; MBBA, $l=1$ mm, and $d=100$ μm.

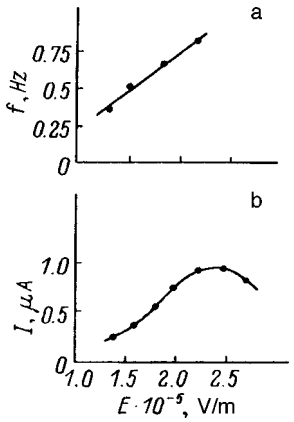

FIG. 3. Amplitude (a) and frequency (b) of current oscillations as a function of average electric field strength in the interelectrode gap.

were generated not at the cathode but in the interelectrode gap. Pulsations of the autosoliton propagation velocity accompanied by complex (multimodal) current oscillations were also observed.

The following hypothetical model may be put forward to explain these current oscillations in a planar diode structure. It is known that the electrical conductivity of nematic liquid crystals is of an ionic impurity type against a background of a negligible electron component. An ion oxidation process takes place at the anode with neutralization (reduction) at the cathode.[5] The process can take place reversibly without the liquid crystal undergoing decomposition. In a twisted nematic the field is concentrated in the central part of the gap between the plates.

Anisotropy of the electrical conductivity of the nematic liquid crystal is required for the formation of an autosoliton (for MBBA this is $\sigma_{\parallel}/\sigma_{\perp}=1.3$). Molecules oriented with the director perpendicular to the field impede the ion motion and form a kind of gate. When a large number of ions accumulate at the cathode, the ion pressure increases, causing the liquid crystal molecules to rotate parallel to the direction of action of the field and the ''gate'' opens, releasing the accumulated ions (the conductivity increases from $\sigma_{\perp}$ to $\sigma_{\parallel}$). The formation of autosoliton fronts is caused by the ion mobility gradient. After the accumulated ions have been released, the initial orientation of the liquid crystal molecules is restored.

[1] A. S. Sonin, *Introduction to the Physics of Liquid Crystals* [in Russian], Nauka, Moscow (1983).
[2] L. K. Vistin', A. Yu. Kabaenkov, and S. S. Yakovenko, Kristallografiya **31**, 360 (1986) [Sov. Phys. Crystallogr. **31**, 212 (1986)].
[3] G. Nicolis and I. Prigogine, *Self-Organization in Non-Equilibrium Systems* [Wiley, New York, 1977; Mir, Moscow, 1979].
[4] A. E. Borovik, V. Yu. Popkov, and V. N. Robuk, Dokl. Akad. Nauk SSSR **305**, 841 (1989) [Sov. Phys. Dokl. **34**, 339 (1989)].
[5] A. Adamchuk and Z. Strugal'skiĭ, *Liquid Crystals* [in Russian], Sov. Radio, Moscow (1979).

Translated by R. M. Durham

Limiting role of desorption in hydrogen transport across a deposited beryllium film

A. V. Samsonov, A. Yu. Koren'kov, I. E. Gabis, and A. A. Kurdyumov

St. Petersburg State University, Scientific-Research Institute of Physics, 198904 St. Petersburg, Russia
(Submitted May 19, 1997)
Zh. Tekh. Fiz. **68**, 128–130 (January 1998)

Hydrogen transport across a deposited beryllium layer has been investigated using the hydrogen permeability and concentration pulse methods. A layer of beryllium was deposited on a prepurified nickel membrane by cathode sputtering in a glow discharge plasma in ''especially pure'' grade hydrogen. An analysis of the experimental results showed that the main limiting process for hydrogen transport is desorption from the layer rather than diffusion in the bulk of the layer. A mathematical transport model is proposed and used to determine the rate constant of hydrogen desorption from beryllium. © *1998 American Institute of Physics.* [S1063-7842(98)02601-4]

The proposed use of beryllium as a protective coating in the ITER fusion reactor, and also plasma disruptions which lead to sputtering of the protective material, have stimulated increased interest in the interaction between hydrogen and beryllium. This problem is scientifically topical because beryllium is one of the *s* metals, whose interaction with hydrogen has clearly not been sufficiently well studied. However, the interaction between hydrogen and beryllium is difficult to study because of the nonremovable oxide film present on the surface and because of its toxicity and related technological difficulties. This accounts for the relative lack of studies, the wide scatter in the results, and the lack of any unified physical picture of the interaction.

Our previous studies of the thermal desorption of hydrogen from polycrystalline beryllium[1] allowed us to formulate a probable model of the transport process. According to this model, the hydrogen atom exhibits high diffusion mobility in the bulk of the grain and the limiting transport stage is the transfer from one grain to another, i.e., apparently, the escape of the diffusate from the bulk of the grain to its surface. Here we propose to study hydrogen transport across a sputtered layer of beryllium which, not being polycrystalline, can serve as a model for a single grain of a polycrystalline material, and we propose to check the accuracy of this model.

The investigations were carried out using an ultrahigh-vacuum automated experimental system[2] by the concentration pulse method, which was described in detail in Ref. 3. The substrate for this layer was made of NVK-grade vacuum-melted nickel. The parameters of its interaction with hydrogen are known in detail, and the high rates of the adsorption and desorption processes allow high-quality rectangular concentration pulses of dissolved hydrogen to be obtained in the subsurface region by switching on and off a hydrogen dissociator (incandescent tungsten filament) along the surface. The substrate was purified by high-temperature annealing in vacuum ($T=1023$ K) and by bombardment with hydrogen ions from a glow discharge plasma. A layer of beryllium was then deposited on the outer surface of the membrane by cathode sputtering using extra-pure hydrogen in a glow discharge plasma. Since no oxygen is present in the vacuum chamber, this procedure eliminates any oxide formation on the surface of the film. Hot-pressed TGP-grade beryllium was used for the deposition process. The composition of the components was 97.8% Be, 1.6% O, and 0.12% C. The deposition was carried out in runs of 4–6 h, and after each run the sample was annealed in vacuum at $T=673$ K until the hydrogen background pressure was established.

This nickel–beryllium system was investigated in the temperature range 523–723 K at intervals of 25 K. Higher temperatures were not used in order to avoid dissolution of the beryllium layer in the nickel. The experiments showed that an upper temperature of 723 K does not cause any drift of the results. At temperatures below 500 K, the nickel forming the base of the two-layer system shows some deviation from the classical limitation of a penetrating diffusion flux, which makes the transport model extremely complex.

After each deposition run, concentration-pulse experiments were carried out. Hydrogen was supplied to the uncoated (inner) side of the membrane and rectangular concentration pulses of hydrogen dissolved in the subsurface region were generated. At the end of the transport processes, a periodically varying penetrating flux was recorded on the outside of the membrane, this was Fourier expanded, and the phase–frequency characteristic of the membrane was determined.

An analysis of the phase–frequency characteristic of the system as a function of the deposition time (Fig. 1a) revealed the following. Deposition of beryllium for 14 h produces a continuous film, since subsequent deposition runs do not alter the phase–frequency characteristic. Diffusion across the deposited layer is not the main limiting factor for the transport process in the Ni–Be system since the phase–frequency characteristics do not depend on the deposition time, i.e., on the layer thickness.

The hydrogen permeability isotherms, i.e., the dependence of the steady-state penetrating flux on the root of the hydrogen pressure p, were measured before and after deposition of the beryllium film. Figure 1b gives the permeability isotherms at $T=673$ K for pure nickel and for a Ni–Be system after deposition for 56 h. It can be seen that deposition of the beryllium film substantially reduces the penetrating flux and changes the profile of the hydrogen permeability isotherm. The isotherms are not approximated by the dependence $J\sim\sqrt{p}$, typical of hydrogen transport limited only by

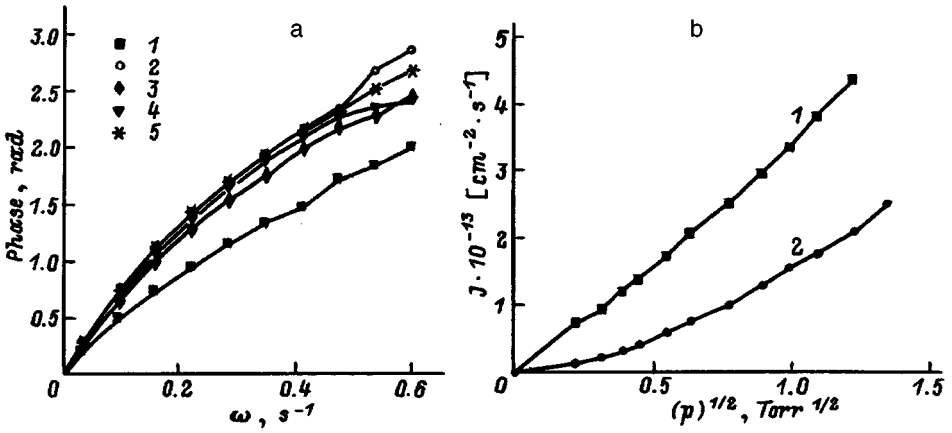

FIG. 1. a — Phase–frequency characteristics of system at $T=673$ K: *1* — pure nickel, *2–5* — deposition for 14, 30, 46, and 58 h, respectively; b — hydrogen permeability isotherms at $T=673$ K: *1* — pure nickel, *2* — after deposition for 56 h.

diffusion (as for a membrane without a film), which suggests[4] that the flux across the system is slowed substantially because of the low rate of hydrogen desorption from the beryllium film. This is also indicated by the appearance of even harmonics in the Fourier spectrum of the concentration pulse curves after deposition of the layer. These harmonics do not appear in the spectrum of pure nickel.

The results may be explained in terms of the following model of hydrogen transport across a sputtered layer of beryllium. The hydrogen atoms exhibit high diffusion mobility in the bulk of the layer and repeatedly cross the layer, being reflected from the exit boundary, since the process of escaping from the dissolved state into vacuum involves overcoming a high potential barrier. Thus the beryllium layer deposited on the substrate, while not presenting any diffusion resistance to the hydrogen, is responsible for its low rate of desorption.

Our experimental data can be used to check the accuracy of this model and to obtain the rate constant for desorption of hydrogen from beryllium b_{Be}. We consider a nickel membrane of thickness l in which the diffusion of hydrogen is described by Fick's law with the diffusion coefficient D

$$\frac{\partial C(x,t)}{\partial t}=D\frac{\partial^2 C(x,t)}{\partial x^2},\quad x\in[0,l]. \tag{1}$$

A first-order boundary condition is satisfied on the inside of the membrane

$$C(l,t)=f(t). \tag{2}$$

In the case of the concentration pulse method, the function f describes rectangular pulses with an inverse duty cycle of 2. The density of the steady-state penetrating flux is clearly given by

$$J=b_{\mathrm{Ni}}C^2(0)=b_{\mathrm{Be}}C_{\mathrm{Be}}^2, \tag{3}$$

where C_{Be} is the hydrogen concentration in the beryllium, which we assume to be independent of the coordinate, and b_{Ni} is the effective rate constant for the desorption of hydrogen from the nickel, which allows for the influence of the beryllium film.

The introduction of the constant b_{Ni} reduces the boundary-value problem to a simple form — diffusion over a homogeneous membrane with a low rate of desorption on the exit side.

Assuming, as always that the local equilibrium condition $C(0)/C_{\mathrm{Be}}=\Gamma_{\mathrm{Ni}}/\Gamma_{\mathrm{Be}}$ is satisfied at the interface between the

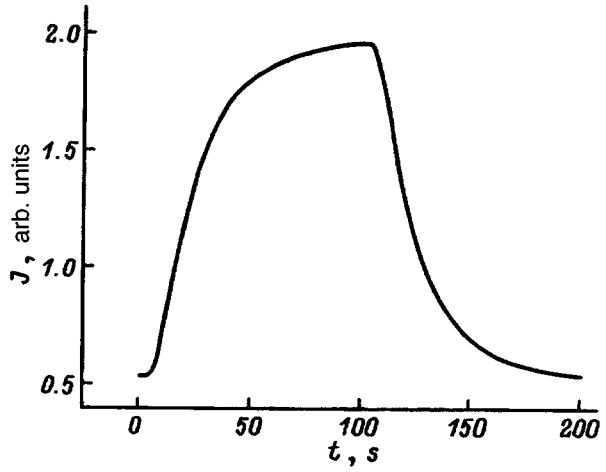

FIG. 2. Experimental and theoretical kinetic curves of penetrating flux at $T=573$ K.

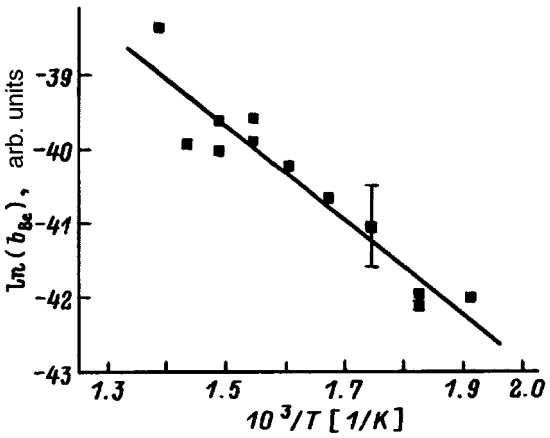

FIG. 3. Arrhenius curve of the rate constant for desorption of hydrogen from beryllium.

metals, where Γ_{Ni} and Γ_{Be} are the solubilities of hydrogen in nickel and beryllium, respectively, we can easily derive a relation to link the rate constants b_{Be} and b_{Ni}:

$$b_{Be}\Gamma_{Be}^2 = b_{Ni}\Gamma_{Ni}^2. \quad (4)$$

The solubilities of hydrogen in beryllium and nickel were taken from Refs. 5 and 4, respectively. The boundary condition on the outside is expressed as the balance of the diffusion and desorption fluxes,

$$b_{Ni}C^2(0,t) = D\frac{\partial C}{\partial x}(0,t). \quad (5)$$

This boundary-value problem (1), (2), (5) gives exceptionally good agreement with the experiment. As an example, Fig. 3 gives the experimental curve obtained at $T = 573$ K for a pulse period of 200 s, together with the superimposed model curve. The experimental curves were used to determine the temperature dependence of the rate constant b_{Ni}. Using Eq. (4), we then used this dependence to plot the Arrhenius curve of the rate constant for the desorption of hydrogen from beryllium plotted in Fig. 3, and we determined the numerical values of the pre-exponential factor and the activation energy $b_0 = 7.7 \pm 0.9 \times 10^{-14}$ [cm^4/s] and $E = 13 \pm 1$ [kcal/mol]. This activation energy is the potential barrier preventing the hydrogen dissolved in the beryllium from escaping to the surface, with a subsequent association to form a molecule and then desorption.

[1] I. E. Gabis, A. A. Kurdyumov, and A. V. Samsonov, *Noble and Rare Metals* [in Russian], Donetsk (1994), Part 3, pp. 89–90.
[2] I. E. Gabis, A. A. Kurdyumov, and N. A. Tikhonov, Vestn. St. Petersburg Gos. Univ. Ser. 4, No. 2, 11 (1993).
[3] I. E. Gabis and A. V. Ermakov, Fiz. Khim. Met. Metalloved., No. 4, 64 (1989).
[4] I. E. Gabis, T. N. Kompanets, and A. A. Kurdyumov, *Interaction Between Hydrogen and Metals*, edited by A. P. Zakharov [in Russian], Nauka, Moscow (1987), pp. 177–208.
[5] V. I. Shapovalov and Yu. M. Dukel'skiĭ, Dopov. Akad. Nauk. Ukr. RSR, Ser. A: Fiz.-Tekh. Mat. Nauki **6**, 84 (1988).

Translated by R. M. Durham

Thermal superconducting–normal transition of superconducting films, induced by the heating of nonsuperconducting defects in microwave fields

A. A. Zharov and A. N. Reznik

Institute of Physics of Microstructures, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia
(Submitted May 19, 1997)
Zh. Tekh. Fiz. **68**, 131–133 (January 1998)

A study is made of the thermal superconducting–normal (S–N) transition induced in high-temperature superconducting films by the Joule heating of small nonsuperconducting defects in a microwave field. It is shown that the loss of superconductivity either leads to the formation of a finite region of normal phase localized near the defect or encompasses the entire film. Estimates of the threshold and S–N switching time show good agreement with the experimental data. © *1998 American Institute of Physics.* [S1063-7842(98)02501-X]

1. Various recently published experimental studies have reported observations of the transition of high-temperature superconducting (HTSC) films from the superconducting (S) to the normal (N) state (S–N switching) stimulated by microwave radiation.[1–3] The observed S–N switching processes had a threshold amplitude of the incident wave field and the N state either occurred locally as small regions (domains) on the surface of the film[1] or encompassed the entire film (breakdown of the superconductor).[2,3] Loss of superconductivity may be caused by magnetic and thermal effects,[4] for which the S–N transition has different thresholds and exhibits different overall space–time behavior. An adequate interpretation of the experimental data cannot be given without studying each of these mechanisms.

Studies of the thermal mechanism for loss of superconductivity have established that the interphase boundary propagates over the surface of the superconductor in the form of a thermal autowave.[5–7] However, estimates of the threshold rf field strengths needed to excite such an autowave in perfect samples yielded a result between one and two orders of magnitude higher than that observed.[1)] The times of the transition processes were also between two and three orders of magnitude higher than those obtained from the measurements.[2,5] It will be shown below that one mechanism which can explain the comparatively low thresholds and short breakdown times for HTSC films is the heating of structural defects in the films.

2. We shall analyze a thin HTSC film deposited on a thermally stabilized dielectric substrate of thickness H, as shown in Fig. 1. Incident electromagnetic radiation having the energy flux density P_0 is partially absorbed in the film and is completely shielded from the substrate. The heat transfer processes in this system are described by the heat conduction equation in the substrate

$$c_s \frac{\partial T}{\partial t} = \varkappa_s \Delta T + q, \tag{1}$$

where c_s and $\varkappa_s$ are the specific heat and thermal conductivity of the substrate, and q is the heat source associated with heating of the film by the electromagnetic radiation.

Since the HTSC film is thin, we assume that the source q is localized in a δ layer on the surface of the substrate at $z=H$. Converting to the dimensionless temperature $\Theta=(T-T_0)/(T_c-T_0)$ and expressing q in explicit form, we have

$$c_s \frac{\partial \Theta}{\partial t} = \varkappa_s \Delta \Theta + \frac{\alpha(\Theta)}{T_c - T_0} P_0 \delta(z-H), \tag{2}$$

where T_c is the critical temperature of the S–N transition, T_0 is the coolant temperature, and $\alpha(\Theta)$ is the total absorption coefficient of a HTSC film with defects.

The absorption coefficient α_F of a film without defects depends on the temperature Θ and increases rapidly near $\Theta=1$. The dependence $\alpha_F(\Theta)$ in the wavelength range $\lambda>1$ cm may be approximately described by a step function.

Let us assume that the structure of the HTSC film contains a foreign inclusion of radius a (Fig. 1) (we shall subsequently call this a defect) whose absorption coefficient α_D does not depend on temperature and may differ substantially from that of the film in the S and N states. For $\alpha(\Theta)$, we finally obtain the following expression:

$$\alpha(\Theta) = \alpha_F(\Theta) + \alpha_D, \tag{3}$$

where

$$\alpha_F = \begin{cases} \alpha_n, & \Theta > 1, \\ 0, & \Theta \leqslant 1, \end{cases} \quad r > a,$$

$$\alpha_D = \begin{cases} \eta \alpha_n, & r \leqslant a, \\ 0, & r > a. \end{cases}$$

The parameter η characterizes the difference between the absorption coefficient of the defect and the absorption coefficient of the superconductor in the normal state.

Let us assume that in the absence of the electromagnetic field the HTSC film is in the S state with the temperature $\Theta=0$. When exposed to microwave radiation, only the defect absorbs energy [see Eq. (3)]. As a result of heat conduction, the superconductor is heated in the vicinity of the defect, and the film temperature may exceed the critical value $\Theta=1$. As a result, the film, being switched to the N state, itself begins to absorb electromagnetic waves and becomes

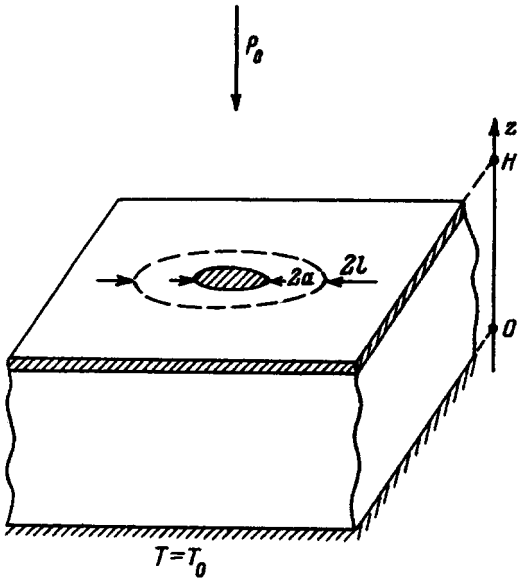

FIG. 1. Geometry of the problem. The shaded area on the surface of the film is a defect and the zone bounded by the dashed line is the region of destroyed superconductivity (thermal domain).

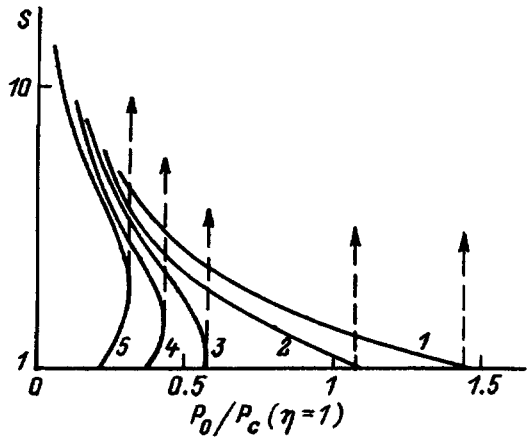

FIG. 2. Size of thermal domain as a function of the energy flux density of the incident radiation: η: *1* — 0.7, *2* — 1.0, *3* — 2, *4* — 3, and *5* — 5. The arrows indicate the critical values of $P_0/P_s(\eta=1)$ corresponding to thermal breakdown.

an additional source of heat. Ultimately, a steady state is established, achieved by a balance of heat release and heat removal when the N state occupies a finite region of radius l around the defect (Fig. 1). The evolution of thermal breakdown arbitrarily corresponds to the absence of any steady states with finite l in the solution of Eq. (2). Thus, the solution of the stated problem reduces to determining the steady-state boundary l of the region of destroyed superconductivity. In the assumed approximations the steady-state solution of the nonlinear Eq. (2) with the boundary conditions $\Theta=0$ at $z=0$ and $d\Theta/dz=\alpha P_0/\varkappa_s(T_c-T_0)$ at $z=H$ may be determined accurately by means of a Fourier transformation with respect to the coordinates along the surface of the film. Omitting the cumbersome, but fairly obvious calculations, we directly give the final result for the steady-state temperature distribution over the film:

$$\Theta(H,r)=\frac{\alpha_n P_0}{(T_c-T_0)\varkappa_s}\left\{a^2(\eta-1)\times\int_0^\infty J_0(kr)J_1(ka)\tanh(kH)\frac{dk}{ka}+l^2\int_0^\infty J_0(kr)J_1(kl)\tanh(kH)\frac{dk}{kl}\right\}, \tag{4}$$

where $J_{0,1}(x)$ is a Bessel function and the radius l of the thermal domain (region of destroyed superconductivity) is obtained from the condition $\Theta(H,r=l)=1$.

If the size of the defect a and the domain l are small compared with H, the integrals in Eq. (4) may be taken in the explicit form:

$$\Theta(H,r)=\frac{\alpha_n P_0}{(T_c-T_0)\varkappa_s}\left\{\frac{a^2(\eta-1)}{2r}F\left(\frac{1}{2},\frac{1}{2},2,\frac{a^2}{r^2}\right)+lF\left(\frac{1}{2},-\frac{1}{2},1,\frac{r^2}{l^2}\right)\right\}, \tag{5}$$

where $F(\alpha,\beta,\gamma,\delta)$ is a hypergeometric function.

Expression (5) corresponds to the temperature distribution inside the domain (a similar expression can obviously be obtained for $r>l$). We find the dependence of the domain radius l on the radiation energy flux density P_0, for which we assume in Eq. (5) $r=l$ and $\Theta(H,r=l)=1$. We then have

$$P_0=P_c\frac{\eta s}{\pi/4(\eta-1)F\left(\frac{1}{2},\frac{1}{2},2,\frac{1}{s^2}\right)+s^2}, \tag{6}$$

where $s=l/a$; $P_c=\pi\varkappa_s(T_c-T_0)/2\eta\alpha_n a$ is the threshold for nucleation of a domain corresponding to $s=1$.

Figure 2 gives curves of $s(P_0)$ for various values of the parameter η. It can be seen that the typical value of η is $\eta_0\approx 2$. For $\eta<\eta_0$ the condition $P_0>P_c(\eta)$ immediately leads to thermal breakdown of the superconductor. However, if $\eta>\eta_0$, a steady-state normal-phase domain is formed initially, and breakdown sets in only at values of P_0 slightly in excess of P_c. In this last case, an increase in P_0 for $P_0\geqslant P_c(\eta)$ corresponds to an increase in the average surface resistance of the film, as was observed in Refs. 2 and 3 (it is natural to assume that superconductors may have many defects of similar size and properties at which domains appear independently).

We estimate the magnetic field strength in the incident wave which corresponds to the threshold energy flux density of the wave $P_0=(c/8\pi)B_m^2=P_c$ (c is the speed of light). Taking $\varkappa_s\approx 3\times 10^6$ (erg/s·cm·K), $a\approx 10^{-4}$ cm, $T_c-T_0=10$ K, $\eta=1$, and $\alpha_n\approx 10^{-2}$, we obtain $B_m\approx 200$ Gs, which agrees with the experimental data.[2–4] Estimates of the time taken to heat the regions of film near defects to $T\sim T_c$ for $P_0\leqslant P_c$ also yield values similar to those observed in Refs. 2 and 3:

$$\tau\approx\frac{c_s l^2}{2\varkappa_s}f(P_0)\sim 10^{-7}s,$$

where $c_s \approx 10^7$ erg/(K·cm^2), $f(P_0)$ is a form factor of ~ 1 which depends logarithmically on P_0.

3. To conclude, if allowance is made for the defects in HTSC films, the thresholds for thermal S–N switching stimulated by microwave fields and the characteristic times of the transition processes are similar to those observed experimentally. Depending on the flux density of the incident radiation and the power released at the defect as a result of absorption, the S–N transition may either lead to thermal breakdown, destroying the superconductivity over the entire surface of the film, or it may result in the formation of thermal domains.

This work was supported by the Russian Fund for Fundamental Research (Grants Nos. 95-02-04996 and 96-02-16997).

[1)] It was shown in Ref. 7 that the threshold may be lowered by local heating of the superconductor by an external source.

[1] G. Hampel, P. Kolodner, P. L. Gammel *et al.*, Appl. Phys. Lett. **69**, 571 (1996).
[2] W. Diete, M. Getta, M. Hein *et al.*, *Proceedings of the Applied Superconductivity Conference ASC′96*, Pittsburg, 1996.
[3] J. Wosik, L. M. Xie, J. H. Miller *et al.*, *Proceedings of the Applied Superconductivity Conference ASC′96*, Pittsburg, 1996.
[4] M. Hein, W. Diete, M. Getta *et al.*, *Proceedings of the Applied Superconductivity Conference ASC′96*, Pittsburg, 1996.
[5] A. N. Reznik and A. A. Zharov, *Proceedings of the Ninth Trilateral German–Russian–Ukrainian Seminar on High-Temperature Superconductors*, Gabelbach, Germany, 1996.
[6] N. A. Buznikov and A. A. Pukhov, Pis'ma Zh. Tekh. Fiz. **22**(22), 45 (1996) [Tech. Phys. Lett. **22**, 499 (1996)].
[7] A. A. Pukhov, Pis'ma Zh. Tekh. Fiz. **22**(12), 55 (1996) [Tech. Phys. Lett. **22**, 504 (1996)].

Translated by R. M. Durham

Three-body recombination of electrons and ions in the presence of two-level atoms

A. N. Tkachev and S. I. Yakovlenko

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia
(Submitted August 20, 1996)
Zh. Tekh. Fiz. **68**, 15–19 (January 1998)

The distribution function of bound electrons and the recombination rate of electrons and ions in the presence of two-level atoms is considered within a diffusion model. Two cases are considered: (a) it is assumed in accordance with traditional theories that relaxation occurs as a result of binary collisions; (b) it is assumed that the anomalous drift previously discovered on the basis of a first-principles simulation takes place. It is shown that the distribution of bound electrons obtained on the basis of the theory of binary Coulomb collisions is not consistent with the results of a numerical many-particle dynamics simulation, while a kinetic model which utilizes the theory of anomalous drift is consistent with the simulation results. © *1998 American Institute of Physics.* [S1063-7842(98)00301-8]

INTRODUCTION

A many-particle dynamics (MPD) simulation for Coulomb particles bears fruitful results, if it is accompanied by an analysis of corresponding kinetic models.[1–5] In Refs. 6 and 7 the results of an MPD simulation of the relaxation of a system of Coulomb particles in a heat bath of two-level atoms were compared with kinetic models that contradict the principle of detailed balance, and good agreement was obtained. In view of the different opinions expressed in Refs. 8–12 in regard to the interpretation of the results of the MPD simulation in Refs. 1–3, it would be useful to check once again to what extent the simulation results contradict the traditional models of recombination based on the theory of binary Coulomb collisions and the principle of detailed balance.

In this context, below we examine the form of the distribution function following from a diffusion model with kinetic coefficients which are obtained on the basis of traditional models of binary Coulomb collisions and obey the relations following from the principle of detailed balance. This function is compared with the results of an MPD simulation and the function obtained with kinetic coefficients that do not correspond to the principle of detailed balance.

FORMULATION OF THE PROBLEM IN THE BINARY-COLLISION MODEL

The Fokker–Planck equation. For the distribution function $f(\varepsilon)$ of electrons with respect to their total energy ε we use the Fokker–Planck equation

$$\partial f/\partial t = -\partial\Gamma/\partial\varepsilon,$$

$$\Gamma = Af - \partial(Bf)/\partial\varepsilon \equiv \widetilde{A}f - B\partial f/\partial\varepsilon.$$

Here Γ is the electron flux along the energy axis (in the case of recombination $\Gamma<0$); A and B are the kinetic coefficients for mobility and diffusion along the energy axis, respectively; and $\widetilde{A}=A-\partial B/\partial\varepsilon$ is the modified mobility coefficient. The use of the diffusion approximation presumes that the motion of electrons along the energy axis is accomplished in small steps and that the kinetic coefficients vary weakly in these energy intervals.

As usual, below we shall use the steady-state approximation $\partial f/\partial t=0$, $\Gamma=\text{const}$, whence it follows that

$$\Gamma = \widetilde{A}(\varepsilon)f(\varepsilon) - B(\varepsilon)df(\varepsilon)d\varepsilon = \text{const}. \tag{1}$$

The boundary conditions for this equation depend on which process (recombination or ionization) is being described; they will be formulated below. For now, we only note that the determination of the relaxation flux (i.e., the number of electrons which recombine or detach from ions per unit time) on the basis of the steady-state Fokker–Planck equation (1) does not require precise specification of the boundary between the bound and free electrons. This is important, because, as a rule, highly excited electrons cannot be unequivocally classified as bound or free particles, and the position of the boundary between them can be indicated only approximately.

Principle of detailed balance. At thermodynamic equilibrium the distribution function of plasma electrons with respect to the total energy is of the Boltzmann type:

$$f_B(y) = g(y)\exp(-y),$$

$$g(y) = \frac{C}{T_t}\begin{cases} \dfrac{2}{\pi^{1/2}}y^{1/2} & \text{for } y \gg \delta^{1/3}, \\ \dfrac{\pi^{3/2}}{4}\delta|y|^{-5/2} & \text{for } |y| \gg \delta^{1/3}, \quad y<0, \end{cases}$$

where $y=\varepsilon/T_e$ is the reduced energy and $\delta=2e^6N_i/T_e^3$ is the nonideality parameter of the plasma.

According to the principle of detailed balance, substitution of the Boltzmann distribution into the diffusion equation should cause the expression for the flux to vanish identically. This imposes the following relation on the expressions for the diffusion and mobility coefficients:

$$\widetilde{A}(\varepsilon)f_B(\varepsilon/T_e) - B(\varepsilon)df_B(\varepsilon/T_e)/d\varepsilon = 0. \tag{2}$$

As we know, in the case of binary collisions, the principle of detailed balance follows from the temporal revers-

ibility of a collision between particles. In the general case the principle of detailed balance is a consequence of the assumption that the system under consideration is ergodic.

Kinetic coefficients. The expression used below for the diffusion coefficient of a bound electron in Coulomb collisions with plasma electrons was derived in Refs. 13 and 14. It can be represented in the form

$$B_c=\frac{8\sqrt{\pi}}{3}e^4N_e\Lambda\sqrt{\frac{2T_e}{m_e}}x^{1/2}\mu(x),$$

$$\mu(x)=\frac{x^{1/2}}{\sqrt{1+6x+0.75x^2+\pi x^3/16}}.$$

Here $x=-y$ is the binding energy of an electron normalized to the electron temperature, and $\Lambda=\sqrt{1+9/4\pi\delta}$ is the Coulomb logarithm. The expression for the diffusion coefficient of a bound electron under the action of inelastic collisions with two-level atoms was obtained in Refs. 5 and 6 (see also Ref. 2):

$$B_a=\frac{4}{3\pi}\Delta\varepsilon^2\sigma_0N_a\sqrt{\frac{2T_e}{m_e}}\sqrt{x}.$$

Here $\Delta\varepsilon>0$ is the energy difference between the atomic levels, N_a is the atom density, and σ_0 is the cross section for the excitation of an atom by electron impact near the threshold. For simplicity, it is assumed within the model from Refs. 5 and 6 that the cross section does not vary when the threshold is exceeded. We note that this leads to some overestimation of the relaxation rate for tightly bound electrons. It is assumed, in addition, that the populations of the ground and excited states obey a Boltzmann distribution with an atom temperature T_a; this simulates the action of a heat bath of two-level atoms.

Justification of the use of the diffusion approximation requires fulfillment of the condition that the transition energy be small compared with the characteristic scale for variation of the distribution function. In particular, it requires that $\Delta\varepsilon\ll T_e, T_a$.

Stationary drainage equation and boundary conditions. We assume that the diffusion of an electron along the energy axis as a result of collisions with atoms and with electrons takes place independently. Then the kinetic coefficients in Eq. (1) are the sums of the corresponding quantities, and the equation for the distribution function takes the form

$$\Gamma=(\widetilde{A}_a+A_c)f(\varepsilon)-(B_a+B_c)df(\varepsilon)d\varepsilon=\text{const}.$$

Introducing a function which characterizes the deviation of the recombination distribution from the Boltzmann distribution

$$\Phi(\varepsilon/T_e)=f(\varepsilon)/f_B(\varepsilon/T_e),$$

and using the principle of detailed balance for $e-e$ and $e-a$ collisions

$$\widetilde{A}_c(\varepsilon)f_B(\varepsilon/T_e)-B_c(\varepsilon)df_B(\varepsilon/T_e)/d\varepsilon=0,$$

$$\widetilde{A}_a(\varepsilon)f_B(\varepsilon/T_a)-B_a(\varepsilon)df_B(\varepsilon/T_a)/d\varepsilon=0,$$

we have

$$\frac{d\Phi(x)}{dx}+\Phi(x)\frac{1-\theta}{1+B_c/B_a}+\frac{-\Gamma T_e}{f_B(x)B_a(1+B_c/B_a)}=0,$$

where $x=-\varepsilon/T_e$ and $\theta=T_e/T_a$.

We write the boundary conditions in the form $\lim_{x\to\infty}\Phi(x)=0$, $\lim_{x\to 0}\Phi(x)=1$. They reflect the fact that the distribution should go over to the Boltzmann distribution at small electron binding energies and that it should be much smaller than the Boltzmann distribution at large binding energies. These boundary conditions correspond to the steady-state drainage of electrons from the continuum to tightly bound states. The time for establishment of the distribution function corresponding to stationary drainage is of the order of the time between Coulomb collisions (for further details, see Refs. 1–3, and 14).

DISTRIBUTION OF ELECTRONS ALONG THE ENERGY AXIS AND RECOMBINATION RATE IN THE BINARY-COLLISION MODEL

Starting equation. Using concrete expressions for the diffusion coefficients, we can bring Eq. (1) into the form

$$\frac{d\Phi(x)}{dx}+\Phi(x)\frac{1-\theta}{1+\mu(x)/c_1}+\frac{\text{const}\cdot x^2e^x}{1+\mu(x)/c_1}=0,$$

where the parameter

$$c_1=\frac{1}{2\pi^{3/2}}\frac{\Delta\varepsilon^2\sigma_0}{e^4\Lambda}\frac{N_a}{N_e}$$

characterizes the ratio between the rates of Coulomb collisions and inelastic collisions of electrons with atoms. Since this equation is linear, it is simple to write down its solution in quadratures. However, the corresponding integral expressions are not convenient for obtaining concrete results. It is simpler to analyze limiting cases and to perform direct numerical integration of the differential equation in the intermediate region.

Limiting cases. When $c_1\to 0$, neglecting the inelastic collisions with atoms, we have the familiar expressions for the distribution function and the recombination flux[13,14]

$$f_c(x)=f_B(x)\times\left[\int_x^\infty dzz^2\,\exp(-z)/\mu(z)\right]\Bigg/\left[\int_0^\infty dzz^2\,\exp(-z)/\mu(z)\right],$$

$$|\Gamma_c|=\frac{4}{5.004}\frac{2^{5/2}\pi^{3/2}}{9}\frac{e^{10}\Lambda}{\sqrt{m_e}}\frac{N_e^2}{T_e^{9/2}}.$$

The distribution function is approximated to within an accuracy no poorer than 5% by the expression

$$f_c(x)=\frac{\pi^{3/2}}{4}\delta\frac{1+x+0.476x^2+0.0657x^3}{x^{5/2}},$$

which gives correct asymptotic behavior. When $c_1\to\infty$, neglecting the Coulomb collisions, we can obtain (compare Refs. 6 and 7)

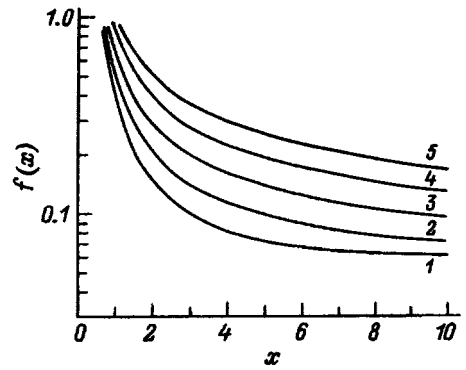

FIG. 1. Distribution functions of particles with respect to the total energy for the model with fulfillment of the principle of detailed balance: *1*,*5*—limiting cases of the absence and predominance of inelastic collisions; *2*–*4*—calculation for $c_1=0.1$, 0.3, and 1, respectively.

$$f_a(x,\theta)=\frac{\pi^{3/2}}{4}\,\delta\frac{2+2\theta x+\theta^2x^2}{2x^{5/2}},$$

$$|\Gamma_a|=\frac{2^{1/2}\pi^{1/2}}{3}\,\frac{e^6\sigma_0\Delta\varepsilon^2N_eN_a}{\sqrt{m_e}T_a^3T_e^{3/2}}.$$

Calculation results. As the results of the calculations show (Fig. 1), in cases in which neither the Coulomb nor the inelastic collisions can be neglected, the distribution function lies between the limiting expressions presented above. The resultant recombination flux can be represented in the form

$$\Gamma=\Gamma_c[1+c_1\theta^3\xi(c_1,\theta)], \tag{3}$$

where the function $\xi(c_1,\theta)$ describes the transition region for the recombination rate (Fig. 2).

DISTRIBUTION OF ELECTRONS ALONG THE ENERGY AXIS AND RECOMBINATION RATE IN THE MODEL OF ANOMALOUS DRIFT

Microfield distribution. It was conjectured in Ref. 1 (see also Refs. 2 and 3) on the basis of the results of numerical simulation that a quasisteady-state distribution of electrons (for $\Gamma=0$) differing from the equilibrium Boltzmann distribution can arise. It forms because the mobility and diffusion coefficients of an electron along the energy axis do not obey

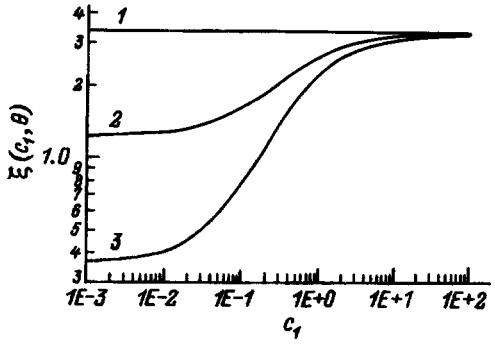

FIG. 2. Dependence of the fraction of the recombination flux due to inelastic collisions on the coefficient characterizing the efficiency of the inelastic collisions. $\theta=1$ (*1*), 2.5 (*2*), and 5 (*3*).

Eq. (2), which follows from the principle of detailed balance. In this case there is anomalous electron drift along the energy axis: it is anomalously directed (from the region of electrons with negative energies into the region of positive energies) and anomalously strong (it exceeds the Coulomb drift). On the basis of approximate arguments we obtained an expression for the ''microfield'' distribution function

$$f_f(y)=\frac{2C}{\sqrt{\pi}}\begin{cases}\sqrt{y}\exp(-y), & y>\alpha\delta^{1/3},\\ C_3\,\exp(C_1y+C_2y^2/2), & |y|\leqslant\alpha\delta^{1/3},\\ C_4\,\exp(\beta y/\delta^{1/3}), & y<-\alpha\delta^{1/3},\end{cases} \tag{4}$$

which takes into account the anomalous drift along the energy axis. Here

$$C_1=[-1+1/(2\alpha\delta^{1/3})+\beta/\delta^{1/3}]/2,$$

$$C_2=[-1+1/(2\alpha\delta^{1/3})-\beta/\delta^{1/3}]/(2\alpha\delta^{1/3}),$$

$$C_3=\alpha^{1/2}\delta^{1/6}\,\exp[-\alpha\beta^{1/3}(1+C_1+C_2\alpha\delta^{1/3}/2)],$$

$$C_4=\alpha^{1/2}\delta^{1/6}\,\exp[\alpha\beta-\alpha\beta^{1/3}(1+2C_1)],$$

$$C^{-1}=1-(2/\sqrt{\pi})\,\gamma(3/2,\alpha\delta^{1/3})+(2C_3/\sqrt{\pi})\times\int_{-\alpha\delta^{1/3}}^{\alpha\delta^{1/3}}\exp(C_1y+C_2y^2/2)\,dy+(2C_4\delta^{1/3}/\beta\sqrt{\pi})\exp(-\alpha\beta),$$

where $\alpha=1.5$ and $\beta=0.4$ are coefficients, whose numerical values are selected such that the microfield distribution function would describe the results of the numerical calculations most faithfully.

Starting equation. On the basis of the expressions for the kinetic coefficients obtained in Refs. 6 and 7 (see also Ref. 3), the equation for the diffusion of an electron along the energy axis under the action of plasma microfields and inelastic collisions with two-level atoms can be written in the following form

$$\frac{df(x)}{dx}+f(x)\frac{a(x)}{b(x)}+\frac{\text{const}}{b(x)}=0,$$

where

$$a(x)=\frac{\beta}{\delta^{1/3}}+c_1'\left(\frac{5}{2\sqrt{x}}-\sqrt{x}\,\theta\right),\quad b(x)=1+c_1'\sqrt{x}$$

are the dimensionless mobility and diffusion coefficients;

$$c_1'=\frac{4}{0.75\cdot3\pi}\,\frac{\Delta\varepsilon^2\sigma_0}{e^4}\,\frac{N_a}{N_e}$$

is a parameter which characterizes the efficiency of the inelastic collisions; and the solution should transform into the microfield distribution when $c_1'\to0$.

The boundary conditions are determined from the condition that the distribution function sought be small compared with the Boltzmann distribution at large binding energies, as well as from the condition for matching with the microfield distribution at small binding energies:

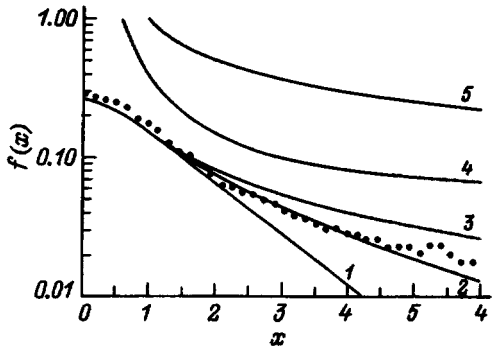

FIG. 3. Distribution functions of particles with respect to the total energy for the model with anomalous drift for a plasma with the parameters $\delta=0.11$ and $\theta=2.5$: *1*—case of the absence of inelastic collisions; *2,3* calculation for the collision rates $c_1'=0.1$ and 0.2; *4,5*—limiting distributions for the model with fulfillment of the principle of detailed balance. Points—results of the MPD simulation in Ref. 3 (see also Ref. 6) for a plasma with the parameters $\delta=0.11$, $\theta=2.5$, and $c_1'\approx 0.1$.

$$\lim_{x\to\infty} f(x)/f_B(x)=0, \qquad \lim_{x\to\alpha\delta^{1/3}} f(x)=f_f(\alpha\delta^{1/3}).$$

Calculation results. Analytical expressions for the cases of predominance of the Coulomb interaction and inelastic collisions were presented in Refs. 3,6, and 7. Here we shall consider only some results of the numerical calculations. As the calculation results show (Fig. 3), in the case of the microfield distribution function a ''tail'' appears at negative energies due to the inelastic collisions. However, the form of the distribution of the electrons in the tail differs significantly from the form which would be observed when the principle of detailed balance holds. When anomalous drift takes place, the drop in the distribution function with increasing binding energies is, of course, significantly steeper. In this model the recombination flux is a function of three parameters and can be represented in the form

$$|\Gamma_f|=N_e(2T_e/m_e)^{1/2}(e^2/T_e)^2\eta(c_1',\theta,\delta).$$

As the efficiency of the inelastic collisions increases, the recombination flux increases monotonically (Fig. 4).

A comparison with the data from the MPD simulation[6,7] shows that the traditional model based on theories of binary Coulomb collisions and the principle of detailed balance cannot account for the form of the distribution obtained in the calculations (Fig. 3).

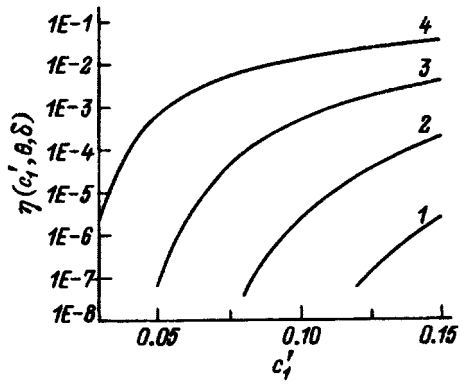

FIG. 4. Dependence of the recombination rate (for the model with anomalous drift) on c_1' for $\delta=0.11$ and $\theta=1$ (*1*), 1.5 (*2*), 2.5 (*3*), 5 (*4*).

CONCLUSIONS

The results of the present work are consistent with the concept reviewed in Refs. 1–3. The essential points of this concept are as follows. The translational and positional degrees of freedom in a system of Coulomb particles mix for an anomalously long time in comparison with the situation which would be observed if relaxation of the bound states were described by binary Coulomb collisions. On the basis of the known entropy conservation law for Hamiltonian systems, it would be natural to conclude that relaxation to a state of statistical equilibrium in a dynamic system takes place only in the presence of external (with respect to the dynamic equations) stochastic disturbances. Translational degrees of freedom are unstable with respect to external disturbances; therefore, even small errors in the numerical calculation lead to the establishment of a Maxwellian distribution. At the same time, the establishment of equilibrium between free and bound states for classical Coulomb particles does not correspond to the theories of binary Coulomb collisions[1)] and the applicability of the principle of detailed balance in the transitional formulation in the present case.

The simulation of both a system of classical Coulomb particles protected from stochastic disturbances and a system that is subjected to external stochastic disturbances shows that there is anomalous electron drift along the energy axis. It is possibly caused by collective plasma oscillations.[5] The occurrence of anomalous drift leads to the formation of a metastable state of a classical Coulomb plasma.

[1)]As was shown in Refs. 5 and 6 (see also Refs. 2 and 3), the discrete nature of the bound states promotes predominance of the role of binary collisions for light Coulomb particles.

[1] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **34**, 3 (1991).
[2] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Usp. Fiz. Nauk **164**, 298 (1994) [Phys. Usp. **37**, 279 (1994)].
[3] S. A. Mayorov, A. N. Tkachev, and S. I. Yakovlenko, Phys. Scr. **51**, 498 (1995).
[4] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Kratk. Soobshch. Fiz. **9–10**, 28 (1995).
[5] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Kratk. Soobshch. Fiz. **9–10**, 35 (1995).
[6] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Mat. Model., No. 7, 3 (1992).
[7] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **36**, 68 (1993).
[8] A. I. Gorskiĭ, V. A. Dzenzerskiĭ, and Yu. P. Kuchugurnyĭ, Pis'ma Zh. Tekh. Fiz. **20**(16), 25 (1994) [Tech. Phys. Lett. **20**, 655 (1994)].
[9] A. M. Ignatov, A. I. Korotchenko, V. P. Makarov, A. A. Rukhadze, and A. A. Samokhin, Usp. Fiz. Nauk **165**, 113 (1995) [Phys. Usp. **38**, 109 (1995)].
[10] S. A. Maĭorov, A. N. Tkachev, and S. I. Yakovlenko, Usp. Fiz. Nauk **165**, 117 (1995) [Phys. Usp. **38**, 113 (1995)].
[11] A. A. Rukhadze, Kratk. Soobshch. Fiz. **9–10**, 40 (1995).
[12] A. N. Tkachev and S. I. Yakovlenko, Kratk. Soobshch. Fiz. (1996).
[13] A. N. Tkachev and S. I. Yakovlenko, Kratk. Soobshch. Fiz. **7**, 10 (1990).
[14] A. N. Tkachev and S. I. Yakovlenko, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **37**, 3 (1994).

Translated by P. Shelnitz

Sputtering of gold by krypton ions in the inelastic region of energy loss

G. I. Akap'ev, A. N. Balabaev, N. A. Vasil'ev, S. V. Latyshev, V. M. Nazarov,[a] A. R. Piuto, I. V. Rudskoĭ, and Yu. N. Cheblukov

Institute of Theoretical and Experimental Physics, 11725 Moscow, Russia
(Submitted November 11, 1996)
Zh. Tekh. Fiz. **68**, 134–135 (January 1998)

The sputtering coefficient of gold was determined experimentally as a function of the incident krypton ion energy in the inelastic region of energy loss. It is shown that this dependence does not differ from that predicted by the cascade theory. The work was carried out using the U-400 cyclotron beam at the Joint Institute for Nuclear Research, Dubna. © *1998 American Institute of Physics.* [S1063-7842(98)02701-9]

It was shown for the first time in Ref. 1 that the coefficient of sputtering of metals by multiply charged ions, for which the inelastic energy loss $(dE/dx)_e$ is considerably greater than the elastic loss in atomic collisions, differs from that predicted by the cascade theory. The sputtering coefficient S of coarse-grained gold by 5.5 MeV/nucleon uranium ions was found to be (12 ± 2) atoms/ion. An estimate of S for this uranium ion energy, allowing only for elastic interactions, gives $S\sim 1$. This experimental result provided the stimulus for a more thorough investigation of the sputtering of metals in the inelastic range of energy losses using heavy ion beams in the U-400 cyclotron at JINR, Dubna.

Here we report the experimentally determined dependence of the sputtering coefficient S of gold on the energy E of the incident krypton ions (see Fig. 1). This dependence was obtained by collecting the sputtered gold using a special target consisting of many alternating layers of gold, carbon, and aluminum foils. The carbon foils collected the sputtered gold atoms and the aluminum foils were used to vary the krypton ion energy. The gold foils, 0.2 μm thick, were annealed beforehand in a vacuum furnace at 400 °C to eliminate any intergranular structure. This design allowed us to determine the sputtering coefficient of gold at various krypton ion energies during a single exposure of the target in a monoenergetic beam. The quantity of sputtered gold at the carbon collectors was determined by activation analysis. A Faraday cup was positioned behind the target to measure the quantity of ions transmitted by the multilayer target. During the experiment the target was heated to $\sim$200 °C to clean the surface of the gold foils.

According to a model proposed earlier to describe the sputtering of a metal by fast multiply charged ions,[2] a hot electron gas with a temperature of 20–25 eV, which exceeds the ionization potentials of the lattice ions, forms near the ion trajectory. At the metal–vacuum interface, as a result of the high pressure inside the hot electron gas the spatial distribution of the electrons differs from the ion density profile, which leads to the formation of an electric double layer. A surface metal ion which has zero binding energy above a certain critical temperature of the electron gas[2] is accelerated by the double layer without being confined by the restoring force. If a surface ion located in a hot spot has time to acquire an energy higher than the binding energy during cooling of the electron gas to the critical temperature, the ion leaves the metal and will contribute to the sputtering. The sputtering process therefore has a threshold nature. Using the results of Ref. 2, it can be shown that for $\sim$100 MeV krypton ions, the threshold value of the inelastic loss in gold is $(dE/dx)_t\approx 2.1$ keV/Å. The energy loss of $\sim$100 MeV krypton ions in gold $(dE/dx)_e\approx 2.7$ keV/Å (Ref. 3) exceeds the threshold value. It can be seen from Fig. 1 that the experimental dependence shows no deviation from the cascade theory. In the particular energy range studied the krypton ions (see Fig. 1) have similar parameters to fission fragments so that the sputtering coefficient of gold by fragments cannot be influenced by the inelastic loss, as was suggested by the authors of Refs. 2 and 4 when comparing the theoretical sputtering coefficient with the experimentally measured value.

In our view, the energy dependence of the sputtering coefficient of gold by 100–500 MeV uranium ions, where the energy loss $(dE/dx)_e$ increases abruptly and is quite appreciable, must be determined experimentally to establish definitively how the sputtering process of metals is influenced by the inelastic energy loss. This experiment is scheduled using the U-400 cyclotron beam at JINR.

The authors would like to thank Yu. Ts. Oganesyan and V. A. Shchegolev for support of the experiments, and Yu. N. Yavlinskiĭ for discussions of the results.

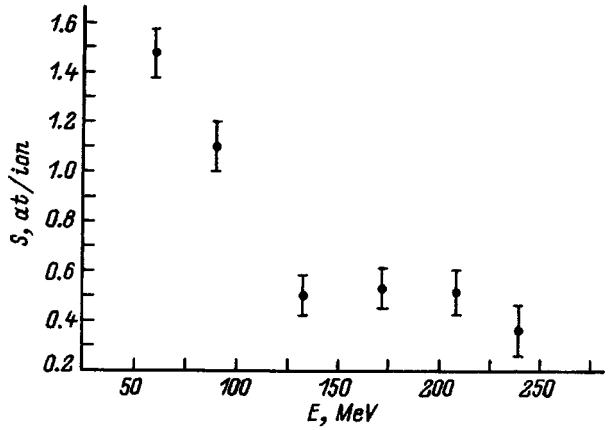

FIG. 1. Sputtering coefficient of gold versus incident krypton ion energy.

 1063-7842/98/010120-02$15.00

[a)]Deceased.

[1] Yu. N. Cheblukov, D. G. Koshkarev *et al.*, Part. Accel. **37–38**, 351 (1992).
[2] Yu. V. Martynenko and Yu. N. Yavlinskiĭ, Zh. Tekh. Fiz. **58**, 1164 (1988) [Sov. Phys. Tech. Phys. **33**, 683 (1988)].
[3] *Nuclear Data Tables*, Vol. 7, No. 3–4 (1970).
[4] I. A. Baranov, Yu. V. Martynenko, S. O. Tsepelevich, and Yu. N. Yavlinskiĭ, Usp. Fiz. Nauk. **156**, 477 (1988) [Sov. Phys. Usp. **31**, 1015 (1988)].

Translated by R. M. Durham

Magnetic properties of an amorphous microconductor in the microwave range

S. A. Baranov
(Submitted December 20, 1996)
Zh. Tekh. Fiz. **68**, 136–137 (January 1998)

The frequency dispersion of the magnetic permeability is taken into account in a calculation of the impedance of an amorphous microconductor. Possible methods of measuring the magnetic parameters are analyzed, and the main characteristics of an amorphous microconductor in the microwave range are described. © *1998 American Institute of Physics.*
[S1063-7842(98)02801-3]

Materials possessing high values of the components of the magnetic permeability μ are needed to develop absorbing screens in the microwave range (see Ref. 1, for example). As far as we are aware, a cast amorphous microconductor is a unique material because, as a result of the natural ferromagnetic resonance in the frequency range up to 10 GHz (Refs. 2–4), composites based on this material possess a radio absorbing property.[5]

Note that because μ has a resonant character[2–8] in this frequency range, model calculations of radio absorbing media which neglect the frequency dispersion cannot really be applied to composites based on a cast amorphous microconductor. The present study is confined to questions of measurement of the magnetic characteristics of an amorphous microconductor.

An important difference when allowance is made for the dispersion is that an expansion is obtained for the electric and magnetic fields in a long cylindrical section of the microconductor:[9]

$$\begin{matrix} E \\ H \end{matrix} \rightarrow \sum_n \begin{matrix} A_n \\ B_n \end{matrix} f_{n,k}(r,\phi),$$

$$f_{n,k}(r,\phi)=e^{\pm in\phi}J_n(kr), \tag{1}$$

where $J_n(x)$ is a Bessel function, r,ϕ are cylindrical coordinates, and k is the wave vector, which tends to the following value[9] in the case under study, i.e., for the mode of uniform precession of the natural ferromagnetic resonance:

$$k \rightarrow (1+i)/\delta, \tag{2}$$

where δ is the skin layer depth.

Unlike the dispersion-free case when only one mode exists and the wave impedance is expressed by[7]

$$Z_1 = \text{const}\ k\frac{J_0(kr_c)}{J_1(kr_c)}, \tag{3}$$

where r_c is the radius of the central core of the microconductor, in our case we have

$$Z=\sum_n Z_n, \tag{4}$$

where

$$Z_n \sim \text{const}\ k\frac{J_{n-1}(kr_c)}{J_n(kr_c)}. \tag{5}$$

The magnetic permeability is not an additive quantity, but for each mode we can introduce μ_n which, according to the method proposed in Ref. 7, may be calculated numerically using the equation

$$\frac{d\mu_n}{d\tau} \approx \frac{2\mu_n}{2\tau - n\tau^2 - n\mu_n} \tag{6}$$

(the boundary conditions and choice of roots are derived from physical reasoning[7,9]).

Specific calculations show that we can only confine ourselves to μ_1 $(n=1)$ for a fairly thick $(r_c \sim 5\ \mu\text{m})$ microconductor. Note that modes with $n>1$ are attenuated more rapidly in a waveguide. Standard VSWR meters are not generally suitable for measuring the parameters of natural ferromagnetic resonance in thin microconductors, since if $r_c \lesssim \delta$, the microconductor is ''transparent'' for the field.

The reflectometric method proposed in Refs. 6 and 7 has another fundamental limitation. When allowance is made for dispersion at a frequency close to those of the maximum of the imaginary part of the magnetic permeability μ'', the real part μ' tends to zero. This necessitates taking account of the reactance of the conductor–waveguide contact which contributes to the systematic error. Without giving the equivalent circuit (see Ref. 10, for example), we merely observe that in this case

TABLE I.

Characteristics of amorphous alloy	Resonance frequencies GHz	Imaginary part of magnetic permeability at resonance	Estimated half-width, GHz
Iron-based alloy	7–10	2×10^2–10^3	0.5–2
Alloy of cobalt and iron in the ratio 5/2	6–3.5	2–5×10^2	1–2
Alloy of cobalt and iron in the ratio 6/1.5	up to 3	1.5–3×10^2	2

$$T \neq 1 - R \tag{7}$$

(T and R are the transmission and reflection coefficients) and the moduli $|T|$ and $|R|$ are not sufficient to calculate Z, but we need to know the phase between R and T. However, an estimate of the imaginary part of the magnetic permeability for a thick microconductor is more accurate in this method of measurement (as was noted in Refs. 4 and 8).

Table I gives characteristics of the magnetic properties of amorphous microconductors in the natural ferromagnetic resonance range, which may be used in radio absorbing materials.

[1] A. G. Alekseev and A. E. Kornev, *Magnetic Elastomers* [in Russian], Khimiya, Moscow (1987).
[2] S. A. Baranov, V. N. Berzhanskiĭ, S. K. Zotov *et al.*, Fiz. Met. Metalloved. **67**, 73 (1989).
[3] S. A. Baranov, S. K. Zotov, V. S. Larin, and A. V. Torkunov, Fiz. Met. Metalloved. **12**, 172 (1991).
[4] S. A. Baranov, S. K. Zotov, V. S. Larin, and A. V. Torkunov, *Magnetic Resonances in an Amorphous Microconductor and Magnetic Structures*, Proceedings of Conference of Young Scientists at the Physics Faculty of L'vov University [in Russian], L'vov University Press, L'vov (1990), pp. 5–7, Deposited Paper 763-UK91 (1991).
[5] S. A. Baranov, Vestn. Prednest. Univ. No. 1(2), 126 (1994).
[6] L. G. Gazyan and L. M. Suslov, Radiotekhnika, No. 7, 92 (1988).
[7] V. I. Ponomarenko, V. N. Berzhanskiĭ, I. V. Dzodolik *et al.*, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **32(3)**, 38 (1989).
[8] V. N. Berzhanskiĭ, L. G. Gazyan, V. L. Kokoz, and D. N. Vladimirov, Pis'ma Zh. Tekh. Fiz. **16**(12), 14 (1990) [Sov. Tech. Phys. Lett. **16**, 447 (1990)].
[9] S. A. Baranov, Vestn. Prednest. Univ. No. 6, 110 (1996).
[10] A. Vaĭsflokh, *Circuit Theory and Measurement Techniques in the Decimeter and Centimeter Ranges* [in Russian], Sov. Radio, Moscow, 1961.

Translated by R. M. Durham

Asymptotic fractality and the anomalous transport of particles having finite velocity

V. V. Uchaĭkin

Ulyanovsk State University, 432700 Ulyanovsk, Russia
(Submitted January 22, 1997)
Zh. Tekh. Fiz. **68**, 138–139 (January 1998)

A generalized time-dependent transport equation is obtained for particles whose free motion has a finite velocity, which includes ''Lévy flights'' and the effect of ''traps.'' It is shown that as a result of allowing for the finite velocity, the asymptotic (with respect to time) distribution of a particle walking in one dimension has a fractal nature only when the power-law tails of the mean-free-path distributions and particle residence times in the trap have the same exponents.
 [S1063-7842(98)02901-8]

Studies of the chaotic behavior of dynamical systems have provided the impetus for the development of a new branch of transport theory — anomalous diffusion or ''strange kinetics.''[1–4] The main difference between anomalous transport and normal transport is that the particle mean free paths ξ (''Lévy flights'') have a power-law distribution

$$p(\xi)\sim\alpha\xi_0^{\alpha}\xi^{-\alpha-1},\quad \xi\to\infty,\quad 0<\alpha<2 \tag{1}$$

and the analysis includes temporal residence of the particle in a state of rest (''traps''), also having a wide power-law distribution

$$q(\tau)\sim\beta\tau_0^{\beta}\tau^{-\beta-1},\quad \tau\to\infty,\quad 0<\beta<2. \tag{2}$$

When the first effect predominates, we talk of superdiffusion and in the opposite case, we talk of subdiffusion. The authors of Refs. 5–8 considered integral equations for anomalous transport, whereas the authors of Refs. 9–12 analyzed equations in fractional derivatives. Both approaches neglect the particle transit time between collisions, i.e., the velocity of the free particle motion is assumed to be infinite. In this approximation, the one-dimensional symmetric walking, fractal in space and time ($\alpha<1$ and $\beta<1$), is given by [formula (58) in Ref. 8]

$$\langle|x|\rangle=\int_{-\infty}^{\infty}|x|\psi(x,t)dx\sim\text{const}\ t^{\beta/\alpha},\quad t\to\infty, \tag{3}$$

where $\psi(x,t)$ is the coordinate distribution of a particle walking along the x axis at time t, having been created at time $t=0$ at the point $x=0$.

However, formula (3) is inaccurate, since for $\alpha<1$ the absolute first-order moment is infinite as a result of condition (1). The result obtained in Ref. 8 can be correctly formulated if formula (57) from that study is expressed in the form

$$\psi(x,t)\sim\frac{1}{m(t)}\Psi^{(\alpha,\beta)}\left(\frac{x}{m(t)}\right),\quad t\to\infty, \tag{4}$$

where

$$\Psi^{(\alpha,\beta)}(x)=\frac{1}{(2\pi)^2 i}\int_{c-i\infty}^{c+i\infty}du\int_{-\infty}^{\infty}dq\,\frac{Bu^{\beta-1}}{Bu^{\beta}+A|q|^{\alpha}}e^{u-iqx},$$

$$m(t)=t^{\beta/\alpha}. \tag{5}$$

The authors of Ref. 8 consider formula (5) to be ''...a basic result that provides extensive opportunities for discussion...'' (p. 154). Here we show that allowance for the finite velocity of the particle motion between collisions significantly alters the situation.

We assume that a particle is created at the origin at time $t=0$ in state 0 (trap) with the probability p_0 or in state 1 (flight with the velocity $\mathbf{v}=\mathbf{\Omega}v$, $v=\text{const}$) with the probability p_1, and we assume that the angular distributions of particles created in state 1 and particles leaving the traps are the same and equal to $W(\mathbf{\Omega})$

$$\int W(\mathbf{\Omega})d\mathbf{\Omega}=1$$

Following the logic of the derivation of the time-dependent integral equation in normal transport theory (see Ref. 13), under the assumptions made here we arrive at the following equation for the spatial distribution density of a walking particle at time t:

$$\psi(\mathbf{r},t)=\int_0^{\infty}dt'Q(t')F_0(\mathbf{r},t-t')+v^{-1}\int d\mathbf{r}'P(\mathbf{r}')F_1(\mathbf{r}-\mathbf{r}',t-r'/v), \tag{6}$$

$$F_0(\mathbf{r},t)=\int d\mathbf{r}'p(\mathbf{r}')F_1(\mathbf{r}-\mathbf{r}',t-r'/v)+p_0\delta(\mathbf{r})\delta(t), \tag{7}$$

$$F_1(\mathbf{r},t)=\int_0^{\infty}d\tau q(\tau)F_0(\mathbf{r},t-\tau)+p_1\delta(\mathbf{r})\delta(t), \tag{8}$$

where

$$Q(t)=\int_t^{\infty}q(\tau)d\tau,\quad p(\mathbf{r})=[W(\mathbf{r}/r)/r^2]p(r),$$

$$P(\mathbf{r})=[W(\mathbf{r}/r)/r^2]\int_r^{\infty}p(\xi)d\xi.$$

It can be confirmed that for exponential distributions $p(\xi)$ and $q(\tau)$ this result agrees with the usual single-velocity theory of delayed neutron transport, whereas for $v\to\infty$ it yields the known integral equations of anomalous diffusion.[1,5,8,12]

For a one-dimensional symmetric walk along the x axis with the direction vector $\mathbf{e}_x$, we have

$$W(\mathbf{\Omega}) = (1/2)[\delta(\mathbf{\Omega} - \mathbf{e}_x) + \delta(\mathbf{\Omega} + \mathbf{e}_x)]. \quad (9)$$

Substituting Eq. (9) into Eqs. (6)–(8) and integrating over the transverse coordinates y and z, we can easily obtain an equation for the one-dimensional density

$$\psi(x,t) \equiv \int\int \psi(x,y,z,t)\,dy\,dz.$$

Adopting the same procedure as in Ref. 8, i.e., performing a Fourier–Laplace transformation with respect to the variables x and t, using distributions (1) and (2), and converting to the asymptotic form $t \to \infty$, we obtain an analog of expression (4) for walking with the finite velocity v:

$$\psi(x,t) \sim t^{-\beta/\alpha} \Psi_v^{(\alpha,\beta)}(xt^{-\beta/\alpha}, t^{\beta/\alpha - 1}), \quad (10)$$

where

$$\Psi_v^{(\alpha,\beta)}(x,t) = \frac{1}{(2\pi)^2 i} \int_{-\infty}^{\infty} dq \int_{c-i\infty}^{c+i\infty} du \times \frac{Bu^{\beta-1} + v^{-1}(q\xi_0)^{\alpha-1} U[ut/(vq)]}{Bu^{\beta} + (q\xi_0)^{\alpha} V[ut/(vq)]} e^{u - iqx},$$

$$B = \Gamma(1-\beta)\tau_0^{\beta}, \quad U(z) = \int_0^{\infty} x^{-\alpha} \cos x e^{-zx} dx,$$

$$V(z) = \int_0^{\infty} x^{-\alpha} (\sin x + z \cos x) e^{-zx} dx.$$

It is easy to see that for $v \to \infty$ this result has the form (4)–(6), but when the particle velocity is finite, the self-similar behavior of the distribution (10) can only be obtained for equal exponents $\alpha = \beta$, when

$$\psi(x,t) \sim t^{-1} \Psi_v{}^{(\alpha,\alpha)}(xt^{-1}, 1).$$

Only in this case can we talk of the fractal nature of the anomalous walks, at least in the sense of a monofractal.

We note in passing that allowance for the finite particle velocity under normal diffusion conditions does not alter the self-similar character of the asymptotic form of the spatial distribution,

$$\psi(x,t) \sim \frac{1}{\sqrt{t}} \Phi\left(\frac{x}{\sqrt{t}}\right), \quad t \to \infty,$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi \langle \xi^2 \rangle / \langle \tau + \xi/v \rangle}} \exp\left\{ - \frac{x^2}{2\langle \xi^2 \rangle / \langle \tau + \xi/v \rangle} \right\}.$$

This work was supported by the Russian State Committee for Higher Education (Grant No. 95-0-3.1-23).

[1] J.-P. Bouchard and A. Georges, Phys. Rep. **195**, 127 (1990).
[2] M. B. Isichenko, Rev. Mod. Phys. **64**, 961 (1992).
[3] M. F. Shlesinger, G. M. Zaslavsky, and J. Klafter, Nature (London) **363**, 31 (1993).
[4] J. Klafter, M. F. Shlesinger, and G. Zumofen, Phys. Today **49(2)**, 33 (1996).
[5] E. W. Montroll and G. H. Weiss, J. Math. Phys. **6**, 167 (1965).
[6] E. W. Montroll and M. F. Shlesinger, in *Studies in Statistical Mechanics*, edited by J. Leibovitz and E. W. Montroll (North-Holland, Amsterdam, 1984), pp. 1–121.
[7] M. F. Shlesinger, Physica D **38**, 304 (1989).
[8] V. V. Afanasiev, R. Z. Sagdeev, and G. M. Zaslavsky, Chaos **1(2)**, 143 (1991).
[9] R. R. Nigmatullin, Phys. Status Solidi B **123**, 739 (1984).
[10] G. M. Zaslavsky, in *Topological Aspects of the Dynamics of Fluids and Plasmas*, edited by H. K. Moffatt, G. M. Zaslavsky, P. Comte, and M. Tabor (Kluwer, Dordrecht, 1992), pp. 481–491.
[11] G. M. Zaslavsky, in *Lévy Flights and Related Topics in Physics*, edited by M. F. Shlesinger, G. M. Zaslavsky, and U. Frisch (Springer-Verlag, Berlin, 1995), pp. 216–236.
[12] K. V. Chubar, Zh. Eksp. Teor. Fiz. **108**, 1875 (1995) [JETP **81**, 1025 (1995)].
[13] A. M. Kol'chuzhkin and V. V. Uchaĭkin *Introduction to the Theory of Particle Propagation Through Matter* [in Russian], Atomizdat, Moscow (1978).

Translated by R. M. Durham

Influence of electron irradiation of the NOVER-1 vacuum resist on its resistance to ion-beam etching

Yu. I. Koval' and V. T. Petrashov

Institute of Problems of Microelectronics Technology and High-Purity Materials, Russian Academy of Sciences, 142432 Chernogolovka, Moscow District, Russia
(Submitted May 19, 1997)
Zh. Tekh. Fiz. **68**, 140–142 (January 1998)

The influence of electron irradiation on the resistance of the NOVER-1 resist to ion-beam etching is studied. Etching is carried out by argon ions with energies between 300 and 2500 eV. It is found that, depending on the energy and angle of incidence of the ions on the surface of the resist, electron irradiation may either speed up or slow down the NOVER-1 etching. A clear correlation is observed between the penetration depth of the ions in the resist and the influence of the electron irradiation on the resistance of the resist to etching. At ion energies higher than 500 eV (ion penetration depth $\gtrsim 3.5$ nm) the resistance decreases, passes through a minimum at low electron irradiation doses, and returns to the etching rate of the initial resist at high doses. For glancing etching angles ($\sim 70°$ to the surface normal) and low ion energies (300 eV), i.e., small ion penetration depths ($\lesssim 2.5$ nm), an electron-irradiated resist is etched more slowly than the initial resist at all the electron irradiation doses studied. This effect may be used to enhance the resistance of resist structures whose height exceeds their width, which in this case is determined mostly by the rate of etching of the inclined facets. [S1063-7842(98)03001-3]

One of the important characteristics of resists is their resistance to ion-beam etching. It is known that polymer resists have a relatively low resistance to ion-beam etching, and this is a serious disadvantage when these resists are used as etching masks. Thick resists can be used to satisfy the resistance requirements but for submicron and particularly nanometer dimensions, an increase in the thickness of the resists leads to a deterioration in the resolution of the lithography and therefore is unacceptable in practice. In addition, thick masks give rise to various well known deleterious effects such as sputtering, the formation of grooves around the perimeter of structures caused by retroreflection of ions from the high walls of the resist, and a substantial difference in the resistance of structures of different sizes as a result of the faceting effect (see, e.g., Ref. 1).

The possibility of improving the resistance of resists has been considered on many occasions,[2–4] but our understanding of the mechanisms responsible for the etching of polymer materials is clearly inadequate to enable a specific search to be carried out. At present, it can be confidently asserted that, in many respects, the low resistance of resists is caused by the nature of the polymer materials, and that ion-beam etching of polymers does not involve physical sputtering but it is also largely determined by radiation-stimulated processes in the subsurface layer of material.[5]

The influence of different types of radiation on the resistance of resists to ion-beam etching may also provide information on the etching mechanisms of polymer materials. Thus, a study of the influence of electron irradiation on the resistance of materials to ion-beam etching seems to be urgently required. In addition, electron irradiation of the structures in a resist is of practical interest. For instance, it has been shown that electron irradiation of structures in an electronic resist such as polymethyl methacrylate (PMMA) can enhance their masking properties.[5] The behavior of the resist under heating during ion-beam etching also improved, and the influence of faceting diminished.

For our investigations we selected the NOVER-1 resist (negative organic vacuum e-beam resist). NOVER-1 was developed as a resist for electron-beam lithography capable of performing an entirely dry vacuum cycle — deposition, exposure, development, and etching. It has been shown in recent studies that in addition to the advantages characteristic of a vacuum resist (the possibility of uniform deposition on a surface of complex topology, the absence of any contact with developer fluids, and the possibility of avoiding contact between the samples and air), NOVER-1 has a high resolution (better than 30 nm) and forms continuous stable films less than 30 nm thick.[6,7]

In our experiments a resist around 0.3 μm thick was deposited by thermal evaporation from a boat at 150 °C. The substrates were Si wafers. Electron irradiation was carried out in a BS-300 relativistic electron machine with a programmed beam control system. The electron energy was 16 keV and the beam current 300 pA. Sections of dimensions 15×20 μm were exposed to doses between 3×10^{-3} and 9×10^{-1} C/cm^2. The resist was then etched with 300–2500 eV Ar^+ ions in an ion-beam etching system.[8] A difference between the rates of ion-beam etching of the unirradiated and irradiated regions resulted in the formation of a relief on the Si surface. After the resist had been completely removed by etching, the difference in height Δh between the unirradiated and irradiated regions was measured using an optical interferometer. The Si etching depth h_0 in the sections unprotected by the resist was also measured. The resistance of the resist was defined as $(h_0-\Delta h)/h_0$. This method can be used to obtain experimental points for the same sample at the same ion energy under known identical conditions of expo-

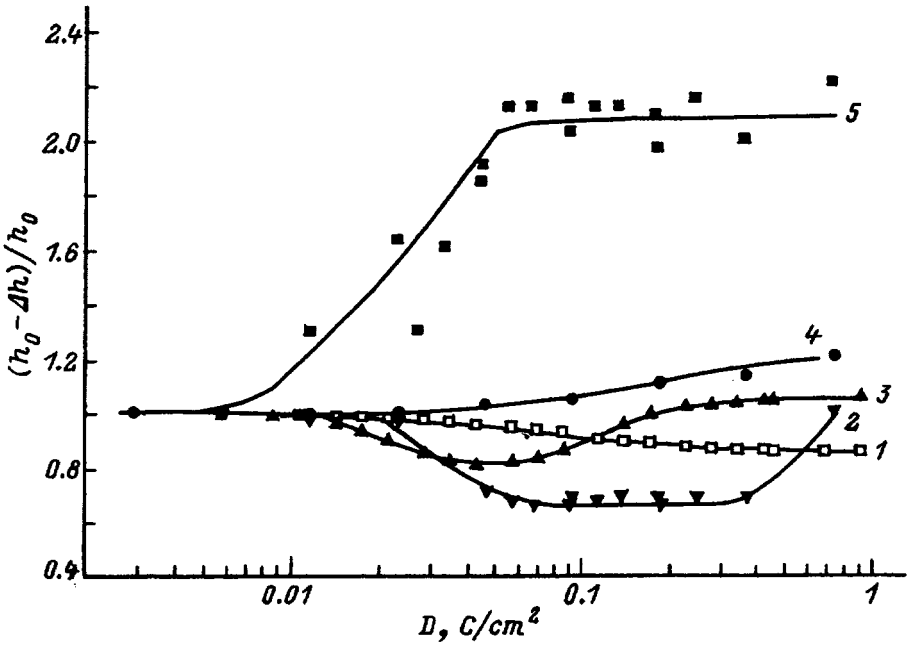

FIG. 1. Thickness of the resist film (*1*) and the resistance to ion-beam etching (*2*–*5*) versus the electron irradiation dose. Etching was carried out along the normal to the surface of the resist using 2500 (*2*), 500 (*3*), and 300 eV (*4*) ions and at an angle of 70° relative to the normal to the surface at 500 eV (*5*).

sure and etching on a resist film with the same initial thickness.

As a result of the electron irradiation, the thickness of the film decreases negligibly (curve *1* in Fig. 1). The kinetics of the decrease in thickness is described by the logarithmic dependence $H(D)=H_0(1-A\ \ln(1+D/D_0))$, where H_0 is the initial thickness of the resist, A and D_0 are parameters, and D is the electron irradiation dose. It is known that the decrease in thickness is a consequence of radiation damage. The low-molecular fragments formed leave the resist and the cavities are filled with surrounding molecules which reduces the volume of the material (in our case, the thickness of the resist). Estimates made using a method proposed in Ref. 9, show that the radiation yield g_e of gas formation in NOVER-1 per atom forming part of the volatile fragments is 0.45 atoms per 100 eV of absorbed energy. A comparison with other resists suggests that NOVER-1 has a relatively high resistance to radiation damage (for example, $g_e=12$ for PMMA).

The rate of ion-beam etching of organic materials correlates with their resistance to radiation damage. This is because resists are etched as a result of the simultaneous action of physical sputtering processes and the formation of weakly bound low-molecular fragments, which may be desorbed. Thus, the high resistance to radiation damage is one reason for the relatively high resistance of NOVER-1 to ion-beam etching. For instance, for 300 eV, 500 eV, and 2500 eV ions the rate of etching is 19 nm/min, 38 nm/min, and 97 nm/min, respectively, for an ion current density of 1 mA/cm^2.

Since the resist under study is a negative, damage under electron irradiation is dominated by cross linking processes. At high irradiation doses the resist is converted into a strongly cross-linked hydrocarbon network in which the formation of weakly bound fragments is less likely in ion-beam etching, and enhanced resistance to ion-beam etching could be expected. However, the experiments have shown that the influence of electron irradiation on the resistance to ion-beam etching is determined to a considerable extent by the etching conditions — by the energy and angle of incidence of the ions.

For ion-beam etching by 2500 eV ions the etching resistance decreases by $\sim$30% at doses higher than 3×10^{-2} C/cm^2, reaching a plateau in the range 5×10^{-2}–4×10^{-1} C/cm^2. A further increase in the preliminary electron irradiation dose increases the etching resistance almost to the initial state. A similar dependence is observed for the 500 eV ions. However, the dependence does not have a plateau near the minimum, the etching resistance decreases by only $\sim$20%, and begins to increase from substantially lower doses (6×10^{-2} C/cm^2). When the ion energy is reduced to 300 eV, no minimum is observed and the etching resistance increases monotonically with increasing electron irradiation dose. Finally, for etching by 500 eV ions at an angle of 70° to the normal to the surface of the sample, the etching resistance is more than doubled, reaching saturation in the dose range of 6×10^{-2} C/cm^2.

Thus, for this resist, the dependence of the resistance to ion-beam etching on the electron irradiation dose clearly correlates with the ion penetration depth. For etching at an angle of 70°, 500 eV ions penetrate to a depth of less than 1 nm, and a substantial increase in the etching resistance is observed. Normally incident 300 eV ions penetrate to a greater depth — around 2.5 nm, and the etching resistance of the resist increases as before, but substantially less. When the ion penetration depth equals 3.5 nm (500 eV ions), the etching resistance begins to vary nonmonotonically, and unlike the previous cases, it decreases in a certain range of doses. This behavior is observed most clearly for 2500 eV ions, whose penetration depth is greater than 7 nm. In this case, the range of electron irradiation doses in which the etching resistance decreases becomes substantially greater, and a plateau appears on the curve, as has already been discussed.

In spite of the correlation observed between the behavior

of the etching resistance of the resist and the ion penetration depth, the available data are insufficient to explain the observed dependences. However, the observed effect may be used to strengthen masks of the NOVER-1 resist with small structure dimensions (less than the thickness of the resist). In this case, the etching resistance of the resists is mainly determined by the etching resistance of the facets at the edges of the structure, whose normals are at 60–70° to the beam. The possibility of more than doubling the etching resistance of the structure edges more than doubles the etching resistance of the masks. In addition, strengthening the walls can reduce the drift of the dimensions for structures with dimensions of the order of or greater than the thickness of the resist films.

[1] *Plasma Processing for VLSI*, edited by N. G. Einspruch and D. M. Brown, Vol. 8 of VLSI Electronics [Academic Press, New York, 1984; Mir, Moscow, 1987].
[2] T. B. Borzenko, A. F. Vyatkin, N. N. Gonchkova *et al.*, Vacuum **38**, 1007 (1988).
[3] M. I. J. Beale, C. Broughton, A. J. Pidduck, and V. G. I. Deshmukh, Nucl. Instrum. Methods **19/20**, 995 (1987).
[4] H. Gokan, K. Tanigaki, and Y. Ohnishi, Solid State Technol. **5**, 163 (1985).
[5] T. B. Borzenko, Y. I. Koval, and V. A. Kudryashov, Microelectron. Eng. **23**, 337 (1994).
[6] V. T. Petrashov, Y. T. Abramenko, V. A. Zarubin *et al.*, Phys. Low-Dimens. Struct. **11/12**, 103 (1994).
[7] V. T. Petrashov, Ju. T. Abramenko, Ju. I. Koval, and L. Aparshina, Microelectron. Eng. **35**, 161 (1997).
[8] Yu. I. Koval', and E. V. Il'ichev, Prib. Tekh. Eksp. No. **3**, 118 (1994).
[9] T. B. Borzenko, Candidate's Dissertation [in Russian], Chernogolovka (1996).

Translated by R. M. Durham

Erratum: Atomic structure of silver clusters on silicon [Tech. Phys. 67, 1429–1432 (December 1997)]

M. V. Gomoyunova, I. I. Pronin, and N. S. Faradzhev

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
Zh. Tekh. Fiz. **68**, 144 (January 1998)

[S1063-7842(98)03201-2]

On pages 1430–1432 the planes {110} and {111} should instead read {111} and {110}, respectively.

In Sec. 2, *Experimental techniques*, 6 lines up from the bottom of p. 1429, the sample dimensions should read $22\times14\times0.25$ mm rather than $22\times4\times0.25$ mm.

In Sec. 3, *Results and discussion*, the phrase ''low-temperature atomic planes'' in the sixth line up from the bottom of the right-hand column should read ''low-index atomic planes.''

Translated by Steve Torstveit

 1063-7842/98/010129-01$15.00

Parameters for the inclined-path sensing of molecular hydrogen in the atmosphere by lidar with a YAG:Nd laser

G. V. Laktyushkin, V. E. Privalov, and V. G. Shemanin

Baltic State Technical University, 198005 St. Petersburg, Russia
(Submitted December 16, 1996)
Zh. Tekh. Fiz. **68**, 20–22 (January 1998)

The lidar equation for the vibrational backscattering of neodymium laser radiation and its harmonics by hydrogen molecules is solved numerically. Inclined paths in the atmosphere are investigated with the aim of selecting the transmitter wavelength for detecting the lowest concentrations of hydrogen. © *1998 American Institute of Physics.* [S1063-7842(98)00401-2]

YAG:Nd laser radiation, especially its second, third, and fourth harmonics, is widely used in remote sensing systems[1] and permits the generation of pulses with a duration of 10 ns and energies from 1 to 100 mJ at pulse repetition rates as high as 50 Hz. The radiation at the third and fourth harmonics of such a laser can be used for the remote sensing of molecular hydrogen in the atmosphere by a Raman lidar.[1] Therefore, it would be interesting to numerically solve the lidar equation for the vibrational Raman backscattering of H_2 molecules for a YAG:Nd laser and inclined paths in the atmosphere at a distance of 6 km and an altitude of 6 km for the purpose of selecting the wavelength of a Raman lidar transmitter for detecting the smallest possible concentration of hydrogen.

We write the equation for Raman backscattering, as in Ref. 2, in the form

$$P(\lambda,R)=P_0(\lambda_0)K\Delta RA_2T(\lambda_0)T(\lambda)\left(\frac{d\sigma}{d\Omega}\right)N_aiR^{-2}, \quad (1)$$

where $P(\lambda,R)$ is the power of the Raman scattering signal in the photodetector at the wavelength λ arriving from a distance R, $P_0(\lambda_0)$ is the power of the laser at its wavelength, K is the lidar constant, ΔR is the spacing with respect to the distance, A_2 is the area of the receiving telescope, $T(\lambda_0)$ and $T(\lambda)$ are the transmission of the atmosphere at the wavelengths of the laser radiation and the Raman backscattering signal, respectively, $(d\sigma/d\Omega)$ is the differential cross section for the vibrational Raman scattering of the molecule under investigation, N_a is the concentration of the molecules, and R is the distance to the sensing point.

The wavelengths of the Raman scattering bands of the H_2 molecules investigated were calculated for different wavelengths of laser radiation from the formula

$$\lambda_{RH}=1\left(\frac{1}{\lambda_0}-\tilde{\nu}\right), \quad (2)$$

where $\tilde{\nu}$ is the frequency of the eigenmodes of H_2, and are presented in the fourth column in Table I.

Following Ref. 1 and replacing the angular frequency ω by the wavelength λ, we can determine the differential cross section for vibrational Raman backscattering from the formula

$$\left(\frac{d\sigma}{d\Omega}\right)_j=\frac{16\pi^4b_j^2g_j}{\lambda^4[1-\exp(-hc/\lambda kT)]}\left\{a_j^2+\frac{7}{45}\nu_j^2\right\}, \quad (3)$$

where b_j is the amplitude of the zero-point vibrations of the jth mode, g_j is its degree of degeneracy, $3a_j$ and ν_j are the trace and the anisotropy of the tensor of the derivative of the polarizability of the molecule with respect to the normal coordinate g_j, T is the vibrational temperature of the molecule, k and h are Boltzmann's and Planck's constants, respectively, and c is the speed of light.

Leaving only the dependence on λ, we can rewrite Eq. (3) in the form

$$\left(\frac{d\sigma}{d\Omega}\right)_j=A/\lambda^4, \quad (4)$$

where the constant A is determined from the known value of the cross section of a H_2 molecule for the emission wavelength of a nitrogen laser $\lambda_0=337.1$ nm, viz., $(d\sigma/d\Omega)_j=8.7\times10^{-30}$ cm^3/sr, which is presented in the last row of Table I. The constant obtained equals 1.13065×10^{-17} cm^{2}nm^4.

The cross section values obtained for the laser wavelengths selected are presented in the second column of Table I. Henceforth, for the specific case of our lidar the constant K_1 depends on the spectral sensitivity $\xi_\nu(\lambda)$ of the photomultiplier photocathode as

$$K_1=K_2\xi_\nu(\lambda). \quad (5)$$

The remaining multipliers in Eq. (1) have the following values: $\Delta R=7.5$ m for the measurement time $t_d=50$; $A_2=0.008$ m^2; $K_2=0.495$ for a wavelength of 532 nm (the measurement result); the energy of a laser pulse $E_0=100$ mJ; the sensing distance $R=1,2,3,4,5$, and 6 km; and the altitude $H=2,3,4,5$, and 6 km. The values of the spectral sensitivity of the FÉU-79 and FÉU-140 (FÉU-124) photocathodes in the ultraviolet region were taken from Ref. 3, and their relative values are listed in the sixth column of Table I. The transmission of the atmosphere was calculated, as in Ref. 1, using the formula

$$T(\lambda,R)=\exp\left[-\int_0^R k(\lambda)dR\right] \quad (6)$$

from the values of the attenuation factor k, which were taken from Ref. 4 and are presented in the third and fifth columns

TABLE I. Values of the differential cross sections, wavelengths of the Raman scattering bands of hydrogen molecules, attenuation factors of the atmosphere, relative spectral sensitivity of the photomultiplier, and spectral brightness of the background solar radiation calculated for the wavelengths of the second, third, and fourth harmonics of a YAG:Nd laser and wavelengths of the Raman scattering bands ($\nu = 4161$ cm^{-1}).

λ, nm	$(d\sigma/d\Omega)\cdot 10^{30}$, cm/sr	k, km^{-1}	λ_{KR}, nm	k, km^{-1}	$\xi_p(\lambda)$	$S_b\cdot 10^3$, W/m·sr
532	1.40	0.16	683.2	0.145	0.25	7.9
355	7.07	0.31	416.5	0.21	0.52	12.4
266	22.41	0.785	299.1	0.45	0.30	6.0
337.1	8.7		392.7			

of Table I for the wavelengths of interest to us. The variation of the attenuation factor with the altitude H was taken into account in accordance with the data in Ref. 4, which were approximated by a function of the form

$$k(H) = k(0)\exp(-0.79H). \quad (7)$$

Using the parameters presented above, we performed numerical calculations of the Raman backscattering power on the basis of Eq. (1) for the wavelengths selected and a laser pulse energy $E_0 = 100$ mJ in the range of sensing distances from 1.0 to 60 km with variation of the altitude also from 1.0 to 6.0 km for the purpose of finding the optimum wavelength for detecting a very small concentration of hydrogen molecules. The results of the calculations for the wavelengths that we selected are presented in Table II. It follows from it that increasing the angle of inclination of the sensing path (or its altitude) leads to a decrease in the Raman scattering power, which, however, is smaller than in the case of sensing on a horizontal path without variation of the spectral dependence of the multipliers appearing in the lidar equation (1). As the distance is increased, the Raman scattering power decreases by almost four orders of magnitude in the range from 1 to 6 km, but when the altitude is increased to 6 km, this decrease amounts to only 14 fold. This difference is attributed to the strong influence of the absorption of the laser radiation in the layer of the atmosphere near the ground. An analysis of these results shows that the use of a laser operating at the third harmonic with a wavelength of 355 nm is optimal, since it provides the maximum value of the Raman backscattering power for the molecules investigated in this range of distances and altitudes.

TABLE II. Results of calculations of the Raman backscattering power of H_2 molecules for wavelengths of harmonics of a YAG:Nd laser with a pulse energy of 100 mJ, a sensing distance of 1–6 km, an altitude of 1–6 km, and a concentration of molecules equal to 10^{20} cm^{-3}.

		R, km					
		1.0	2.0	3.0	4.0	5.0	6.0
H, km	λ, nm	P, nW					
0.0	683.2	7696	1426	469.7	195.5	92.79	47.78
	416.5	65900	9888	2643.0	891.1	342.30	142.70
	299.1	58060	4224	545.6	88.5	16.77	3.34
1.0	683.2	1931	358	117.9	49.05	23.28	11.99
	416.5	16530	2481	663.1	223.60	85.89	35.81
	299.1	14570	1060	136.9	22.21	4.21	0.84
2.0	683.2	1031	191.1	62.9	26.19	12.43	6.40
	416.5	8828	1325.0	354.1	119.40	45.86	19.12
	299.1	7779	565.9	73.1	11.86	2.25	0.45
3.0	683.2	776	143.7	47.33	19.70	9.35	4.81
	416.5	6640	996.3	266.30	89.79	34.49	14.38
	299.1	5851	425.6	54.98	8.92	1.69	0.34
4.0	683.2	682	126.3	41.59	17.31	8.22	4.23
	416.5	5835	875.5	234.00	78.91	30.31	12.64
	299.1	5141	374.0	48.31	7.84	1.48	0.30
5.0	683.2	643	119.1	39.22	16.32	7.75	4.00
	416.5	5502	825.6	220.70	74.41	28.59	11.92
	299.1	4848	352.7	45.56	7.39	1.40	0.28
6.0	683.2	626	116.0	38.19	15.90	7.54	3.88
	416.5	5358	803.9	214.90	72.46	27.83	11.61
	299.1	4721	343.4	44.36	7.20	1.36	0.27

TABLE III. Results of calculations of the minimum power that can be detected by a lidar for wavelengths of Raman scattering bands of hydrogen molecules, sensing distances from 1 to 6 km, and altitudes from 1 to 6 km.

		R, km					
		1.0	2.0	3.0	4.0	5.0	6.0
H, km	λ, nm	P_m, fW					
0.0	683.2	81.52	17.74	6.85	3.35	1.86	1.13
	416.5	248.09	50.32	18.15	8.28	4.29	2.42
	299.1	54.55	8.70	2.46	0.88	0.36	0.14
1.0	683.2	40.81	8.89	3.43	1.68	0.93	0.56
	416.5	124.39	25.21	9.09	4.14	2.13	1.21
	299.1	27.36	4.36	1.24	0.42	0.18	0.08
2.0	683.2	29.83	6.50	2.50	1.23	0.68	0.41
	416.5	91.0	18.39	6.64	3.03	1.57	0.88
	299.1	20.0	3.19	0.90	0.32	0.13	0.06
3.0	683.2	25.89	5.63	2.18	1.06	0.59	0.34
	416.5	78.74	16.09	5.76	2.62	1.36	0.58
	299.1	17.36	2.76	0.76	0.28	0.11	0.05
4.0	683.2	24.30	5.28	2.04	1.00	0.55	0.34
	416.5	73.53	15.02	5.41	2.47	1.27	0.72
	299.1	16.26	2.59	0.73	0.27	0.10	0.05
5.0	683.2	23.56	5.11	1.98	0.97	0.54	0.33
	416.5	71.69	14.56	5.24	2.39	1.26	0.70
	299.1	15.73	2.52	0.71	0.26	0.10	0.05
6.0	683.2	23.27	5.07	1.95	0.96	0.53	0.32
	416.5	70.77	14.33	5.11	2.36	1.22	0.69
	299.1	15.56	2.48	0.70	0.25	0.10	0.05

However, these calculations were performed for the case of the absence of background illumination or for precision sensing. Since the solar background radiation has a strong influence on the Raman scattering power detected by a lidar, we performed calculations of the background power in the photodetector $P_b(\lambda, R)$ and examined the influence of the background illumination on the potential capabilities of a lidar. The value of the spectral brightness of the solar radiation for different dates during the year, times of day, and weather conditions were taken from Ref. 5. The conditions for a bright sunny day were chosen as the severest conditions for lidar operation, and the data in Refs. 1,6, and 7 were used (because of the uncertainty regarding the orientation of the telescope axis relative to the direction to the sun) to construct the spectral distribution of the background radiation $S_0(\lambda)$, whose values are presented in the last column of Table I for the wavelengths selected. Using these values of $S_0(\lambda)$ and the equation

TABLE IV. Results of calculations of the minimum detectable concentrations of H_2 molecules with a lidar for wavelengths of harmonics of a YAG:Nd laser, sensing distances from 1 to 6 km, and altitudes from 1 to 6 km.

		R, km					
		1.0	2.0	3.0	4.0	5.0	6.0
H, km	λ, nm	N_A, cm^{-3}					
0.0	683.2	1.1×10^{15}	1.2×10^{15}	1.5×10^{15}	1.7×10^{15}	2.1×10^{15}	2.4×10^{15}
	416.5	3.8×10^{14}	5.1×10^{14}	6.9×10^{14}	9.3×10^{14}	1.3×10^{15}	1.7×10^{15}
	299.1	9.4×10^{13}	2.1×10^{14}	4.5×10^{14}	9.9×10^{14}	2.2×10^{15}	4.2×10^{15}
1.0	683.2	2.1×10^{15}	2.5×10^{15}	2.9×10^{15}	3.4×10^{15}	4.0×10^{15}	4.7×10^{15}
	416.5	7.5×10^{14}	1.0×10^{15}	1.4×10^{15}	1.9×10^{15}	2.5×10^{15}	3.4×10^{15}
	299.1	2.0×10^{14}	4.1×10^{14}	9.1×10^{14}	1.9×10^{15}	4.3×10^{15}	9.5×10^{15}
2.0	683.2	2.9×10^{15}	3.4×10^{15}	4.0×10^{15}	4.7×10^{15}	5.5×10^{15}	6.4×10^{15}
	416.5	1.0×10^{14}	1.4×10^{15}	1.9×10^{15}	2.5×10^{15}	3.0×10^{15}	4.6×10^{15}
	299.1	2.6×10^{14}	5.6×10^{14}	1.2×10^{15}	2.7×10^{15}	5.8×10^{15}	1.3×10^{16}
3.0	683.2	3.3×10^{15}	3.9×10^{15}	4.6×10^{15}	5.4×10^{15}	6.3×10^{15}	7.5×10^{15}
	416.5	1.2×10^{15}	1.6×10^{15}	2.2×10^{15}	2.9×10^{15}	3.9×10^{15}	4.0×10^{15}
	299.1	3.0×10^{14}	6.5×10^{14}	1.4×10^{15}	3.1×10^{15}	6.5×10^{15}	1.5×10^{16}
4.0	683.2	3.5×10^{15}	4.2×10^{15}	4.9×10^{15}	5.8×10^{15}	6.7×10^{15}	8.0×10^{15}
	416.5	1.3×10^{15}	1.7×10^{15}	2.3×10^{15}	3.1×10^{15}	4.2×10^{15}	5.7×10^{15}
	299.1	3.2×10^{14}	6.9×10^{14}	1.5×10^{15}	3.4×10^{15}	6.8×10^{15}	1.7×10^{16}
5.0	683.2	3.7×10^{15}	4.3×10^{15}	5.1×10^{15}	5.9×10^{15}	7.0×10^{15}	8.3×10^{15}
	416.5	1.3×10^{15}	1.8×10^{15}	2.4×10^{15}	3.2×10^{15}	4.4×10^{15}	5.8×10^{15}
	299.1	3.2×10^{14}	7.1×10^{14}	1.6×10^{15}	3.5×10^{15}	7.1×10^{15}	1.8×10^{16}
6.0	683.2	3.7×10^{15}	4.4×10^{15}	5.1×10^{15}	6.0×10^{15}	7.3×10^{15}	8.3×10^{15}
	416.5	1.3×10^{15}	1.8×10^{15}	2.4×10^{15}	3.3×10^{15}	4.4×10^{15}	5.6×10^{15}
	299.1	3.3×10^{14}	7.2×10^{14}	1.6×10^{15}	3.5×10^{15}	7.4×10^{15}	1.9×10^{16}

$$P_b(\lambda,R)=S_b(\lambda)T(\lambda,R)K_2\xi_\nu(\lambda)A_2\Omega(R)\Delta\lambda \qquad (8)$$

[$\Omega(R)$ is the solid angle of the field of vision of the receiving telescope, and $\Delta\lambda$ is the spectral width of the reception channel], which is similar to the equation in Ref. 7, we calculated the values of the background power $P_b(\lambda,R)$ for our case. Assuming, as in Ref. 1, a minimum permissible signal-to-noise ratio (S/N) equal to 1.5, as in Ref. 1, we can determine the minimum power P_m that can be detected by a lidar according to the equation

$$P_m=(S/N)P_b(\lambda,R). \qquad (9)$$

The calculation results obtained are presented in Table III. A comparison of these results with the data in Table II allows us to conclude that the largest excess of the Raman scattering power above the background level was obtained for the wavelengths of 266 and 355 nm over the entire range of distances and altitudes. A 10 MW laser operating at these wavelengths permits the detection of the concentrations of H_2 molecules listed in Table IV. It should be noted that the values of the minimum detectable concentrations increase for all the wavelengths by a factor of 3–4 as the altitude is varied from 0 to 6 km and by a factor of 2.2 for 523 nm, a factor of 4.5 for 355 nm, and a factor of 44.7 for 266 nm in the range of distances from 1 to 6 km selected. A minimum concentration of 0.4×10^{13} cm^3 is provided by fourth-harmonic radiation at a distance of 1 km.

Thus, the results obtained demonstrate the possibility of the optimum choice of the wavelength of laser radiation for sensing molecular hydrogen in a required concentration on inclined paths in the atmosphere at an assigned distance with consideration of the background conditions.

[1] R. M. Measures, *Laser Remote Sensing: Fundamentals and Applications* [Wiley, New York (1984); Mir, Moscow (1987)].
[2] E. K. Ivanov, K. A. Kolbenkov, L. A. Konopel'ko, and V. V. Rastoskuev, Izmer. Tekh. **5**, 56 (1986).
[3] *Laser Handbook* [in Russian], A. M. Prokhorov (ed.), Sov. Radio, Moscow (1978), Vol. 2, p. 182.
[4] *Laser Handbook* [in Russian, A. M. Prokhorov (ed.), Sov. Radio, Moscow (1978), Vol. 1, p. 382.
[5] P. Camagni, in *Lidar Applications to Aerosols and Particles, Optical Remote Sensing of Air Pollution. Lectures of a Course*, Ispra, Italy (1983), pp. 205–253.
[6] H. Rosen, P. Robish, and O. Chamberlain, Appl. Opt. **14**, 2703 (1975).
[7] H. Inaba and T. Kobayasi, Opto-electronics **4**, 101 (1972).

Translated by P. Shelnitz

Instability of the flat surface of a magnetic fluid in a cylindrical cavity in the presence of a vertical magnetic field

V. M. Korovin and A. A. Kubasov

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted January 19, 1996)
Zh. Tekh. Fiz. **68**, 23–30 (January 1998)

The problem of a magnetic liquid which completely fills a vertical cylindrical cavity in an undeformable horizontal layer of a magnet having the same magnetic properties as the liquid is considered. The entire system is immersed in a uniform vertical magnetic field. in a linear formulation of the problem an approximate solution in the form of series is obtained for the evolution of an initial small deviation of the free surface of the liquid from its flat equilibrium shape. An experiment is performed which shows that the initially flat free surface takes on a stable domed shape as the field strength is increased (from zero) and that a further increase in the field in a certain restricted range leads to the formation of an annular corrugation. The structures observed, which are the result of the nonlinear stage in the development of the initial perturbation, are qualitatively similar to the first two modes of the solution obtained. [S1063-7842(98)00501-7]

INTRODUCTION

The instability of the free surface of a nonviscous magnetic liquid in a fairly strong orthogonal magnetic field, which was discovered and first investigated about 30 years ago,[1,2] is one of the best-known surface phenomena characteristic of magnetic liquids. The experimental study and linear analysis of the influence of a vertical magnetic field on the stability of the initially flat, free surface of a magnetic liquid occupying the lower half space in Ref. 1 prompted both an investigation of this question within complicated formulations based on the linearized equations of ferrohydrodynamics[3] and the study of the appearance and reorganization (as the field is increased) of the periodic hexagonal structure on an initially flat, unbounded surface of a magnetic liquid in a nonlinear formulation (see, for example, the references cited in Refs. 2 and 4).

In this paper we investigate the instability of the flat free surface of a magnetic liquid occupying a vertical cylindrical cavity in a horizontal undeformable layer of a magnet, whose magnetic permeability is equal to the permeability of the liquid, due to a uniform vertical magnetic field. Unlike the previous studies, here the influence of the vertical wall is taken into account approximately in the theoretical analysis. This factor imposes a condition on the surface that the normal component of the velocity vanishes during development of the initial perturbations, which ultimately leads to destruction of the original hydrostatic state of a liquid with a flat free surface. An approximate solution of the problem of the development of an initial perturbation of the shape of the free surface is obtained analytically in a linear formulation; the first two modes exhibit a qualitative similarity to the structures observed in our experiment.

EQUILIBRIUM STATE

A magnetic liquid (Fig. 1), which completely fills a vertical cylindrical cavity *1* of radius a in a horizontal layer of a magnet *2* with undeformable boundaries, whose magnetic permeability μ is equal to that of the liquid, in a uniform vertical magnetic field $\mathbf{H}_\infty$, is considered. A nonmagnetic vertical wall *3* is located above the layer of the magnet strictly along the boundary of the cavity. The liquid is confined from below by a nonmagnetic plate *4*, which is in close contact with the magnet, while the upper surface of the liquid is free. The depth of the liquid is equal to the thickness d of the layer. In the case where the contact angle Θ formed by the liquid in contact with the material of the vertical boundary is equal to $\pi/2$, the free surface is obviously horizontal. In the general case there is a meniscus *5* near the wall due to the capillary rise of the liquid wetting the wall.

Following the reasoning used in Quincke's classical problem[2] to justify the neglect of the menisci near two flat vertical magnetic pole shoes that are parallel to one another and are partially immersed in a reservoir containing a magnetic liquid, we shall consider a cavity of sufficiently large radius so that the bending of the free surface due to capillary rise would be manifested only in an annular region that is narrow compared with a and is adjacent to the wall. In view

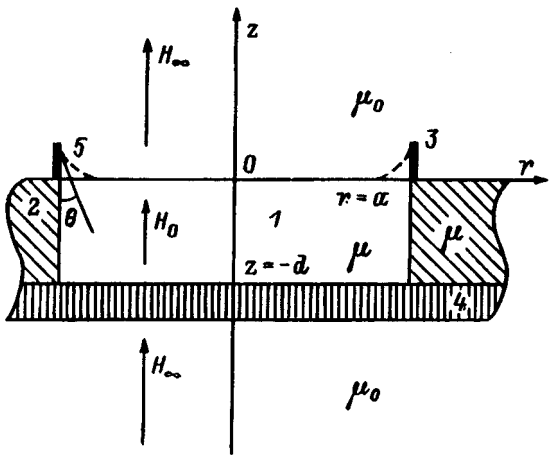

FIG. 1.

 1063-7842/98/010019-08$15.00

of the small size of this region it will henceforth be neglected, and it will be assumed that the free surface of the liquid is flat and lies at the same level as the upper boundary of the magnet.

Utilizing the solution of the problem of the shape of the free surface of a liquid which is in a gravitational field and in contact on one side with a vertical flat wall,[5] in the cases of $\Theta \neq 0$ and $\Theta \neq \pi/2$ we can easily write (within an order of magnitude, i.e., without consideration of the influence of the meniscus on the distribution of the magnetic field) the condition for applicability of such an approach

$$[2B_0^{-1}(1-\sin\Theta)]^{1/2}\tan\Theta \ll 1,$$

where $B_0=\rho g a^2/\alpha$ is the Bond number, α is the surface tension, ρ is the density of the liquid, and g is the acceleration of gravity.

Within the assumptions made the magnetic field $\mathbf{H}_0$ is uniform both in the liquid and within the layer of the magnet. Assuming that the function $\mu=\mu(H)$ is known, we can easily express $\mathbf{H}_0$ in terms of $\mathbf{H}_\infty$ on the basis of the condition of continuity of the normal component of the induction $\mathbf{B}_0=\mu\mathbf{H}_0$ on the horizontal boundaries of the region occupied by the liquid and the magnet.

We introduce the cylindrical coordinate system r,ϑ,z, whose z axis is directed vertically upward and whose $z=0$ plane coincides with the free surface of the motionless magnetic liquid and the upper surface of the layer of the magnet. Owing to the uniformity of the induction, the strong influence of the magnetic field on the liquid is effected in the case under consideration only by the surface ponderomotive force localized at the liquid–air interface, i.e., the magnetic pressure according to the terminology in Ref. 2. Using the equation of hydrostatics and the condition of continuity of the normal stress on the liquid–air interface, we obtain, as in the case of an ordinary liquid, the linear variation of the pressure p_0 with the depth

$$p_0=p_a+\frac{\mu_0}{2}M_0^2-\rho g z, \quad \mathbf{M}_0=\frac{1}{\mu_0}\mathbf{B}_0-\mathbf{H}_0. \tag{1}$$

Here p_a is the pressure of the air near the free surface of the liquid, and $\mu_0 4\pi\times 10^{-7}$ H·m^{-1} is the magnetic constant.

It is known[2] that physical realization of the hydrostatic state of a layer of a magnetic liquid extended in horizontal directions with a flat free surface in the presence of an orthogonal magnetic field requires that the magnetization M_0 be smaller than a critical value. In the case under consideration, of course, the critical magnetization differs from the value calculated in Refs. 1 and 2.

FORMULATION OF THE STABILITY PROBLEM

Assuming that the liquid is nonviscous, we obtain a linear formulation of the problem of the stability of a hydrostatic state $\mathbf{H}_0$, p_0 with a flat free surface (1). Let $z=\eta(r,\vartheta,t)$ be the equation of the free surface of the liquid in the presence of standing waves of small amplitude, where t denotes the time. Introducing the potential of a small perturbation of the magnetic field $\mathbf{h}=(r,\vartheta,z,t)=\Delta\psi$, we write the magnetic fields, the induction, and the magnetization perturbed as a result of the deformation of the free surface caused by the waves, respectively, in the form

$$\mathbf{H}=\Delta(H_0 z+\psi), \quad \mathbf{B}=\mathbf{B}_0+\mathbf{b}, \quad \mathbf{M}=\mathbf{M}_0+\mathbf{m},$$

$$\mathbf{H}_j=\Delta(H_\infty z+\psi_j); \quad j=1,2,$$

where the index 1 refers to the field in the region $z>\eta$, and the index 2 refers to the region $z<-d$.

We also represent the distribution of the pressure in the liquid in the presence of a wave $P=p_0+p$ in a similar manner. With accuracy to within small first-order quantities, we have

$$H-H_0=\frac{\partial\psi}{\partial z}, \quad M-M_0=\chi_t^0\frac{\partial\psi}{\partial z}, \quad \chi_t^0=\left.\frac{dM}{dH}\right|_{H=H_0},$$

$$\mathbf{m}=\chi_c^0\Delta_2\psi+\chi_t^0\frac{\partial\psi}{\partial z}\mathbf{a}_z, \quad \chi_c^0=\frac{M_0}{H_0},$$

$$\mathbf{b}=\mu_c^0\Delta_2\psi+\mu_t^0\frac{\partial\psi}{\partial z}\mathbf{a}_z, \quad \mu_c^0=\frac{B_0}{H_0}, \quad \mu_t^0=\left.\frac{dB}{dH}\right|_{H=H_0},$$

$$\Delta_2=\frac{\partial}{\partial r}\mathbf{a}_r+\frac{1}{r}\frac{\partial}{\partial\vartheta}\mathbf{a}_\vartheta,$$

where $\mathbf{a}_r$, $\mathbf{a}_\vartheta$, and $\mathbf{a}_z$ are unit vectors along the respective coordinate axes.

When $\mu=\mu(H)$, in the case under consideration the magnetic force $\mu_0 M\Delta H$, as well as the motion of the liquid, are related to potentials. Introducing the velocity potential $\mathbf{v}=\Delta\varphi$, from the continuity equation we have

$$\nabla_2^2\varphi+\frac{\partial^2\varphi}{\partial z^2}=0, \quad \nabla_2^2=\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\vartheta^2}. \tag{3}$$

Using (2), from the linearized equation of motion we can easily obtain a linearized Cauchy–Lagrange integral, which can be used to calculate the perturbation of the pressure p,

$$p=-\rho\frac{\partial\varphi}{\partial t}+\mu_0 M_0\frac{\partial\psi}{\partial z}. \tag{4}$$

The distribution of the potential of the perturbation of the magnetic field is described by the Laplace equation. In the region occupied by the liquid and the layer of the magnet, from the equations of magnetostatics with consideration of (2) we obtain

$$\nabla_2^2\psi+\frac{1}{\sigma^2}\frac{\partial^2\psi}{\partial z^2}=0, \quad \sigma=\sqrt{\frac{\mu_c^0}{\mu_t^0}}\geqslant 1, \tag{5}$$

while outside of this region we have

$$\nabla_2^2\psi_j+\frac{\partial^2\psi_j}{\partial z^2}=0; \quad j=1,2. \tag{6}$$

The linearized kinematic and dynamic conditions on the free surface of the liquid with consideration of (2) and (4) are written in the following manner

$$z=0: \quad \frac{\partial\eta}{\partial t}=\frac{\partial\varphi}{\partial z}, \tag{7}$$

$$z=0: \quad \rho\frac{\partial\varphi}{\partial t}+\rho g\eta-\alpha\nabla_2^2\eta-\mu_t^0 M_0\frac{\partial\psi}{\partial z}=0. \tag{8}$$

The last term in the dynamic condition (8) represents the perturbation of the surface ponderomotive force due to the deformation of the free surface caused by the waves.

We impose a condition that the normal component of the velocity must vanish on the impermeable cavity boundaries:

$$r=a, \quad -d\leqslant z\leqslant 0: \quad \frac{\partial\varphi}{\partial r}=0, \tag{9}$$

$$0\leqslant r\leqslant a, \quad z=-d: \quad \frac{\partial\varphi}{\partial z}=0. \tag{10}$$

In the approximation under consideration the continuity conditions of the magnetic field potential and the normal component of the induction on the horizontal interfaces between the magnetic and nonmagnetic media have the form

$$z=0: \quad \psi-\psi_1=M_0\eta \quad \text{for } 0\leqslant r<a,$$

$$\psi=\psi_1 \quad \text{for } a<r<\infty,$$

$$z=0: \quad \mu_t^0\frac{\partial\psi}{\partial z}=\mu_0\frac{\partial\psi_1}{\partial z} \quad \text{for } 0=r<\infty,$$

$$z=-d: \quad \psi=\psi_2,$$

$$\mu_t^0\frac{\partial\psi}{\partial z}=\mu_0\frac{\partial\psi_2}{\partial z} \quad \text{for } 0\leqslant r<\infty. \tag{11}$$

Of course, physical meaning is attached only to the solutions of Eqs. (5) and (6) which ensure fulfillment of the conditions

$$|\psi(0,\vartheta,z,t)|<\infty, \quad |\psi_j(0,\vartheta,z,t)|<\infty; \quad j=1,2,$$

$$\Delta\psi\to 0, \quad \Delta\psi_j\to 0 \quad \text{for } r^2+z^2\to\infty.$$

We assume that in the initial moment in time a nonflat shape is assigned to the free surface and that the liquid is in a state of rest:

$$\eta(r,\vartheta,0)=f(r,\vartheta), \quad \varphi(r,\vartheta,z,0)=0. \tag{12}$$

To simplify the further mathematical manipulations, it is expedient to rewrite the dynamic condition (8) on the free surface. Differentiating (8) with respect to time, with consideration of kinematic condition (7) and Laplace equation (3) we obtain

$$z=0: \quad \rho\frac{\partial^2\varphi}{\partial t^2}+\rho g\frac{\partial\varphi}{\partial z}+\alpha\frac{\partial^3\varphi}{\partial z^3}-\mu_t^0 M_0\frac{\partial^2\psi}{\partial t\partial z}=0. \tag{13}$$

Thus, in the linear formulation the development of the initial perturbation (12) is described by Laplace equations (3), (5), and (6), whose solutions must satisfy boundary conditions (7), (9)–(11), and (13).

APPROXIMATION SOLUTION

In the problem under consideration the magnetic and hydrodynamic fields influence one another because of the conditions for matching of the functions sought on the free surface. Its shape is known only at $t=0$, and it is subject to determination in subsequent moments in time. Even in the initial moment, when the shape of the boundaries separating the magnetized and unmagnetized media is known everywhere, calculating the magnetic field perturbation caused by deformation of the free surface of the liquid is a complex task. The possibility, in principle, of obtaining a solution for it is associated with the use of numerical methods.

Although discussing the computational algorithm is not our purpose, let us assume that there is an iteration method for calculating the functions sought at any t based on their systematic calculation in the regions

$$D_i=\{0\leqslant r<a, \quad 0\leqslant\vartheta\leqslant 2\pi, \quad -\infty<z<\infty\},$$

$$D_e=\{a<r<\infty, \quad 0\leqslant\vartheta\leqslant 2\pi, \quad -\infty<z<\infty\},$$

which have a common boundary Γ in the form of a cylindrical surface of radius a, on which, of course, the normal component of the induction and the field potential are continuous. When the first approximation is calculated in D_i, we assume that

$$z=a: \quad \frac{\partial\psi}{\partial r}=0, \quad \frac{\partial\psi_1}{\partial r}=0, \quad \frac{\partial\psi_2}{\partial r}=0. \tag{14}$$

It should be noted that these boundary conditions correspond to the case of ideal conductivity of the medium in D_e, into which the magnetic field introduced in D_i at $t=0$ does not manage to penetrate within the characteristic time for the development of waves on the free surface. Such a schematic approach is widely used in problems of magnetic hydrodynamics.

A calculation of the functions ψ, ψ_1, and ψ_2 in D_i using (14) gives the values of $\psi|_\Gamma$, $\psi_1|_\Gamma$, and $\psi_2|_\Gamma$ on the D_i side, which are then employed as the boundary conditions at Γ for calculating ψ, ψ_1, and ψ_2 in D_e. After finding these functions in D_e, we can calculate the values of $\partial\psi/\partial r$, $\partial\psi_1/\partial r$, and $\partial\psi_2/\partial r$ at Γ on the D_e side and use them to find the second iteration in D_i. This process is then repeated.

Within the proposed approach the first step on the way to calculating the functions sought in D_i can be implemented using the Fourier method. We set

$$\varphi(r,\vartheta,z,t)=\Phi(r,\vartheta)Y(z)S(t),$$

$$\psi(r,\vartheta,z,t)=\psi(r,\vartheta)Z(z)T(t), \tag{15}$$

$$\psi_j(r,\vartheta,z,t)=\psi_j(r,\vartheta)Z_j(z)T_j(t); \quad j=1,2.$$

After substituting expressions (15) into Laplace equations (3), (5), and (6) and separating the variables, we have

$$\nabla_2^2\Phi+k^2\Phi=0, \quad \nabla_2^2\psi+k_0^2\psi=0,$$

$$\nabla_2^2\psi_j+k_j^2\psi_j=0; \quad j=1,2, \tag{16}$$

$$Y''-k^2Y=0, \quad Z''-k_0^2Z=0,$$

$$Z_j''-k_j^2Z_j=0; \quad j=1,2. \tag{17}$$

With consideration of the condition that the perturbations vanish as $|z|\to\infty$, from Eqs. (17) we find

$$Y=a_1\exp(kz)+a_2\exp(-kz),\quad Z_1=c_1\exp(-k_1z),$$

$$Z=b_1\exp(k_0z)+b_2\exp(-k_0z),\quad Z_2=c_2\exp(k_2z).$$

As a result of the substitution of expressions (15) into boundary conditions (9) and (14), we have

$$r=a:\frac{\partial\Phi}{\partial r}=0,\quad \frac{\partial\psi}{\partial r}=0,\quad \frac{\partial\psi_j}{\partial r}=0;\quad j=1,2. \tag{18}$$

The solution of the original initial-boundary problem can be constructed in the form of series in the eigenfunctions of the two-dimensional Helmholtz equations (16) in analogy to the case of the vibrations of a round membrane. It should, however, be noted that, in contrast to the case of a membrane clamped along its edge,[6] the eigenvalues k^2, k_0^2, k_1^2, and k_2^2 of problems (16) and (18) can be expressed in terms of the roots $\varkappa_{n,m}$, $n=0,1,2,\ldots$; $m=1,2,\ldots$ of the equations $J_n'(x)=0$, where $J_n(x)$ is a Bessel function of the first kind:

$$k_{n,m}^2=\left(\frac{\varkappa_{n,m}}{a}\right)^2;\quad n=0,1,2,\ldots;\quad m=1,2,\ldots. \tag{19}$$

The set of roots $\varkappa_{n,m}$ is denoted by Q. Each eigenvalue (19) corresponds to two eigenfunctions

$$v_{1,n,m}(r,\vartheta)=J_n(k_{n,m}r)\cos n\vartheta,$$

$$v_{2,n,m}(r,\vartheta)=J_n(k_{n,m}r)\sin n\vartheta.$$

Assuming that the initial perturbation of the free surface (12) at $r=a$ satisfies the condition $\partial f/\partial r=0$, we expand $f(r,\vartheta)$ in a series in the eigenfunctions:

$$f(r,\vartheta)=\sum_{n,m}F_{n,m}(r,\vartheta),$$

where

$$F_{n,m}(r,\vartheta)=A_{n,m}v_{1,n,m}(r,\vartheta)+B_{n,m}v_{2,n,m}(r,\vartheta),$$

$$A_{n,m}=\|v_{1,n,m}\|^{-2}\int_0^{2\pi}\int_0^a f(r,\vartheta)v_{1,n,m}(r,\vartheta)r\,dr\,d\vartheta,$$

$$B_{n,m}=\|v_{2,n,m}\|^{-2}\int_0^{2\pi}\int_0^a f(r,\vartheta)v_{2,n,m}(r,\vartheta)r\,dr\,d\vartheta,$$

$$\|v_{2,n,m}\|^2=\|v_{1,n,m}\|^2=\int_0^{2\pi}\int_0^a v_{1,n,m}^2(r,\vartheta)r\,dr\,d\vartheta$$

$$=\frac{\pi a^2}{2}\varepsilon_n\left(1-\frac{n^2}{\varkappa_{n,m}^2}\right)J_n^2(\varkappa_{n,m}),$$

$$\varepsilon_n=\begin{cases}2 & \text{for } n=0,\\ 1 & \text{for } n\neq 0.\end{cases} \tag{20}$$

Taking into account the initial conditions (12), expansion (20), and the form of the solutions of equations (17), we set

$$\eta(r,\vartheta,t)=\sum_{n,m}F_{n,m}(r,\vartheta)\cos(\omega_{n,m}t),$$

$$\varphi(r,\vartheta,z,t)=\sum_{n,m}F_{n,m}(r,\vartheta)[a_{1,n,m}\exp(k_{n,m}z)$$
$$+a_{2,n,m}\exp(-k_{n,m}z)]\sin(\omega_{n,m}t),$$

$$\psi(r,\vartheta,z,t)=\sum_{n,m}F_{n,m}(r,\vartheta)[b_{1,n,m}\exp(k_{n,m}z)$$
$$+b_{2,n,m}\exp(-k_{n,m}z)]\cos(\omega_{n,m}t),$$

$$\psi_1(r,\vartheta,z,t)=\sum_{n,m}F_{n,m}(r,\vartheta)c_{1,n,m}$$
$$\times\exp(-k_{n,m}z)\cos(\omega_{n,m}t),$$

$$\psi_2(r,\vartheta,z,t)=\sum_{n,m}F_{n,m}(r,\vartheta)c_{2,n,m}$$
$$\times\exp(k_{n,m}z)\cos(\omega_{n,m}t). \tag{21}$$

Here $a_{j,n,m}$, $b_{j,n,m}$, and $c_{j,n,m}$ $(j=1,2)$ are arbitrary constants, and $\omega_{n,m}$ denotes the permissible frequencies of the standing waves, which are to be determined in the course of solving the problem. After substituting expressions (21) into the conditions (7), (10), and (11) for matching of the functions on the interfaces between the media, we arrive at a system of algebraic equations with respect to the constants:

$$a_{1,n,m}-a_{2,n,m}=-\frac{\omega_{n,m}}{k_{n,m}},$$

$$\exp(-k_{n,m}d)a_{1,n,m}-\exp(k_{n,m}d)a_{2,n,m}=0,$$

$$b_{1,n,m}+b_{2,n,m}-c_{1,n,m}=M_0,$$

$$\sigma\mu_t^0(b_{1,n,m}-b_{2,n,m})+\mu_0c_{1,n,m}=0,$$

$$\exp(-\sigma k_{n,m}d)b_{1,n,m}+\exp(\sigma k_{n,m}d)b_{2,n,m}$$
$$-\exp(-k_{n,m}d)c_{2,n,m}=0,$$

$$\sigma\mu_t^0[\exp(-\sigma k_{n,m},d)b_{1,n,m}-\exp(\sigma k_{n,m}d)b_{2,n,m}]$$
$$-\mu_0\exp(-k_{n,m}d)c_{2,n,m}=0.$$

Hence we find

$$a_{1,n,m}=-\frac{\omega_{n,m}}{2k_{n,m}}\frac{\exp(k_{n,m}d)}{\sinh(k_{n,m}d)},$$

$$a_{2,n,m}=-\frac{\omega_{n,m}}{2k_{n,m}}\frac{\exp(-k_{n,m}d)}{\sinh(k_{n,m}d)},$$

$$b_{1,n,m}=\frac{\mu_0}{2\tau_{n,m}}M_0(\mu_0+\sigma\mu_t^0)\exp(\sigma k_{n,m}d),$$

$$b_{2,n,m}=-\frac{\mu_0}{2\tau_{n,m}}M_0(\mu_0-\sigma\mu_t^0)\exp(-\sigma k_{n,m}d),$$

$$c_{1,n,m}=-\frac{\sigma}{\tau_{n,m}}\mu_t^0 M_0[\mu_0\ \cosh(\sigma k_{n,m}d)+\sigma\mu_t^0\ \sinh(\sigma k_{n,m}d)],$$

$$c_{2,n,m}=\frac{\mu_0\sigma}{\tau_{n,m}}\mu_t^0 M_0\ \exp(\sigma k_{n,m}d),$$

$$\tau_{n,m}=\sinh(\sigma k_{n,m}d)[\mu_0^2+(\sigma\mu_t^0)^2]+2\sigma\mu_0\mu_t^0\ \cosh(\sigma k_{n,m}d).$$

When these equalities are taken into account, the expressions for potentials (21) take the form

$$\varphi(r,\vartheta,z,t)=-\sum_{n,m}\frac{\omega_{n,m}}{k_{n,m}}\frac{\cosh[k_{n,m}(z+d)]}{\sinh(k_{n,m}d)}F_{n,m}(r,\vartheta)\sin(\omega_{n,m}t),$$

$$\psi(r,\vartheta,z,t)=\mu_0 M_0\sum_{n,m}\frac{\mu_0\ \sinh[\sigma k_{n,m}(z+d)]+\sigma\mu_t^0\ \cosh[\sigma k_{n,m}(z+d)]}{\sinh(\sigma k_{n,m}d)[\mu_0^2+(\sigma\mu_t^0)^2]+2\sigma\mu_0\mu_t^0\ \cosh(\sigma k_{n,m}d)}F_{n,m}(r,\vartheta)\cos(\omega_{n,m}t')$$

$$\psi_1(r,\vartheta,z,t)=-\sigma\mu_t^0 M_0\sum_{n,m}\frac{\sigma\mu_t^0\ \sinh(\sigma k_{n,m}d)+\mu_0\ \cosh(\sigma k_{n,m}d)}{\sinh(\sigma k_{n,m}d)[\mu_0^2+(\sigma\mu_t^0)^2]+2\sigma\mu_0\mu_t^0\ \cosh(\sigma k_{n,m}d)}F_{n,m}(r,\vartheta)\exp(-k_{n,m}z)\cos(\omega_{n,m}t),$$

$$\psi_2(r,\vartheta,z,t)=\sigma\mu_0\mu_t^0 M_0\sum_{n,m}\frac{F_{n,m}(r,\vartheta)\exp[k_{n,m}(z+\sigma d)\cos(\omega_{n,m}t)]}{\sinh(\sigma k_{n,m}d)[\mu_0^2+(\sigma\mu_t^0)^2]+2\sigma\mu_0\mu_t^0\ \cosh(\sigma k_{n,m}d)}.$$

When the series representing φ and ψ are substituted into the transformed dynamical condition (13), which has not yet been used, we obtain a dispersion relation, which can be utilized to calculate the permissible frequencies:

$$\omega_{n,m}^2=\tanh(k_{n,m}d)\left\{fk_{n,m}-\frac{\sigma k_{n,m}^2}{\rho}\times\frac{\mu_0\mu_t^0 M_2^2[\mu_0+\sigma\mu_t^0\ \tanh(\sigma k_{n,m}d)]}{2\sigma\mu_0\mu_t^0+\tanh(\sigma k_{n,m}d)[\mu_0^2+(\sigma\mu_t^0)^2]}+\frac{\alpha}{\rho}k_{n,m}^3\right\}. \tag{22}$$

Thus, when all the defining parameters of the problem, except M_0, are fixed as functions of M_0, each of the permissible frequencies $\omega_{n,m}$ is either a real or a purely imaginary quantity. In the former case the flat free surface is stable, and in the latter case it is unstable. It follows from (22) with consideration of (19) that the critical value of the magnetization of the liquid M_*, above which the onset of instability occurs, is specified by the formula

$$M_*=\min_{\varkappa_{n,m}\in Q}\left\{\frac{\rho g a}{\mu_0\mu_t^0}\frac{1+B_0^{-1}\varkappa_{n,m}^2}{\sigma\varkappa_{n,m}}\times\frac{2\sigma\mu_0\mu_t^0+\tanh(s\sigma\varkappa_{n,m})[\mu_0^2+(\sigma\mu_t^0)^2]}{\mu_0+\sigma\mu_t^0\ \tanh(s\sigma\varkappa_{n,m})}\right\}^{1/2},$$

$$s=\frac{d}{a}.$$

It is not difficult to see that the countable set Q of roots, of which a fairly representative group was presented in Ref. 7, always includes a single element which minimizes the right-hand side of this expression. In the case of a liquid with a linear magnetization law, we have

$$\sigma=1,\quad \mu_t^0=\mu,\quad M_0=H_\infty(1-\mu_r^{-1}),\quad \mu_r=\mu/\mu_0,$$

so that the critical value of the external magnetic field H_∞ is specified by the expression

$$H_\infty^*=\min_{\varkappa_{n,m}\in Q}\left[\frac{\rho g a}{\mu_0}\frac{\mu_r(1+B_0^{-1}\varkappa_{n,m}^2)}{\varkappa_{n,m}(\mu_r-1)^2}\times\frac{2\mu_r+(1+\mu_r^2\ \tanh)(s\varkappa_{n,m})}{1+\mu_r\ \tanh(s\varkappa_{n,m})}\right]^{1/2}. \tag{23}$$

This formula, which, of course, does not claim to be a quantitative definition of the critical field, can be used to estimate it.

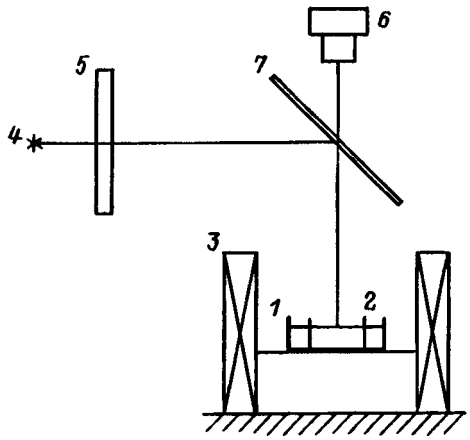

FIG. 2.

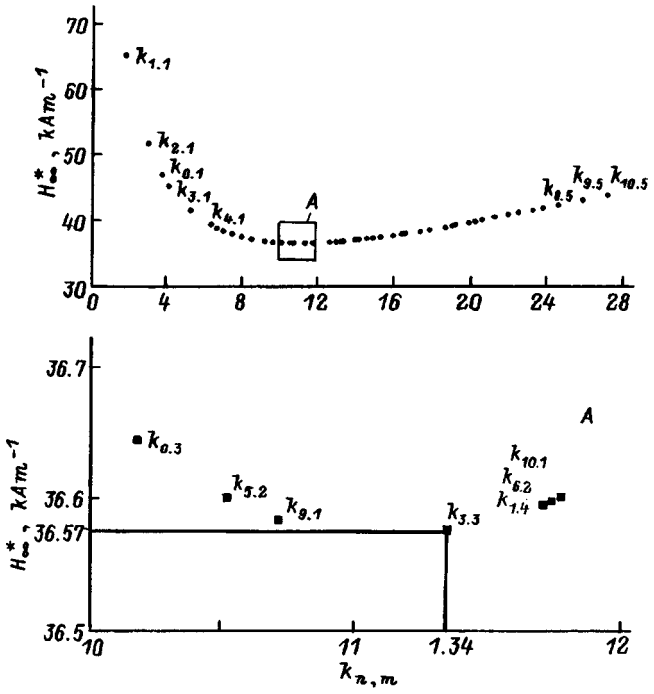

FIG. 3.

EXPERIMENT

An experiment was carried out in the context of the problem considered here for the purpose of comparing the theoretical representation of $\eta(r,\vartheta,t)$ expressed by the first equality in (21) with the physically observed evolution of the shape of the free surface of a magnetic liquid as $\mathbf{H}_\infty$ is increased.

The experimental setup is shown schematically in Fig. 2. A cylindrical cell containing a magnetic liquid *1*, into which a thin-walled nonmagnetic cylindrical insert *2* was lowered coaxially to the cell to the entire depth of the liquid, was placed in solenoid *3*. The surface of the magnetic liquid was illuminated by a beam of light rays created by light source *4* and diffuser *5* and was photographed by camera *6*. Axial illumination in a scheme utilizing a beam splitter,[8] which is labeled by the number *7* in Fig. 2, was employed.

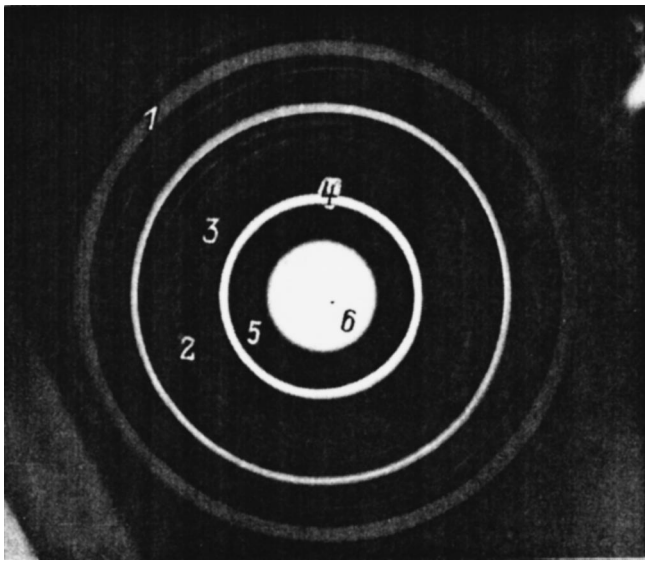

FIG. 5.

A magnetic liquid for which $\rho=992$ kg/m^3, $\alpha=2.7\times10^{-2}$ N/m, and $\mu_r=1.23$ was employed in the experiment; the radius of the insert was $a=1.85\times10^{-2}$ m, the radius of the cell was 3.3×10^{-2} m, and the depth of the layer of the liquid was $d=6\times10^{-3}$ m. The Bond number for these values of the defining parameters is $B_0=123$, and the critical field calculated from formula (23) is $H_\infty^*=36.57$ kA/m. As can be seen from Fig. 3, the minimizing element in the set of roots is then $\varkappa_{3,3}=11.3459$. All the values of $\varkappa_{n,m}$ for which $0\leqslant n\leqslant10$ and $1\leqslant m\leqslant5$ were used to construct the plot in Fig. 3.

Figure 4 presents a downward view of the surface of the liquid in the absence of a magnetic field. In this figure *1* is the wall of the cell, *2* is the wall of the cylindrical insert, *3* is

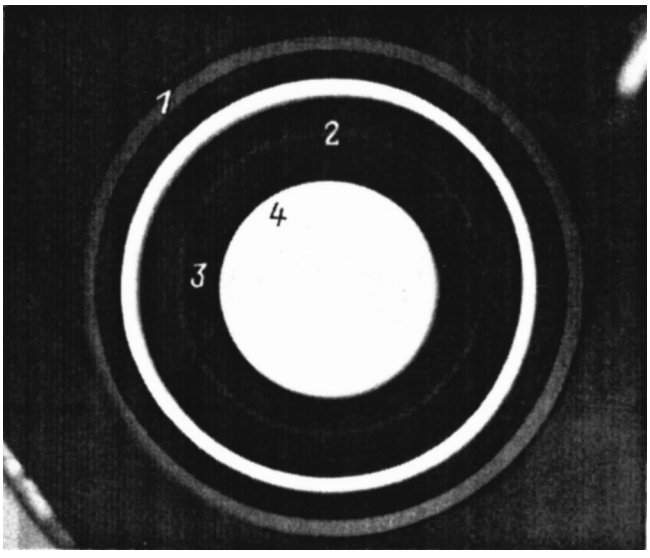

FIG. 4.

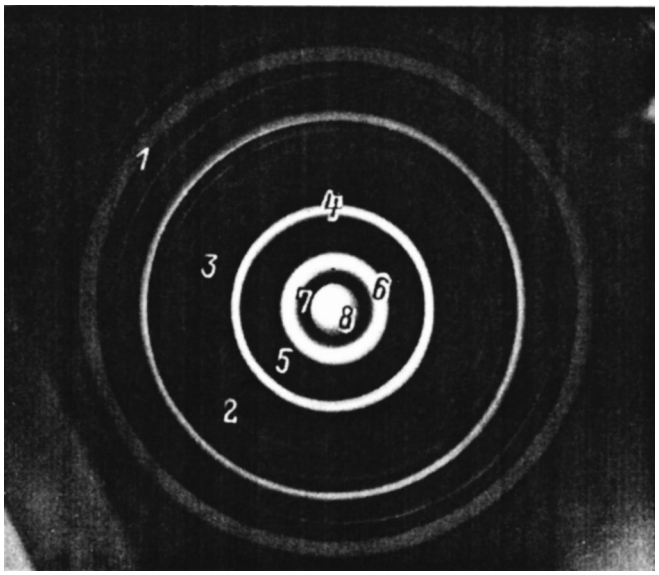

FIG. 6.

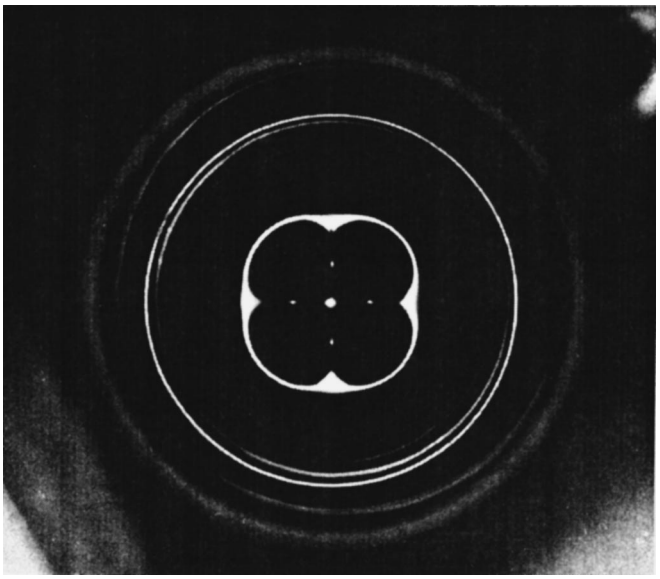

FIG. 7.

the capillary meniscus, and *4* (the bright circular area) is the flat part of the free surface. The bright ring between the wall of the cell *1* and the wall of the insert *2* corresponds to the depression on the surface of the liquid between the menisci, i.e., the dark rings near the walls.

Figure 5 shows the free surface for $H_\infty = 9.31$ kA/m. The number *4* in it marks the relatively flat (in comparison to the meniscus next to the wall of the insert) part of the free surface, which has the form of a bright ring. It is followed in the direction leading to the center of the cell by a rise *5* of the free surface caused by the magnetic field (the dark ring), which gives way to relatively flatter central portion *6* (the bright spot). Thus, in a magnetic field the free surface took on a domed shape rising with distance from the wall of the insert. It should be stressed that the structure shown in Fig. 5 is the result of the nonlinear stage of development of the instability.

Figure 6 shows the free surface for $H_\infty = 11.78$ kA/m. A comparison of Figs. 5 and 6 reveals that increasing the magnetic field causes a depression to form in the central portion of the dome that had been observed in the lower field $H_\infty = 9.31$ kA/m. The incline *7* (the dark ring) and the relatively flat bottom of the depression *8* (the central bright spot) are labeled in Fig. 6. As a whole, the surface structure that arises has the form of an annular corrugation.

A further increase in the field leads to the formation of more complicated surface structures, whose shape depends on the azimuthal angle. Figure 7 presents the free surface for $H_\infty = 13.77$ kA/m as an example. For technical reasons this was the highest value of the magnetic field strength used. As before, in this figure the bright regions correspond to relatively flat portions of the free surface.

The structures presented in Figs. 5 and 6 are qualitatively similar to the first (Fig. 8a) and second (Fig. 8b) modes of the representation (21) of the free surface, which correspond to the roots $\varkappa_{0,1} = 3.8317$ and $\varkappa_{0,2} = 7.0156$. In the former case the critical field (for the respective mode) equals 46.75 kA/m, and in the latter case it equals 38.5 kA/m. The

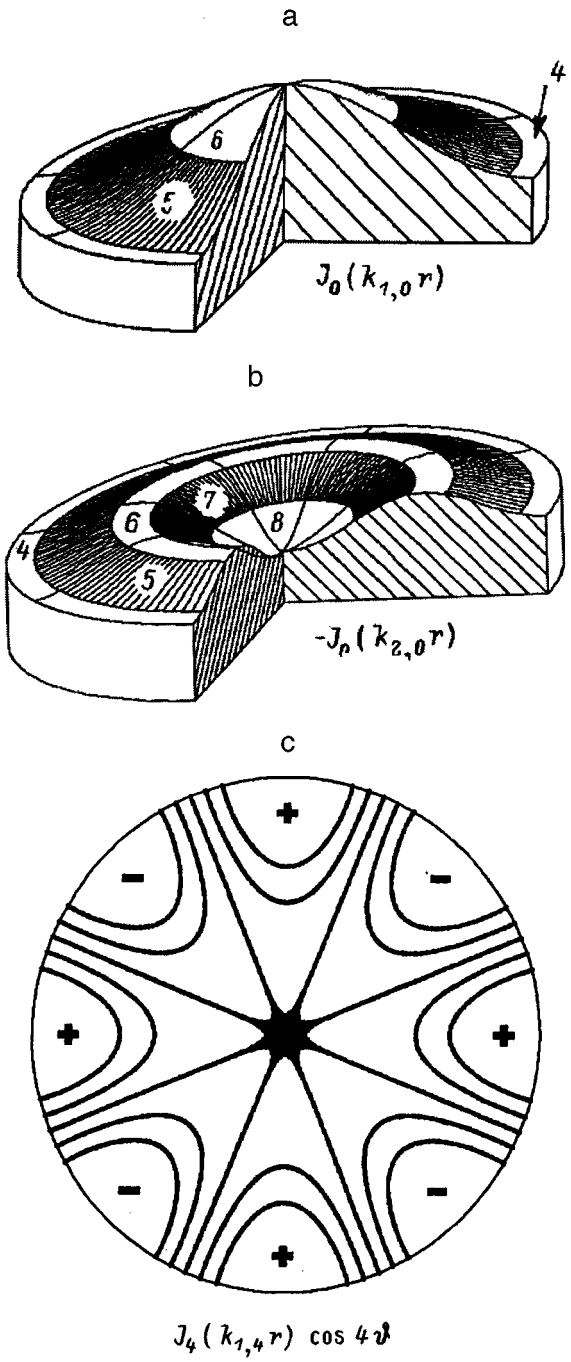

FIG. 8.

structure shown in Fig. 7 exhibits some qualitative similarity to the mode containing $J_4(k_{4,1}r)\cos 4\vartheta$ and $J_4(k_{4,1}r)\sin 4\vartheta$, for which $\varkappa_{4,1} = 5.3174$ and $H_\infty^* = 41.49$ kA/m. Level lines of this mode are shown in Fig. 8c, in which elevated areas are marked by plus signs and depressions are marked by minus signs, the latter being separated from one another by radial nodal lines.

CONCLUSIONS

It follows from the approximate solution of the problem of the stability of an initially flat free surface of a magnetic liquid filling a vertical cylindrical cavity in a flat horizonal layer of a magnet under the action of a vertical magnetic

field that the system considered has axisymmetric unstable modes. The surface structures of the ''dome'' and ''annular corrugation'' types observed in the experiment performed qualitatively confirm the results of the theoretical analysis.

[1] M. D. Cowley and R. E. Rosensweig, J. Fluid Mech. **30**, 671 (1967).
[2] R. E. Rosensweig, *Ferrohydrodynamics*, Cambridge University Press [Cambridge–New York (1985); Mir, Moscow (1989)].
[3] B. M. Berkovsky, V. F. Medvedev, and M. S. Krakov, *Magnetic Fluids: Engineering Applications*, Oxford University Press, Oxford–New York (1993).
[4] É. Ya. Blum, M. M. Maĭorov, and A. O. Tsebers, *Magnetic Fluids* [in Russian], Zinatne, Riga (1989).
[5] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed., Pergamon Press, Oxford (1987).
[6] A. N. Tikhonov and A. A. Samarskii, *Equations of Mathematical Physics*, Pergamon Press, Oxford (1964).
[7] *Tables of Zeros of Bessel Functions. Library of Mathematical Tables, No. 44* [in Russian], VTs Akad. Nauk SSSR, Moscow (1967).
[8] A. A. Blaker, *Handbook for Scientific Photography* [W. H. Freeman, San Francisco (1977); Mir, Moscow (1980)].

Translated by P. Shelnitz

Onset of turbulence in open liquid flows as a nonequilibrium noise-induced second-order phase transition

P. S. Landa

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted June 10, 1996)
Zh. Tekh. Fiz. **68**, 31–39 (January 1998)

It is shown that there is a profound analogy between the transition to turbulence in open liquid flows and the noise-induced excitation of oscillations of a pendulum with a randomly oscillating pivot. It is significant that this analogy is based not on the similarity of the equations describing these processes, but on the generality of the laws of the theory of oscillations. The existence of this analogy makes it possible to understand and account for numerous phenomena observed in both numerical simulations and real experiments. Moreover, this analogy suggests several recommendations to experimenters for achieving a more thorough suppression of undesirable turbulent pulsations in subsonic jets. © *1998 American Institute of Physics.* [S1063-7842(98)00601-1]

INTRODUCTION

Liquid flows in channels are laminar at low flow rates and become turbulent at high flow rates.[1–3] The problem of the nature of turbulence has attracted the attention of investigators for a long time. If the nature of turbulence is discussed from the standpoint of the theory of oscillations, most scientists would classify turbulence as an auto-oscillatory process without always thinking over this question, since they employ investigative methods that are suitable specifically to such processes. The foundation for such an approach was laid in the work of Landau,[4] according to whose ideas turbulence appears in the following manner. At first, the equilibrium state corresponding to laminar flow becomes unstable, and auto-oscillations are excited at one frequency. Landau wrote a phenomenological equation for the amplitude of these auto-oscillations, which has the form of the well-known abridged van der Pol equation. ''As the Reynolds number increases further,'' wrote Landau, ''new periods continue to appear in succession. As to the newly appearing motions themselves, they have increasingly smaller scales.'' As a result, according to Landau's hypothesis, multiple-frequency auto-oscillations with incommensurate frequencies, i.e., quasiperiodic motion, are established. In phase space such auto-oscillations should correspond to an attractor in the form of a multidimensional torus. When the number of frequencies is large, such motion differs only slightly in form from chaotic motion, and, therefore, developed turbulence can be regarded as a random process. Despite the fact that Landau's theory is phenomenological and, in general, does not follow from the equations of hydrodynamics, it has not been challenged for a long time and has been confirmed by almost all investigations of turbulence. Landau's theory was further developed by Stuart,[5–8] who proposed a method for calculating the coefficients appearing in Landau's equation on the basis of an approximate solution of the Navier–Stokes equation. Stuart thereby ''substantiated'' Landau's theory. However, the form of the approximate solution assigned by Stuart, viz., $A(\varepsilon t)\exp\{i(\omega t-kx)\}$, is incorrect from the physical standpoint. In fact, this solution describes a wave which is periodic in space and has an assigned wave number k and an amplitude that varies slowly with time. Strictly speaking, such a solution is valid only for an annular flow of length $L=2\pi n/k$, where n is an integer, i.e., for a flow with feedback. The solution assigned by Stuart does not take into account the convective character of the instability of laminar flow. Nevertheless, even recently, periodic boundary conditions along the longitudinal coordinate have been assigned quite often in numerical studies of turbulent flows (see, for example, Ref. 9).

The opinion that turbulence can be regarded as auto-oscillations in a continuous medium, i.e., in a system with an extremely large number of degrees of freedom, was also held by G. S. Gorelik, as is known from Rytov's recollections.[10]

In the nineteen-seventies, an opinion that regards the development of turbulence as the instantaneous appearance of a strange attractor in a phase space of several dynamical variables became popular following the discovery of dynamical chaos.[11,12] These ideas were presented in the latest editions of *Fluid Mechanics* by L. D. Landau and E. M. Lifshitz[1] and *Statistical Fluid Mechanics* by A. S. Monin and A. M. Yaglom.[2] Since the concept of strange attractors generally refers only to auto-oscillations, it was tacitly assumed in these books that turbulence is specifically an auto-oscillatory process. On the basis of a representation of turbulence as auto-oscillations, a group of investigators published a series of papers on the simulation of the development of turbulence in the form of an infinite chain of unidirectionally coupled oscillators[13,14] and on the use of such characteristics as the correlation dimension of an attractor for a quantitative description of turbulent flows.[15]

It was theorized in Refs. 16–18 that the turbulence appearing in open liquid flows does not consist of auto-oscillations, and all the approaches described above are, therefore, not applicable to it. This hypothesis was based on the fact that the instability of the solutions corresponding to laminar flow in such open flows is convective. This means that a disturbance appearing at a certain point in a flow will not grow with time without bound (in the linear approximation) but will drift downstream. It follows from the proper-

ties of systems with convective instability that such systems, in themselves, are not auto-oscillatory but are only amplifiers of disturbances. To make such a system auto-oscillatory, we must introduce feedback, for example, by closing the system in the form of a ring. Disturbances are unavoidably present in all real systems due to both external factors (so-called technical fluctuations) and internal factors (natural fluctuations). These disturbances can be included in the equation of motion of a system as additional random forces, whose strength in the general case depends on the variables describing the state of the system. The forces created by natural fluctuations in hydrodynamic flows on the basis of the fluctuation-dissipation theorem were calculated by Klimontovich.[19]

If the gain of an amplifier is sufficiently small, the presence of fluctuations can be neglected, and it can be assumed that the output signal of the amplifier derives solely from the input signal. In hydrodynamic flows the gain is generally fairly large. In this case the presence of fluctuations is of fundamental significance, since, in our opinion, they determine the turbulent disturbances observed. Hence it follows that an approach to the investigation of turbulence (or of ordinary amplifiers with a large gain) within the theory of dynamical systems will be inadequate if fluctuations are not taken into account. The amplified fluctuations and nonlinearity in a system can give rise to a phase transition in which the system passes into a qualitatively new state. It can be assumed that the onset of turbulence characterized by the presence of large-scale, highly regular structures against a background of small-scale random motions corresponds precisely to such a transition. In our opinion, application of the theory of noise-induced phase transitions to the investigation of turbulence may be very fruitful.

NIKITIN'S NUMERICAL SIMULATION AND ITS INTERPRETATION FROM THE STANDPOINT OF NOISE-INDUCED PHASE TRANSITIONS

Indirect evidence that the turbulence in open flows is not auto-oscillatory is provided by Nikitin's numerical simulation[20] of turbulent flow in pipes of finite length. He investigated flow in a round pipe of radius R with an assigned velocity in the entrance cross section and so-called ''soft'' boundary conditions in the exit cross section, which were of the form

$$\frac{\partial^2 u}{\partial x^2}=\frac{\partial^2 \xi}{\partial x^2}=\frac{\partial^2 \eta}{\partial x^2}.$$

Here u is the longitudinal component of the flow velocity, and ξ and η are the radial and angular components of the vorticity $\mathbf{\Omega}=\mathrm{curl}\mathbf{V}$, where $\mathbf{V}$ is the flow velocity vector in the cylindrical coordinates x, r, Θ.

The velocity components in the entrance cross section of the pipe were assigned in the form

$$u=u_0\left(1-\frac{r^2}{R^2}\right)+A\mathrm{Re}(u'(r)e^{-i\omega t})\cos\Theta,$$

$$v=A\mathrm{Re}(v'(r)e^{-i\omega t})\cos\Theta,$$

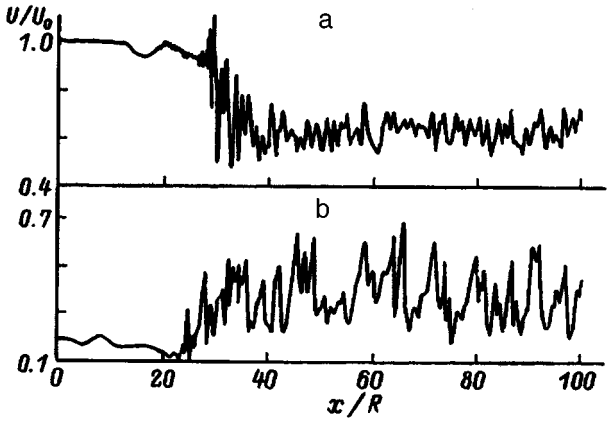

FIG. 1. Instantaneous distributions of the longitudinal velocity component in a steady-state regime: a—near the axis of the pipe ($r/R=0.02$), b—near the wall ($r/R=0.93$).

$$w=A\mathrm{Re}(w'(r)e^{-i\omega t})\sin\Theta,$$

where v and w are the radial and angular components of the flow velocity, respectively; $u'(r)$, $v'(r)$, and $w'(r)$ are eigenfunctions of the Orr–Sommerfeld equation for an assigned real value of the frequency ω; R is the radius of the pipe; and A and ω are the amplitude and frequency of the disturbance.

The frequency of the disturbance was chosen as $\omega=0.36u_0/R$, and the velocity u_0 and the pipe radius R corresponded to a Reynolds number of 4000. At the initial time a Poiseuille velocity profile was assigned throughout the flow, i.e.,

$$\mathbf{V}|_{t=0}=\left\{u_0\left(1-\frac{r^2}{R^2}\right),0,0\right\}.$$

When the amplitude A of the disturbance exceeds a certain critical value, random high-frequency pulsations that cover the entire lower portion of the pipe, beginning at a certain value $x=x_0$, which depends weakly on the distance r from the axis of the pipe, appear in the flow after a short time interval. The value of x_0 is smaller, the larger is the amplitude A of the disturbance. The appearance of turbulent pulsations is accompanied by significant alternation of the profile of the longitudinal component of the mean flow velocity: it decreases on the axis of the pipe and increases near the wall. The instantaneous distributions of the longitudinal component of the velocity in a steady-state regime for $A/u_0=0.04$ are shown in Fig. 1, which was borrowed from Ref. 20. If the amplitude A of the periodic disturbance is gradually decreased, below a certain value the turbulent region is carried along by the flow, and the flow in the tube becomes laminar. As we know (see, for example, Ref. 21), Poiseuille flow in a round pipe, unlike Poiseuille flow in a flat channel, has the property that laminar flow at any Reynolds number is stable against infinitesimal perturbations. However, in the case of fairly large Reynolds numbers, such flow is unstable with respect to disturbances of finite magnitude. If an attractor corresponding to a turbulent regime would be present in the system in the absence of the disturbance, and the role of the disturbance would be reduced to

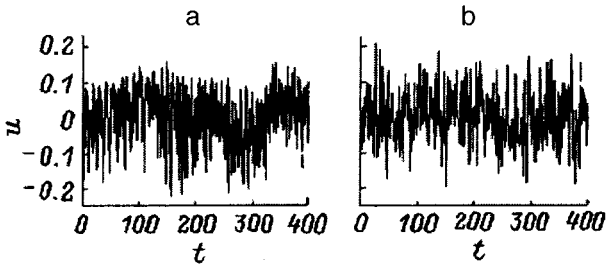

FIG. 2. Form of turbulent velocity pulsations in a pipe with periodic boundary conditions (a) and in a pipe with an assigned harmonic signal at its entrance (b).

only directing phase trajectories into the region of attraction of that attractor, the turbulence should not vanish when the disturbance causing it is removed. Actually, another situation, in which an attractor appears in the presence of an asynchronous disturbance, is also possible, in principle.[22,23] In this case it should vanish when the disturbance is removed. The following arguments against this situation can be cited: first, it follows from the general theory of the asynchronous excitation of auto-oscillations[22,23] that such a situation is possible only in a narrow range of parameters, whereas a transition to turbulence was observed by Nikitin over a broad range of Reynolds numbers; second, the asynchronous excitation of auto-oscillations is equally possible both for positive and negative detuning of the excitation frequency relative to the frequency of the oscillations that arise, while a transition to turbulence was observed only at low excitation frequencies.

It can be postulated that the observed development of turbulence when $A \geqslant A_{cr}$ is attributable to the appearance of a noise-induced phase transition, which leads to the formation of an induced attractor. The similarity of both the external and statistical characteristics of the turbulence appearing in this case to turbulence in a pipe under periodic boundary conditions,[9] where there is feedback and auto-oscillations are excited, supports the latter hypothesis. The outward similarity is demonstrated in Fig. 2, which was constructed on the basis of Nikitin's data. If these hypotheses are correct, the role of the periodic disturbance at the pipe entrance in the development of turbulence reduces to only stimulating a phase transition, as will be shown in the next section in the example of a pendulum.

NOISE-INDUCED PHASE TRANSITION IN A PENDULUM WITH A RANDOMLY OSCILLATING PIVOT

A comparatively simple example of a noise-induced phase transition that leads to the appearance of undamped random oscillations, which are very reminiscent of turbulent pulsations, was considered in our earlier studies.[24,25] The equation describing the oscillations of a pendulum with a randomly oscillating pivot was investigated analytically and numerically. If an additive noise is also taken into account, which was not done in the studies just cited, the equation takes the form

$$\ddot{\varphi} + 2\beta(1+\alpha\dot{\varphi}^2)\dot{\varphi} + \omega_0^2(1+\xi_1(t))\sin\varphi = k\xi_2(t), \qquad (1)$$

where φ is the angular deviation of the pendulum relative to its equilibrium position; $2\beta(1+\alpha\dot{\varphi}^2)\dot{\varphi}$ is a quantity that is proportional to the moment of the forces of friction, which is assumed to be nonlinear; ω_0 is the eigenfrequency of the small oscillations; $\xi_1(t)$ is the acceleration of the pivot, which is a comparatively broad-band random process with a nonzero spectral density at $2\omega_0$; and $k\xi_2(t)$ is the additive noise, whose intensity can be varied by varying the coefficient k.

Let us first discuss the case of $k=0$. It was shown in Refs. 24 and 25 that when the intensity of the pivot oscillations exceeds a certain critical value, which is proportional to the coefficient of friction β, parametric excitation of pendulum oscillations occurs, which is manifested by the fact that the rms deviation of the rotation angle of the pendulum becomes nonzero. Examples of such oscillations and the dependence of the mean square of the rotation angle of the pendulum on the ratio of the spectral density of the noise $\varkappa(2\omega_0)$ to its critical value, which were obtained by numerically solving Eq. (1), are presented in Fig. 3. As can be seen from the figure, near the excitation threshold the pendulum oscillations have the property of intermittency,[1)] i.e., the pendulum oscillates about its equilibrium position over long time intervals (these are so-called ''laminar'' phases); these segments give way to short spikes (''turbulent'' phases). As we move away from the threshold, the duration of the laminar phases shortens, and then the duration of the turbulent phases increases, and the laminar phases disappear completely as a result. During this process, the rms deviation of the rotation angle increases. We note that turbulence also has the property of intermittency in the region of transitional Reynolds numbers (see, for example, Refs. 9 and 26–29). It is no accident that researchers specializing in turbulence were the first to consider the theory of intermittency in detail[30] and that the terminology (laminar and turbulent phases) was borrowed from the theory of turbulence.

When an additive noise is present, the excitation threshold is obliterated, and the dependence of $\overline{\varphi^2}$ on $\varkappa(2\omega_0)/\varkappa_{cr}(2\omega_0)$ becomes smooth. This is demonstrated in Fig. 4a for the case in which the rms deviation of the additive noise is proportional to that of the multiplicative noise, with a proportionality factor equal to 0.05. The weak additive noise causes intermittency to begin to be observed at intensities of the multiplicative noise that are smaller than the critical value in the absence of the additive noise (Fig. 4b). It is quite surprising that in this case the pendulum oscillations are similar in form to turbulent pulsations in the presence of intermittency. An example of such pulsations in a subsonic jet near a nozzle is shown in Fig. 4c.[2)] We note that if the intensity of the multiplicative noise is sufficiently high, even a very appreciable additive noise has practically no influence on the form of the oscillations excited (compare Figs. 3c and 4c).

If there is no additive noise, and the intensity of the multiplicative noise is less than the threshold value, the excitation of pendulum oscillations can be caused by weak additional low-frequency oscillations of the pivot. These oscillations can be taken into account if we replace ξ_1 by $\xi_1 + a\cos\omega_a t$ in Eq. (1), where a and ω_a are quantities

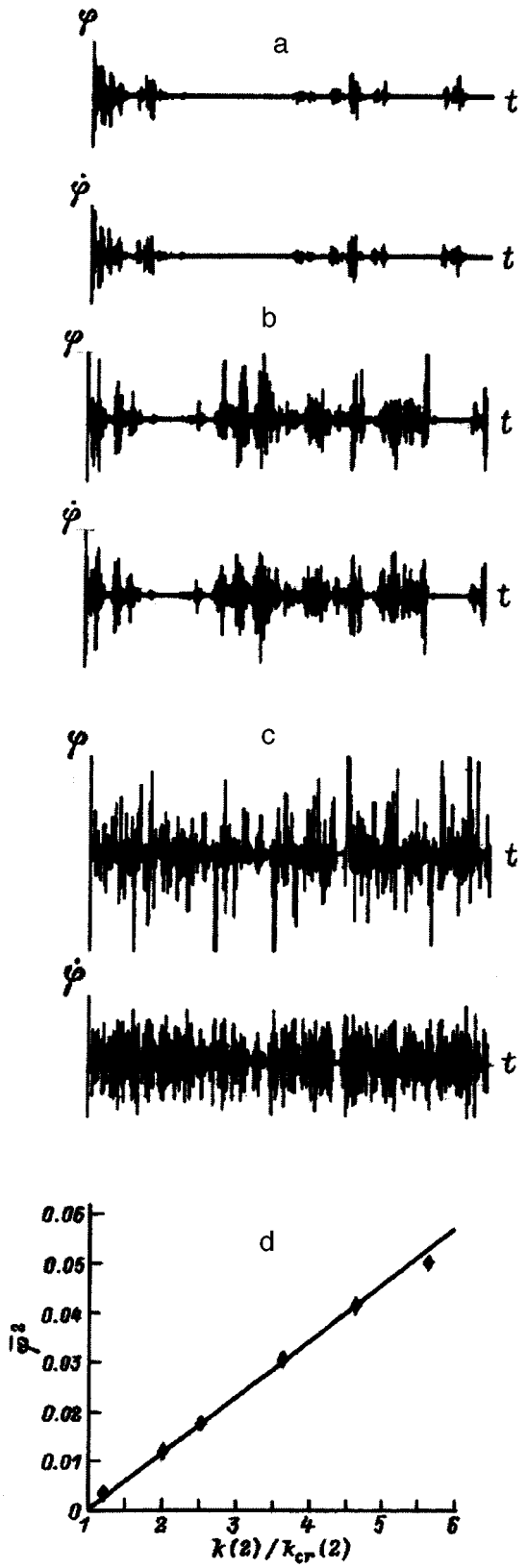

FIG. 3. Plots of $\varphi(t)$ and $\dot{\varphi}(t)$. $\omega_0=1$, $\beta=0.1$, $\alpha=100$, $k=0$; $k(2)/k_{cr}(2)=1.02$ (a), 1.2 (b), and 5.6 (c); d—dependence of $\overline{\varphi^2}$ on $k(2)/k_{cr}(2)$. Solid line—$\overline{\varphi^2}=0.01151(k(2)/k_{cr}(2)-1)$.

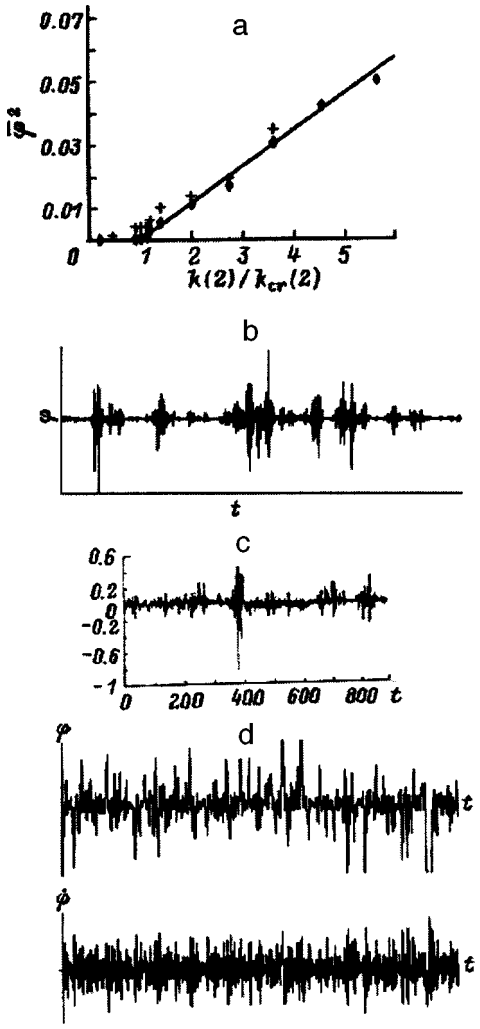

FIG. 4. Dependence of $\overline{\varphi^2}$ on $k(2)/k_{cr}(2)$ without an additive noise (squares) and with an additive noise $k^2\xi_2^2=0.05\overline{\xi_1^2}$ (squares) [solid line—$\overline{\varphi^2}=0.01151(k(2)/k_{cr}(2)-1)$] (a), dependence of $\varphi(t)$ for $\overline{k^2\xi_2^2}=0.000125\overline{\xi_1^2}$ and $k(2)/k_{cr}(2)=0.92$ (b), pulsations of the longitudinal component of the flow velocity on the axis of a round subsonic jet at a distance from the nozzle amounting to 0.1 of the nozzle diameter (c), and plots of $\varphi(t)$ and $\dot{\varphi}(t)$ for $k(2)/k_{cr}(2)=5.6$ and $\overline{k^2\xi_2^2}=0.05\overline{\xi_1^2}$ (d).

which are proportional to the amplitude and frequency of the additional oscillations of the acceleration of the pivot. In the case in which the intensity of the multiplicative noise exceeds the critical value, the additional low-frequency signal increases the intensity of the noise-induced oscillations. The results of the numerical solution of Eq. (1) for two values of $\varkappa(2)/\varkappa_{cr}(2)$ and various values of a are presented in Fig. 5. We see that in the case of $\varkappa(2)<\varkappa_{cr}(2)$ the excitation of oscillations exhibits a threshold character as a is increased. When $\omega_a=0.318$ and $\varkappa(2)/\varkappa_{cr}(2)=0.51$, the threshold value of a is equal to 1.1. The dependence of the rms deviation of the rotation angle on the difference between the amplitude of the low-frequency signal and its critical value is nearly linear (Fig. 5d). When $a>a_{cr}$, the oscillations excited are virtually indistinguishable from the oscillations which appear only as a result of noise, the intensity being higher, the greater is the value of a. This means that the low-frequency signal stimulates the appearance of a phase transition and the creation of an induced attractor.

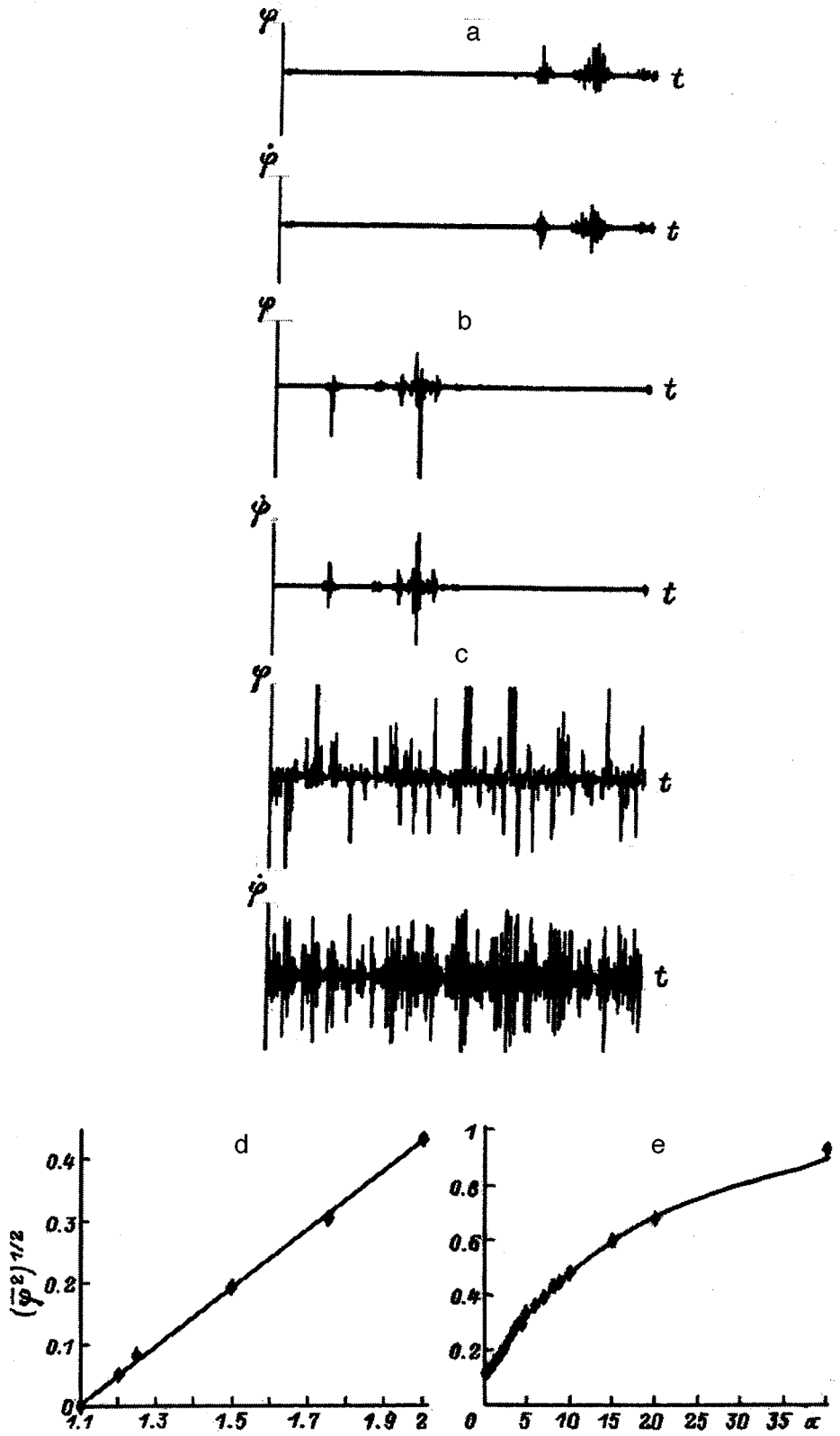

FIG. 5. Plots of $\varphi(t)$ and $\dot{\varphi}(t)$ for $\omega_0=1$, $\beta=0.1$, $\alpha=100$, $k=0$, $k(2)/k_{cr}(2)=0.51$, $\omega_a=0.318$, and $a=1.1$ (a), 1.2 (b), and 1.5 (c); dependence of $(\overline{\varphi^2})^{1/2}$ on a [solid line—$(\overline{\varphi^2})^{1/2}=0.48(a-1.1)$] (d) and dependence of $(\overline{\varphi^2})^{1/2}$ on a for $k(2)/k_{cr}(2)=2.23$ and $\omega_a=1.5$ (e).

NOISE-INDUCED PENDULUM OSCILLATIONS AND TURBULENCE IN JET FLOWS. CONTROLLING THESE PROCESSES

The analogy between turbulent processes and noise-induced pendulum oscillations can also be traced in the case of the development of turbulence and the control of its development in jet flows. Its is known[17,31,32] that large-scale turbulence (coherent structures) appears in jets at a definite distance from the nozzle, mainly in a boundary layer, whose width increases almost linearly as this distance increases. The appearance of coherent structures is accompanied by variation of the mean flow velocity.

It is also known that a weak acoustic signal acting on a jet in the region where it issues from the nozzle can strongly

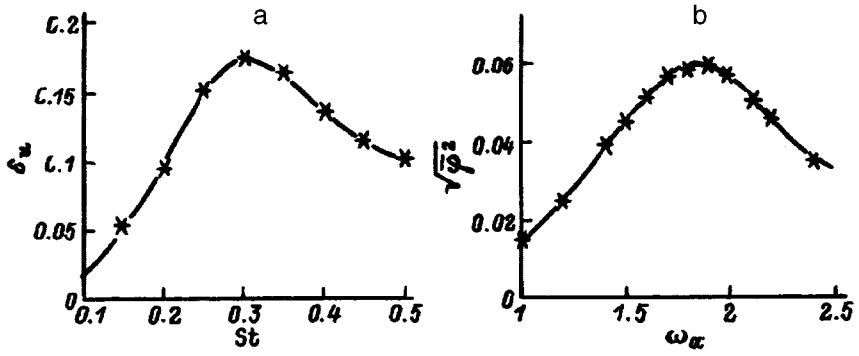

FIG. 6. Dependence of ε_u on St for $\varepsilon_{ua}=0.02$ and $x/D=4$, where x is the distance from the nozzle rim (a), and dependence of $(\overline{\varphi^2})^{1/2}$ on ω_a for $a=0.5$ and $k(2)/k_{cr}(2)=1.01$ (b).

influence the hydrodynamic processes in the jet. If the frequency of this signal f_a lies in a certain region, which can be called the resonance region, it transforms into an intensifying hydrodynamic wave. This is confirmed by the experimental results presented in Ref. 33. It follows from these results that above a certain value of the amplitude of the acoustic signal, the dependence of the rms value of the pulsations of the hydrodynamic velocity on the frequency of the signal has a resonance character. As an example, Fig. 6a presents the plot for $\varepsilon_{u_a}=\sqrt{\overline{u_a^2}}/U_0=0.02$, where u_a is the oscillatory velocity in the acoustic wave and U_0 is the mean flow velocity on the axis of the jet near the nozzle, of the dependence of the rms value of the relative pulsations of the longitudinal component of the hydrodynamic velocity $\varepsilon_u=\sqrt{\overline{u^2}}/U_0$ on the frequency of the acoustic signal expressed in terms of the

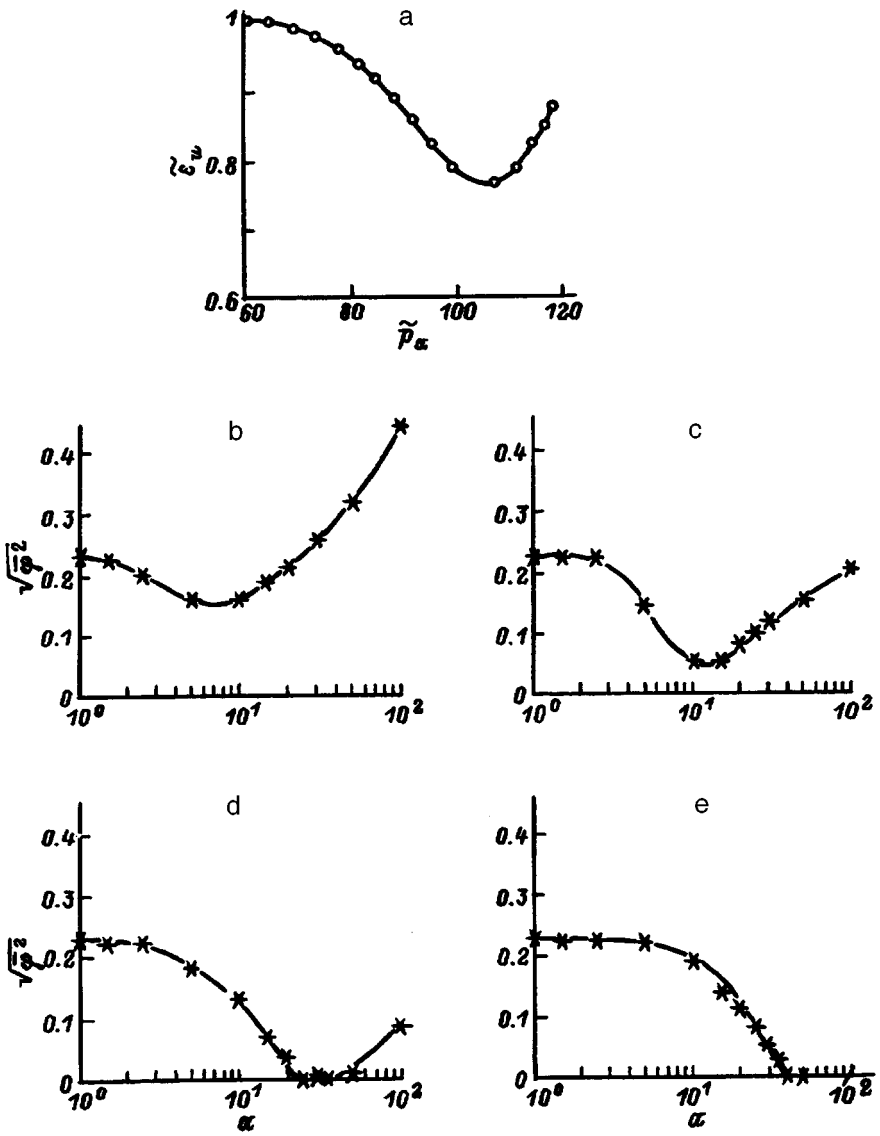

FIG. 7. Experimental dependence of ε_u on the relative amplitude of the acoustic pressure $\tilde{p}_a$ measured in decibels for St=2.35 and $x/D=8$ (a) and dependence of $(\overline{\varphi^2})^{1/2}$ on a for a pendulum when $\omega_0=1$, $\beta=0.1$, $\alpha=100$, $k(2)/k_{cr}(2)=5.6$, and $\omega_a=3.5$ (b), 6 (c), 11 (d), and 19.75 (e).

Strouhal number $\mathrm{St}=f_aD/U_0$, where D is the diameter of the nozzle. We see that ε_u reaches a maximum when $\mathrm{St}\approx 0.3$. A similar dependence is obtained for the rms deviation of the rotation angle of the pendulum under consideration for additional harmonic vibration of the pivot. The corresponding dependence for $a=0.5$ and $\varkappa(2)/\varkappa_{\mathrm{cr}}(2)=1.01$ is presented in Fig. 6b. We note that in the absence of the additional signal $\varepsilon_u\approx 0.04$ and $\sqrt{\overline{\varphi^2}}\approx 0.0317$.

It has been shown in several publications (see, for example, Ref. 17 and 34) that, depending on its frequency, an acoustic signal can either stimulate or suppress the development of large-scale turbulence: stimulation occurs when the frequency of the signal is low, and suppression occurs when it is high. This fact alone suggests that the turbulence in jets is not an auto-oscillatory process. As we know,[22,23] the stimulation or suppression of oscillations in auto-oscillatory systems does not depend on whether the frequency of the acoustic signal is lower or higher than the frequency of the auto-oscillations. These effects are determined only by whether the auto-oscillatory system is a system with hard or soft excitation.

The experimental dependence of the rms magnitude of the pulsations of the longitudinal component of the hydrodynamic velocity ε_u on the acoustic pressure at $\mathrm{St}=2.35$ is presented in Fig. 7a. We see that the turbulent pulsations at first decrease with increasing amplitude of the acoustic signal and then increase.

We have already discussed the stimulation of a noise-induced phase transition in a pendulum in the preceding section. Let us now consider the possibility of suppressing noise-induced oscillations by a high-frequency harmonic signal. Numerical simulation of Eq. (1) with the replacement of ξ_1 by $\xi_1+a\cos\omega_a t$, where $\omega_a>2$, showed that such suppression actually occurs. The results of the simulation are presented in Figs. 7b–7c and 8. We see that if the amplitude of the high-frequency signal is small, it has practically no influence on the existing oscillations. In addition, it is seen from Fig. 7 that in the case of signals with not very high frequencies the intensity of the noise-induced oscillations at first decreases with increasing amplitude, as in the experiment described above, to a certain minimum value, which is smaller, the higher is the frequency of the signal, and then begins to increase. True, the higher the frequency, the higher is the amplitude of the signal at which this minimum value is achieved. When the frequency of the signal is sufficiently high, the oscillations are completely suppressed. Thus, on the basis of the analogy with the oscillations of a pendulum, experimental investigators can be advised to increase the frequency of the acoustic signal to avoid an undesirable increase in the turbulent pulsations as the amplitude of the signal is increased. As far as we know, this advice has been found to hold true.

The suppression of noise-induced oscillations of a pendulum by an additional signal of sufficiently high frequency is demonstrated in greater detail in Fig. 8, which presents plots of the dependence of φ and $\dot{\varphi}$ on the time for various values of the amplitude of a signal whose frequency equals 19.757. It is seen that the intensity of the noise-induced oscillations decreases with increasing amplitude, while the duration of the segments corresponding to ''laminar'' phases increases. When the amplitude exceeds a certain critical value (it was found to be equal to 42 for the case under consideration), the oscillations are completely suppressed. When the amplitude of the signal is increased further, the oscillations reappear, but now because the corresponding parametric resonance conditions begin to be satisfied.

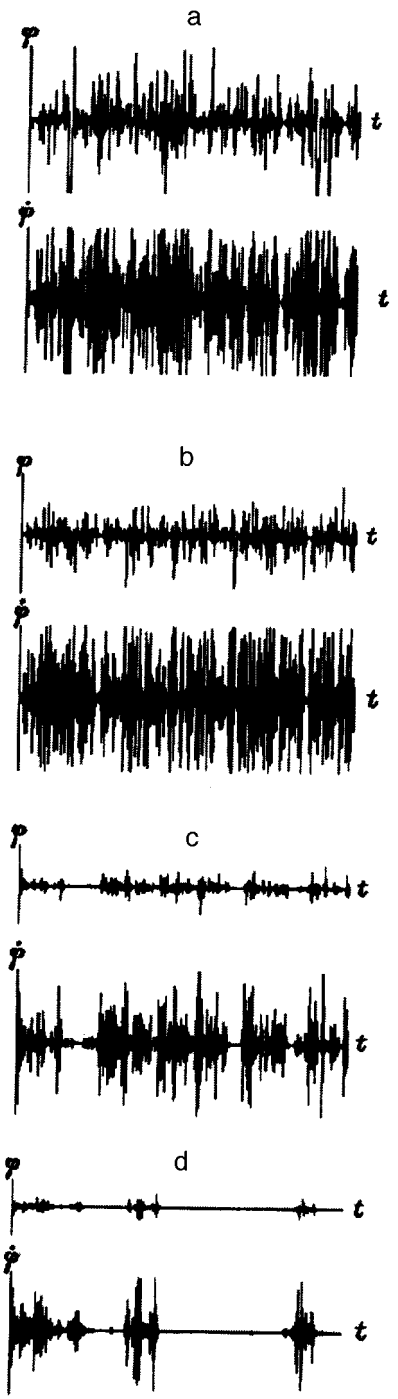

FIG. 8. Plots of $\varphi(t)$ and $\dot{\varphi}(t)$. $\omega_0=1$, $\beta=0.1$, $\alpha=100$, $k(2)/k_{\mathrm{cr}}(2)=5.6$, $k=0$, $\omega_a=19.757$, and $a=5$ (a), 15 (b), 30 (c), and 40 (d).

It should be noted that the stimulation and suppression of noise-induced pendulum oscillations occur not only under a parametric harmonic signal, but also under a force signal,

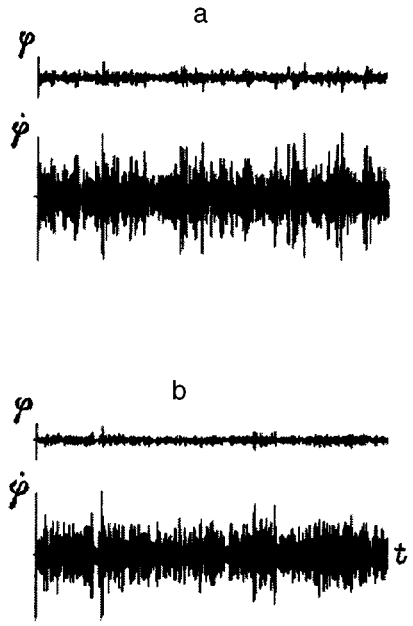

FIG. 9. Plots of $\varphi(t)$ and $\dot{\varphi}(t)$. $\overline{k^2\xi_2^2}=0.05\overline{\xi_1^2}$; $a=40$ (a), 50 (b); the remaining parameters are the same as in Fig. 8.

the effectiveness of such a signal being even greater.

When an additive noise is present, complete suppression no longer occurs, but it is very significant, especially with respect to the variable φ. This is evidenced by Fig. 9, which presents plots of the time dependence of φ and $\dot{\varphi}$ for two values of the amplitude of the signal, whose frequency, as in Fig. 8, is equal to 19.757.

CONCLUSIONS

In summary, we have shown that there is a profound analogy between such apparently dissimilar phenomena as turbulence and noise-induced oscillations of a pendulum. It is very interesting that this analogy is based on the general nature of the laws of the theory of oscillations, rather than on a similarity between the equations of motion. Serious research is certainly required to prove that this analogy is not superficial and that it reflects the essence of the phenomena taking place. Nevertheless, the existence of such an analogy once again confirms the well-known fact that nature obeys common laws.

We thank N. N. Nikitin and A. S. Ginevskiĭ and their co-workers for supplying the data which they obtained in numerical simulations and real experiments.

[1)]For further information on the phenomenon of intermittency, see, for example, Ref. 26.

[2)]These data were obtained and supplied to us by A. S. Ginevskiĭ and co-workers.

[1] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed., Pergamon Press, Oxford (1987) [Russian original, Nauka, Moscow (1987)].
[2] A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics (The Theory of Turbulence), Vol. 1* [in Russian], Gidrometeoizdat, St. Petersburg (1992).
[3] P. D. Drazin and W. H. Reid, *Hydrodynamic Stability*, Cambridge University Press, Cambridge (1981).
[4] L. D. Landau, Dokl. Akad. Nauk SSSR **44**, 339 (1944).
[5] J. T. Stuart, J. Fluid Mech. **4**, 1 (1958).
[6] J. T. Stuart, J. Fluid Mech. **9**, 353 (1960).
[7] J. T. Stuart, Appl. Mech. Rev. **18**, 523 (1965).
[8] J. T. Stuart, Annu. Rev. Fluid Mech. **3**, 347 (1971).
[9] N. V. Nikitin, Izv. Ross. Akad. Nauk Mekh. Zhidk. Gaza, No. 6, 14 (1994).
[10] S. M. Rytov, Usp. Fiz. Nauk **62**, 485 (1957).
[11] D. Ruelle and F. Takens, Commun. Math. Phys. **20**, 167 (1971).
[12] D. Ruelle, in *Statistical Models and Turbulence (Lecture Notes in Physics, Vol. 12)*, M. Rosenblatt (ed.), Springer-Verlag, Berlin (1972), p. 292.
[13] A. V. Gaponov-Grekhov, M. I. Rabinovich, and I. M. Starobinets, JETP Lett. **39**, 688 (1984).
[14] G. V. Osipov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **31**, 624 (1988).
[15] V. V. Kozlov, M. I. Rabinovich, M. P. Ramasanov *et al.*, Phys. Lett. **128**, 479 (1988).
[16] P. S. Landa, Izv. Vyssh. Uchebn. Zaved. Prikl. Nelin. Dyn. **3**(2), 37 (1995).
[17] A. S. Ginevskiĭ and P. S. Landa, Izv. Vyssh. Uchebn. Zaved. Prikl. Nelin. Dyn. **3**(2), 42 (1995).
[18] P. S. Landa, in *Nonlinear Oscillations and Waves in Dynamical Systems*, Kluwer Academic Publishers, Dordrecht–Boston (1996).
[19] Yu. L. Klimontovich, *Turbulent Motion and the Structure of Chaos* [Kluwer Dordrecht (1991); Nauka, Moscow (1990)].
[20] N. V. Nikitin and G. C. Chernyi, Dokl. Akad. Nauk **343**, 767 (1995) [Dokl. Akad. Nauk **40**, 434 (1995)].
[21] M. A. Gol'dshtik and V. N. Shtern, *Hydrodynamic Stability and Turbulence* [in Russian], Nauka, Novosibirsk (1997).
[22] K. F. Teodorchik, *Auto-oscillatory Systems* [in Russian], Gostekhizdat, Moscow (1952).
[23] P. S. Landa, *Auto-oscillations in Systems with a Finite Number of Degrees of Freedom* [in Russian], Nauka, Moscow (1980).
[24] P. S. Landa and A. A. Zaikin, Zh. Éksp. Teor. Fiz. **111**, 358 (1997) [JETP **84**, 197 (1997)].
[25] P. S. Landa and A. A. Zaikin, Phys. Rev. E **54**, 3535 (1996).
[26] Yu. I. Neimark and P. S. Landa, *Stochastic and Chaotic Oscillations* [Kluwer, Dordrecht (1992); Nauka, Moscow (1987)].
[27] L. Prandtl and O. K. G. Tietjens, *Hydro- und Aeromechanik nach Vorlesungen, Vol. 2* [Springer, Berlin (1931); Gostekhizdat, Moscow –Leningrad (1935)].
[28] S. Corrsin, NACA Advis. Conf. Rep. **312**, (1943).
[29] J. Rotta, Ing.-Arch. **24**, 258 (1956).
[30] Y. Pomeau and P. Manneville, Commun. Math. Phys. **74**, 189 (1980).
[31] A. S. Ginevskiĭ, *Theory of Turbulent Jets and Wakes* [in Russian], Mashinostroenie, Moscow (1969).
[32] S. M. Belotserkovskiĭ and A. S. Ginevskiĭ, *Simulation of Turbulent Jets and Wakes by the Discrete-Vortex Method* [in Russian], Nauka, Moscow (1995).
[33] S. C. Crow and F. H. Champagne, J. Fluid Mech. **48**, 547 (1971).
[34] A. S. Ginevskiĭ, E. V. Vlasov, and A. V. Kolesnikov, *Aeroacoustic Interactions* [in Russian], Mashinostroenie, Moscow (1978).

Translated by P. Shelnitz

Influence of an electric field on the dynamic viscosity of liquid dielectrics

A. A. Ostapenko

Scientific-Research Institute of Radiophysics, St. Petersburg State University, 198904 St. Petersburg, Russia
(Submitted December 11, 1996; resubmitted July 2, 1997)
Zh. Tekh. Fiz. **68**, 40–43 (January 1998)

Experimental dependences of the dynamic viscosity of dielectric liquids on the applied voltage are considered. The experiment is remarkable because of the elimination of the electric current through the liquid in a series of measurements by insulating one of the electrodes from the liquid. The change in the viscosity of the liquid media can be a consequence of alteration of the structure of the liquid and the formation of ion–molecule groups differing from the previously existing ones. The variation of the mechanical properties of the medium with variation of the applied field strength points out the influence of charge formation on them. The dependences for polar and nonpolar liquids are considered. © *1998 American Institute of Physics.* [S1063-7842(98)00701-6]

Previously performed investigations of the behavior of a liquid dielectric subjected to the action of a transverse electric field have shown that the viscosity of the dielectric changes. This phenomenon, which has been termed ''electroviscosity,'' has been described by various investigators.[1–3] However, the poor reproducibility of the data and the often apparent contractions between the data of different investigators have made it impossible to clearly single out the main mechanism among the numerous phenomena that might be the cause of the viscosity change. The possible candidates include the transport of momentum by means of the electrophoresis of ions, the orientation of polar molecules,[10] the formation of clusters near an electrode,[1,4] and the influence of space charge and the electrohydrodynamic effect.[5] Despite the diverse interpretations of the phenomenon, all the investigators agree with the need for the passage of an electric current through the liquid. The absence of this phenomenon in nonpolar media was pointed out in some papers.[1,2] However, significant changes in the viscosity of nonpolar liquids in an electric field were noted when poorly purified liquids were used.[2,7] This and several other phenomena indicate that the appearance of space charge at the electrode can be the cause of the viscosity change.

Some investigators believe that the viscosity change is caused by electrohydrodynamic flows appearing in the liquids and that the influence of the field on the molecular transport of momentum is negligibly small. On the other hand, there is an opinion that the action of electric fields can lead to perceptible viscosity changes. A basis for this can be provided by data on the influence of electric fields on the thermal conductivity,[6] since it, like the viscosity, is a characteristic of molecular transport in a liquid.

The influence of charge formation on a medium can be investigated by insulating one of the electrodes from the liquid, which does not eliminate the influence of the electric field on the medium. It is known[8] that electrohydrodynamic flows do not appear when no current passes through a liquid dielectric. Thus, the study of the behavior of a liquid flowing through a cell with one insulated electrode allows us to rule out the influence of electrohydrodynamic flows, on the one hand, and to effect charge injection, on the other hand. If both electrodes are insulated from the liquid, there will be no injection, and only the influence of the electric field on the liquid will be realized.

The equipment and techniques of capillary viscometry were previously employed in Refs. 1 and 2. A diagram of the measuring apparatus is shown in Fig. 1. The time of passage of a liquid dielectric between the marks was measured using a stopwatch or a photorecording device to within 0.1 s. The part of the system to which a high voltage was supplied has the form of a flow-through cell of rectangular cross section placed in a housing of an insulating material (Figs. 1b,c). The dimensions of the channel are as follows: length, 15 mm; width, 3.5 m; height 0.7 mm. The upper and lower walls of the channel are copper electrodes. A cell with a lower electrode insulated from the liquid (Fig. 1c) was employed during the development of the experiment. The electric field was recalculated to find the field actually existing in the channel. The thickness of the dielectric layer was 0.7 mm. The length of the portion of the capillary before the cell was selected such that the flow would have a steady-state character over the entire length of the cell.

Then the dynamic viscosity is

$$\eta = A\frac{\Delta P}{Q}, \tag{1}$$

where A is a constant of the apparatus and ΔP is the pressure drop which causes the flow rate Q.

When an electric field is supplied,

$$\eta_{\mathrm{el}} = A\frac{\Delta P}{Q_{\mathrm{el}}}. \tag{2}$$

Hence the viscosity change is

$$\frac{\eta_{\mathrm{el}} - \eta}{\eta} = \frac{t_{\mathrm{el}} - t}{t},$$

where t_{el} is the time of emergence of the liquid when the electric field E is applied, and t is the time of emergence of the liquid in the absence of an electric field.

It should be noted that the viscosity value obtained without application of a field differed from the values known in

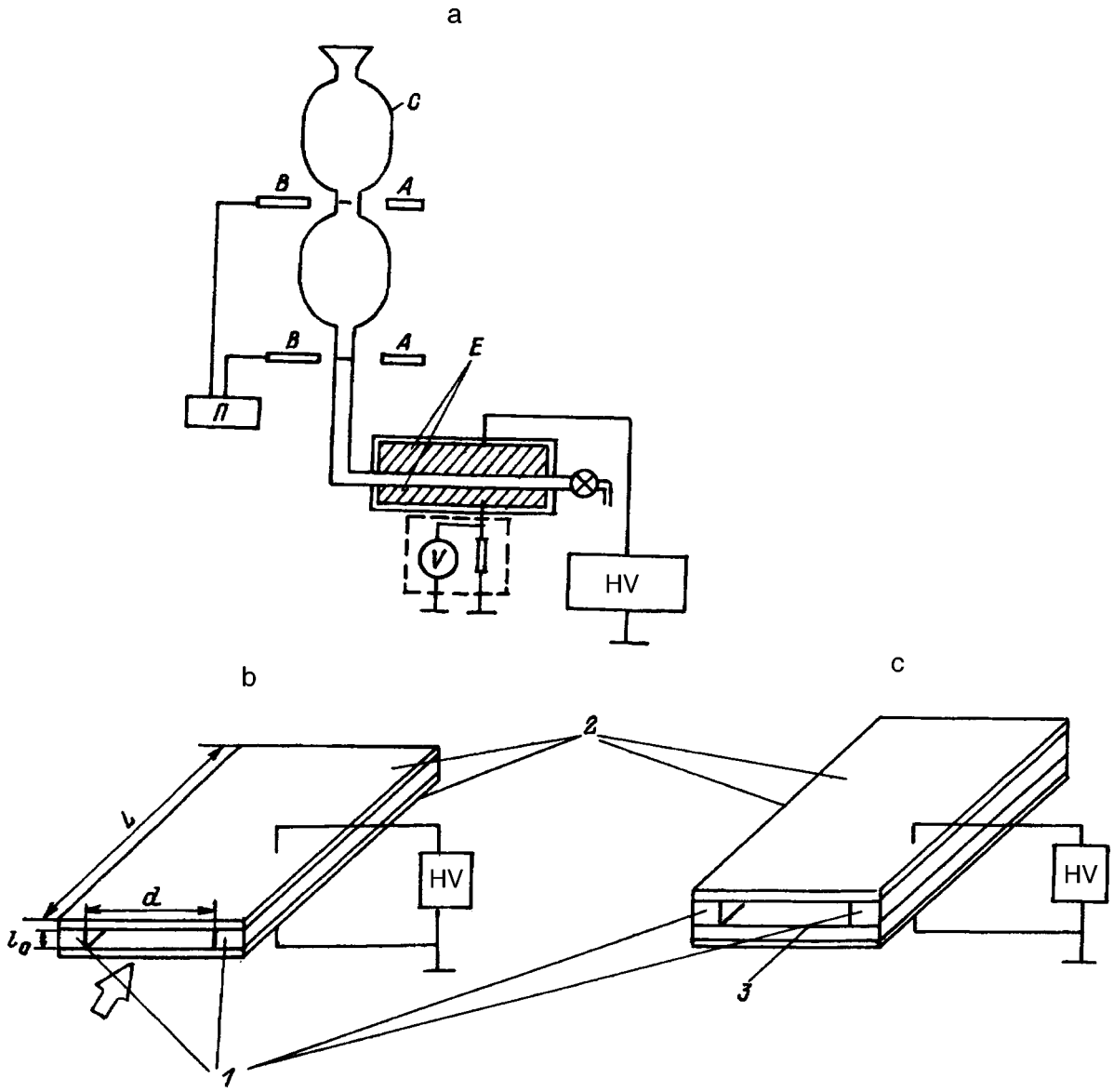

FIG. 1. Diagram of the experimental system. a—overall view; b, c—flow-through cell with an uninsulated electrode and an electrode insulated from the liquid; *A*—source of the light signal, *B*—regenerator, *C*—vessel, *E*—electrodes; *1*—insulating layers, *2*—metallic (copper) layers, *3*—insulating plate.

the literature by no more than 3%. In both cases the passage of a current through the cell was monitored by an electrometric amplifier, which permits the measurement of currents as small as 0.1 nA. The range of field strengths used was 0.8–120 kV/cm at a temperature of 20 °C.

The objects of investigation were liquids with different dielectric parameters [for example, acetone, which is a polar liquid ($\varepsilon=20.1$, and its dipole moment $\mu=9.47\times10^{-30}$ C $\cdot$m), and hexane, which is a nonpolar liquid ($\varepsilon=1.9$, $\mu=0$) (Table I)]. The liquids are all dielectrics and are fairly typical representatives of considerable groups of the latter. As a result of the experiment, we obtained plots of the viscosity change as a function of the applied voltage for both the cell with one insulated electrode and the cell with uninsulated electrodes. The liquids were divided into two groups according to the magnitude of the viscosity change: in one group (the polar liquids) the viscosity changes in the two kinds of cells (with uninsulated and insulated electrodes) ranged from 10 to 160% when the electrodes were uninsulated and from 5 to 40% when one electrode was insulated (Figs. 2a,b). In the other group (the nonpolar liquids) the viscosity changes did not exceed 6% in the system with uninsulated electrodes and 4% in the system with insulated electrodes (Figs. 2c,d). For the most part, all the $\eta(E)$ curves have two segments: a rapid increase and slow variation (saturation). The curves for

TABLE I.

Liquid	$\eta=10^3$, Pa·s	ε	μ, C·m	J, eV	ΔH	$\overline{Z}$
Nitrobenzene	0.18	35	10.2×10^{-30}	7	8.2	20.1
Acetone	0.32	21.2	9.47×10^{-30}	9.7	8.9	9.7
Hexane	0.35	1.9	0	10.5	10.2	33.5
Toluene	0.59	2.4	1.7×10^{-30}	9	10.8	23.5
Chlorobenzene	0.83	5.3	5.2×10^{-30}	9.4	12.4	19.7
Decane	0.92	1.2	0	11	12.2	40.2

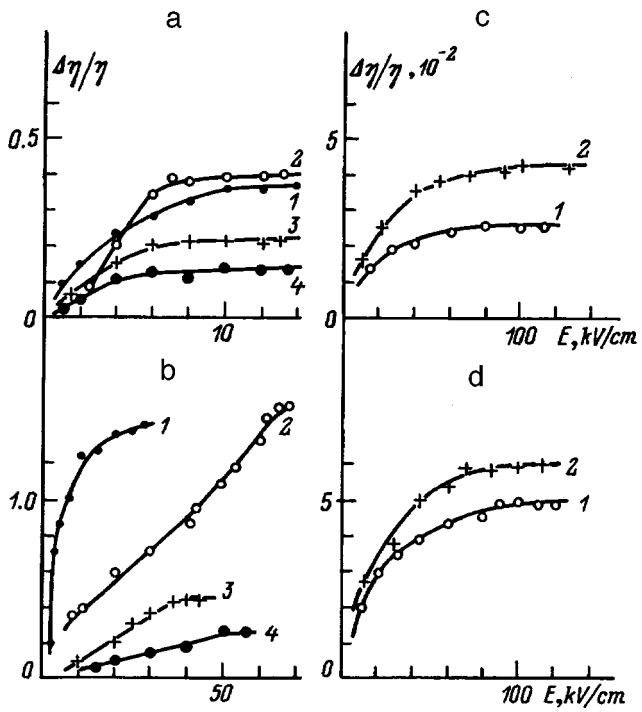

FIG. 2. Relative viscosity change as a function of the strength of the transverse electric field in polar (a, b) and nonpolar (c, d) liquids in cells with one electrode insulated from the liquid (a, c) and uninsulated electrodes (b, d). a, b: *1*—acetone, *2*—nitrobenzene, *3*—chlorobenzene, *4*—toluene; c, d: *1*—hexane, *2*—decane.

acetone and benzene are clearly distinguished from the others by their steep ascending segment. The curves for the cells with one insulated electrode are similar in form. The curves for the nonpolar liquids can also be divided into two segments: a rapid increase and slow variation. The curves for the two types of cells are similar. It should be noted that the magnitude of the viscosity change is higher when a current flows though the cell than in the absence of a current (for example, the corresponding values for decane differ by a factor of 1.5 when the field strength equals 60 kV/cm, and the values for chlorobenzene differ by a factor of 4 when $E=10$ kV/cm). When a current is passed through the cell at $E>E_{cr}$, the increase in viscosity is probably caused by combined electroconvective and molecular momentum transport, while electroconvection does not occur in the cell with the insulated electrode. Qualitative explanations can be proposed for the viscosity change (in this case). It is known[9] that

$$\eta=\left(\frac{hN}{V_m}\right)\exp\left(\frac{\Delta Q}{RT}\right)=A\ \exp\left(\frac{\Delta Q}{RT}\right),$$

where η is the dynamic viscosity, T is the absolute temperature, Q is the free enthalpy, h is Planck's constant, K is Boltzmann's constant, R is the universal gas constant, N is Avogadro's number, and V_m is the molar volume.

In this experiment A and T are constants. Then the viscosity change in an electric field can be associated only with ΔQ: $\eta_{el}=A\ \exp(\Delta Q_{el}/RT)$. From the experiment we have $\eta_{el}>\eta$; therefore, $\Delta Q_{el}>\Delta Q=\Delta H-T\Delta S$, where ΔS and ΔH are the enthalpy and entropy of activation for flow. The estimates in Ref. 10 showed that the change in the free enthalpy in an electric field is attributable mainly to the change in ΔH. The values of ΔQ_{el} and ΔH_{el} were calculated from the formulas[10]

$$\Delta Q_{el}=RT\ \ln\left(\frac{V_m\eta_{el}}{hN}\right),$$

$$\Delta H_{el}=RT\left[\ln\frac{V_m\eta_{el}}{hN}-\left(\frac{T}{V_m}\right)\frac{\partial V_m}{\partial T}\right].$$

It is also known[9] that $\Delta H>\overline{\varphi_m}$, where $\overline{\varphi_m}$ is the molar energy density of the molecular interaction, i.e., flow should be accompanied by deformation of the ''supramolecular structure'' of the liquid and the structure of the local environment, which can be described by the parameter

$$\overline{Z}=\frac{\Delta H}{\overline{\varphi_p}\,\overline{\frac{N}{2}}}\ \text{(Ref. 10)},$$

where $\overline{Z}$ is the mean number of molecules for which correlation of their interactions is observed when the liquid flows and $\overline{\varphi_p}$ is the mean energy of the intermolecular pair interaction.

To estimate the energy of the interaction of molecules of the liquids investigated within the first coordination sphere, we use a variant of the London-Debye-Keesom potential[11]

$$\overline{\varphi_1}-\overline{Z_1}\,\overline{\varphi_p}=\overline{Z_1}\left[\frac{4}{3}J\alpha^2+2\mu^2\alpha+2\frac{\mu^4}{3KT}\right]\frac{1}{\overline{r_1^6}},$$

where J is the ionization potential, α is the polarizability of the molecules, μ is the dipole moment of the molecules, $\overline{\varphi_p}$ is the mean energy of the intermolecular pair interaction, $\overline{Z_1}$ is the number of molecules averaged over the entire volume, and $\overline{r_1}$ is the radius of the first coordination sphere.

The application of an electric field should lead to an increase in the dipole moment of the liquid

$$\mu_{el}=\mu+\left[\alpha+\frac{\mu^2}{3KT}\right]E$$

and, accordingly, in $\overline{\varphi_p}$, but the estimates showed that the change in the dipole moment of the liquid is small (no greater than 1%) for the parameters used and can be neglected. We can still postulate that the increase in viscosity in an electric field is due mainly to enhancement of the correlation between the molecules in the local environment. Then we can write by analogy

$$\Delta H_{el}=\overline{Z_{el}}\,\overline{\varphi_p}\,\overline{\frac{N}{2}},$$

where $\overline{Z_{el}}$ is the mean number of molecules for which the correlation is observed when the liquid flows in an electric field.

It was found experimentally (Table II) that $\overline{Z_{el}}>\overline{Z}$ both in the presence of a current flowing through the cell and in the absence a current. Moreover, $\overline{Z_{el}}$ is greater in the presence of a current than in its absence.

Thus, this research has shown that a change in viscosity occurs in both polar and nonpolar liquids when an electric

TABLE II.

Liquid	No current flowing through the liquid				Current flowing through the liquid			
	E, kV/cm	$\Delta\eta/\eta$	ΔH_{el}, J/mol	$\overline{Z_{el}}$	E, kV/cm	$\Delta\eta/\eta$	ΔH_{el}, J/mol	$\overline{Z_{el}}$
Nitrobenzene	10	0.40	9.0	22.0	10	0.4	8.8	23.1
Acetone	10	0.35	10.0	11.5	10	1.26	10.9	11.8
Hexane	60	0.025	10.0	33.1	60	0.05	10.6	34.5
Toluene	10	0.10	12.0	24.8	30	0.15	12.1	25.1
Chlorobenzene	10	0.20	13.2	21.5	20	0.20	12.9	20.6
Decane	60	0.04	12.9	41.5	60	0.045	13.3	42.3

field is applied across the flow of a liquid dielectric. The effect is observed both with and without the passage of a current through the liquid. The effect is stronger in the presence of a current. In this case it can be assumed that momentum is transferred by combined convective and molecular transport. The main viscosity change occurs because of an increase in the number of molecules for which a correlation is observed, i.e., because of the formation of new structures in the liquid. The presence of an injecting electrode allows us to theorize that the structure-formation centers may be ions, i.e., that ion–molecule complexes appear.

[1] E. N. Andrade and C. Dodd, Proc. R. Soc. London, Ser. A **187**, 296 (1946).
[2] P. T. Sokolov and S. L. Sosinskiĭ, Dokl. Akad. Nauk SSSR **4**, 1037 (1937).
[3] B. N. Dikarev, A. A. Ostapenko, and G. G. Karasev, in *Eleventh International Conference on Conduction and Breakdown in Dielectric Liquids, Zurich*, IEEE (1993), pp. 509–514.
[4] Yu. M. Richkov, V. A. Liona *et al.*, Elektron. Obrab. Mater. (5), 34 (1994).
[5] J. Honda and T. Sasada, Jpn. J. Appl. Phys. **16**, 1775 (1977).
[6] M. K. Bologa, F. P. Grossu, and I. A. Kozhukhar', *Electroconvection and Heat Transfer* [in Russian], Shtiintsa, Kishinev (1977).
[7] *The Electrorheological Effect* [in Russian], A. V. Lykov (edited), Nauka Tekhnika, Minsk (1972).
[8] Yu. K. Stishkov and A. A. Ostapenko, *Electrohydrodynamic Flows of Liquid Dielectrics* [in Russian], Izd. LGU, Leningrad (1989).
[9] S. Glasstone, K. J. Laidler, and H. Eyring, *The Theory of Rate Processes, The Kinetics of Chemical Reactions, Viscosity, Diffusion, and Electrochemical Phenomena* [McGraw–Hill, New York (1941); IL, Moscow (1948)].
[10] M. I. Shakhparonov, *Methods for Investigating the Thermal Motion of Molecules and the Structure of Liquids* [in Russian], Moscow (1963).
[11] A. E. Lutskiĭ, V. S. Klimenko, and E. M. Obukhova, Zh. Fiz. Khim. **52**, 955 (1978) [J. Phys. Chem. **52**, 543 (1978)].

Translated by P. Shelnitz

Generation and transport of high-current, low-energy electron beams in a system with a gas-filled diode

V. N. Devyatkov, N. N. Koval', and P. M. Shchanin

Institute of High-Current Electronics, Siberian Branch of the Russian Academy of Sciences, 634055 Tomsk, Russia
(Submitted July 26, 1996)
Zh. Tekh. Fiz. **68**, 44–48 (January 1998)

Investigations of the generation and transport of a high-current, low-energy electron beam are performed in a system with a gas-filled diode based on a plasma cathode. At accelerating voltages of up to 20 kV and pressures of $(1–5)\times 10^{-1}$ Pa, a beam with an emission current of 600 A, emission current density of 12 A/cm^2 and pulse duration of 30 μs if obtained in a diode with a grid-stabilized emission opening having a diameter equal to 8 cm. The beam is transported in the absence of an external magnetic field over a distance of 20 cm. The beam is compressed by its self-magnetic field, and the current density at the collector reaches 100 A/cm^2 when the beam diameter is 3 cm. © *1998 American Institute of Physics.* [S1063-7842(98)00801-0]

INTRODUCTION

One of the effective ways to improve the service characteristics of structural materials is to treat the surface with pulsed beams of charged particles. It is desirable to carry out surface modification with highly efficient utilization of the energy transported by the beam and a uniform distribution of the current density over a beam cross section with an area up to several tens of square centimeters. These conditions are satisfied if the beam energy is released in a thin (of the order of several microns) layer upon irradiation with a power density of 10^6-10^9 W/cm^2. The release of energy in such a thin layer is ensured by the pulsed high-current ion beams used to investigate surface modification processes, where the ion energy usually amounts to $100-300$ keV.[1]

High-current electron beams may be an alternative to the ion beams used in pulsed heat treatment if they can provide the required power density at electron energies not exceeding 20–30 keV. However, when the electron energy is low, it is extremely difficult to obtain high beam current densities because of the restriction of the current in vacuum diodes according to a 3/2 power law, as well as because of the appearance of both radial and longitudinal dips in the potentials created by the self-charge of the beam during the transport of intense beams. On the other hand, in this case the dimensions of the source can be reduced significantly, and the means of radiation protection, which is an important factor in developing industrial equipment, can be simplified considerably.

Gas-filled and plasma-filled diodes, in which the electrons are accelerated in a space-charge layer formed under definite conditions between the cathode and the anode plasma, have a higher perveance than vacuum diodes. For example, current densities up to 1 kA/cm^2 have been obtained in a diode with an exploding cathode and in a plasma diode[2] in the microsecond range of pulse durations. Plasma emitters with a grid-stabilized emission surface have a high emission capacity.[3,4] Emission current densities up to 60 A/cm^2 have been achieved[4] in a quasistationary regime with a pulse duration of 300 μs and a current of 30 A.

This paper presents the experimental results on the generation and transport of a quasistationary low-energy electron beam in a system with a plasma cathode that provides a current of hundreds of amperes.

DESIGN AND OPERATING PRINCIPLE OF THE ELECTRON SOURCE

The design of the electron source is represented schematically in Fig. 1. Cylindrical magnesium cathode *1* with a diameter of 4 mm and a length of 15 mm and permanent magnet *3*, which creates a magnetic field with an induction equal to 0.02 T are mounted on a Teflon insulator in the shell of cathode section *2*, which simultaneously serves as the trigger electrode. Cylindrical hollow anode *5* with a diameter of 150 mm and a length of 160 mm has an emission window *6* with a diameter of 80 mm, which is covered by a fine metallic grid with holes measuring 0.1×0.1 mm and a geometric transparency equal to 40%, on its end. The hollow anode is connected to the trigger electrode through a resistor with $R=75\ \Omega$. Within the hollow anode there is a spherical distributing electrode *4* with a diameter of 10 mm, which is at the potential of the trigger electrode, at a distance of 20 mm from the cathode section. Electrode *4* improves the uniformity of the distribution of the density of the emission plasma near the grid electrode and, accordingly, the uniformity of the distribution of the emission current density and also permits reduction or elimination of the intrusion of the cathode material into the beam-formation region and its appearance on the treated surfaces of samples. Accelerating electrode *7*, which is fashioned in the form of a diaphragm with an opening having a diameter of 85 mm, is located at a distance of 5 mm from the emission electrode and is connected to drift tube *8*, which has a diameter of 100 mm and a length of 90 mm. Collector *9* is located at a distance of 100–600 mm from the emission electrode. The working gas (air, helium or argon) is injected with a flow rate up to 30 mPa·m^3/s

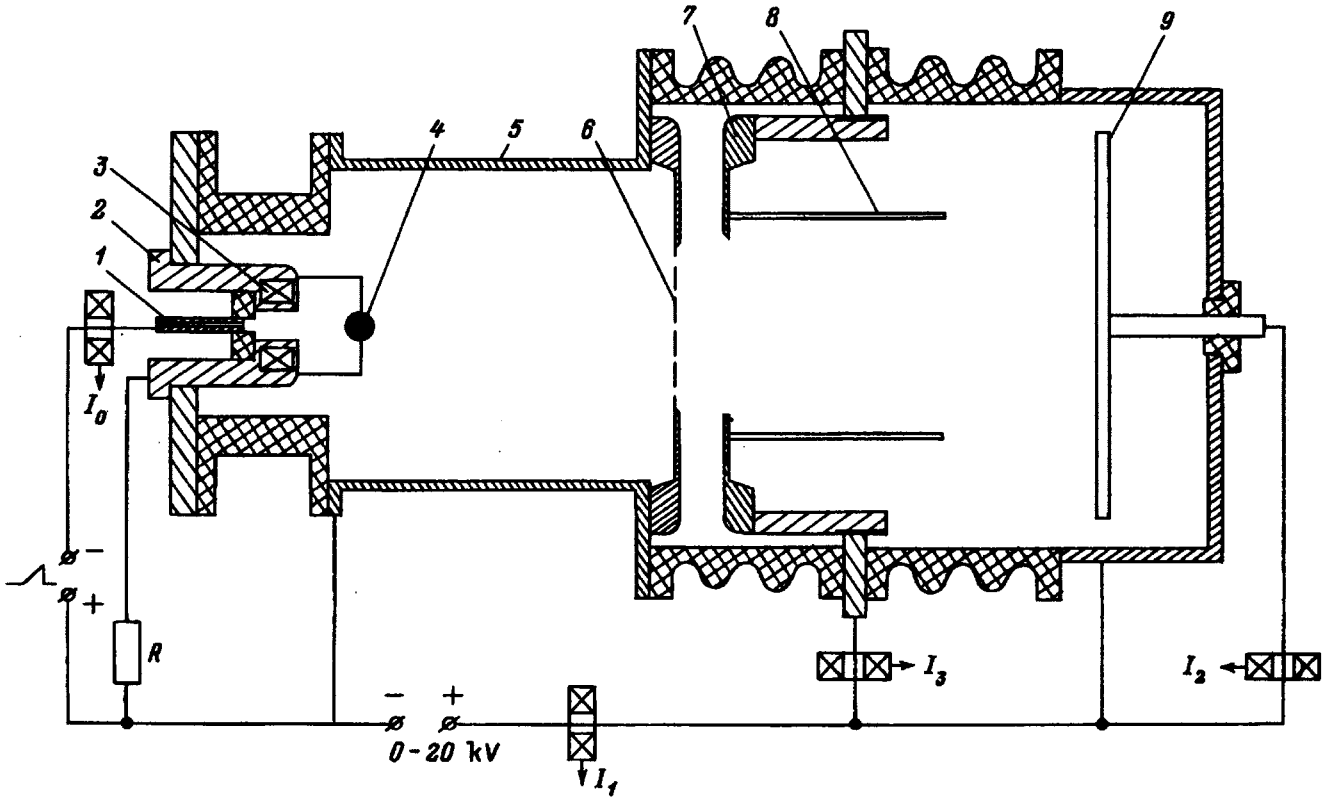

FIG. 1. Diagram of the electron source.

through a channel in the magnesium cathode. The pressure in the system is regulated in the range from 5×10^{-3} to 1 Pa by adjusting the gas injection rate.

The discharge current I_0, the emission current I_1, the beam current at the collector I_2, and the current reaching the accelerating electrode I_3 were measured by Rogowski loops placed in the respective circuits with consideration of the influence of the plasma formed by the electron beam at the collector.[5] The collector was equipped with a special protective ring, which prevents closing of the electrode gaps by the plasma.

When a regulated current pulse in the range from 100 to 1000 A with a duration of 45 μs, which is formed upon the discharge of capacitor into the primary winding of a pulse transformer, is supplied to the cathode, a discharge appears at first over the surface of the insulator between the cathode and the trigger electrode, which initiates the appearance of a cathode spot and ignition of an arc discharge between the cathode and the hollow anode. The plasma fills the hollow anode and the electrons drawn through the holes in the grid enter the accelerating gap. In the electron source under consideration the use of a permanent magnet in the cathode section and the separation of the cathode and anode regions of the discharge by a contracting opening in the trigger electrode permit considerable lowering of the ignition voltage of the arc discharge due to the increase in the pressure in the cathode region as a result of the pressure drop in the contracting channel and prolongation of the residence time of the electrons in the magnetic field.

Under the conditions of a gas-filled diode, electrons extracted from the cathode plasma effectively ionize the gas, creating an anode plasma in the region of the accelerating electrode and the collector. The plasma acquires a potential close to the anode potential,[6] and a space-charge layer, in which electrons are accelerated, forms between the grid emission electrode and the anode plasma. The perveance of the plasma-filled gap exceeds the perveance of a vacuum gap due to compensation of the negative charge by ions supplied from the anode plasma. The mechanism for formation of the space-charge layer and variation of the perveance were considered in Ref. 2 for the case in which the accelerating gap is filled with a plasma before the accelerating voltage is supplied. In the case considered here, in contrast to the case in Ref. 2, the anode plasma, which is responsible for compensation of the beam space charge and its formation, is created in the absence of an external magnetic field as a result of ionization of an injected working gas by electrons extracted from a plasma cathode based on the quasistationary low-pressure arc discharge and compression of the beam by its self-magnetic field.

EXPERIMENTAL RESULTS AND DISCUSSION

Figure 2 presents typical oscillograms of the emission current I_1, the current in the collector circuit I_2, and the current in the accelerating electrode I_3, and Fig. 3 presents experimental current–voltage characteristics of the gas-filled diode for various values of the discharge current (curves *1–3*) and the dependence of the current on the voltage calculated according to the 3/2 power law for a plane-parallel

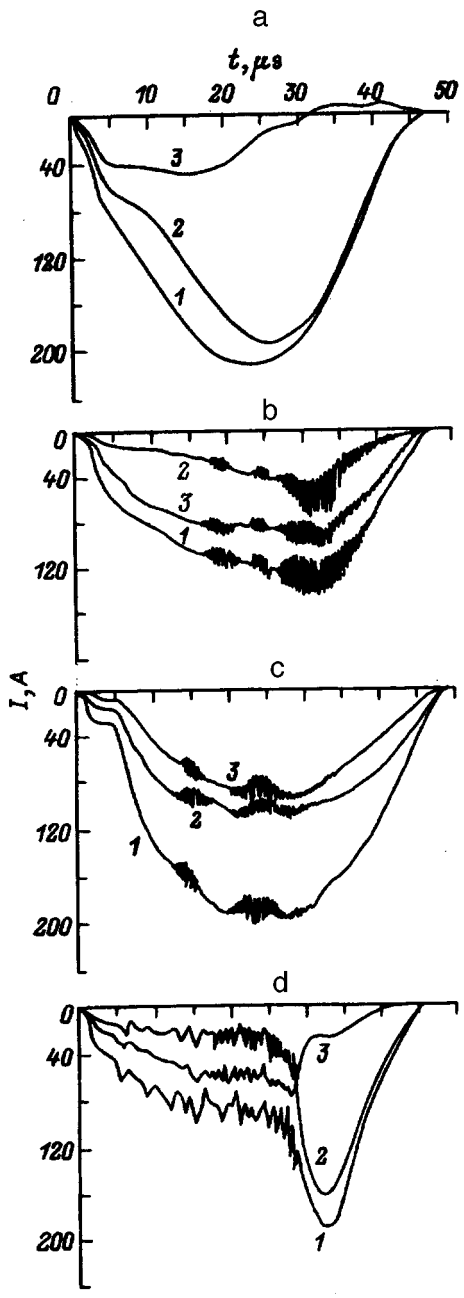

FIG. 2. Oscillograms of the emission current I_1 (*1*), the collector current I_2 (*2*), and the current reaching the accelerating electrode I_3 (*3*) in the absence (a, b) and in the presence of an accelerating grid (c). $I_0=300$ A; p, Pa: a—5×10^{-1}; b, c—5×10^{-2}; d—1×10^{-2}.

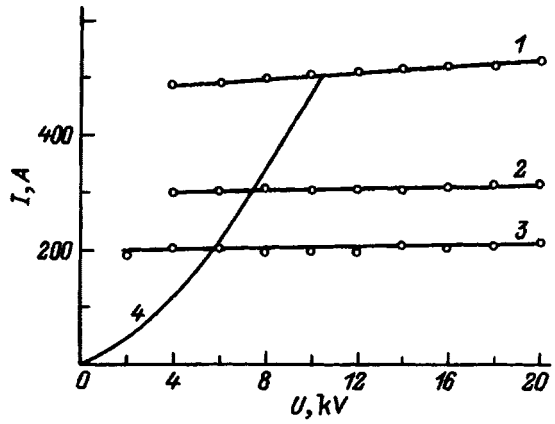

FIG. 3. Current–voltage characteristics of the electron source. $p=10^{-1}$ Pa; I_0, A: *1*—600, *2*—400, *3*—270.

vacuum diode with a 5 mm gap (curve *4*). As is seen from Fig. 3, the diode operates in a saturation regime, and its emission current is determined by the discharge current and can significantly exceed the current of a vacuum diode. The electrons are accelerated in the space-charge layer between the grid of the plasma emitter and the movable boundary of the anode plasma, which is created by ionization of the gas by electrons supplied from the gas-discharge plasma through the opening in the emitter electrode. The increase in the perveance of the diode as the accelerating voltage is lowered is a result of displacement of the anode plasma toward the emitter electrode. The boundary of the anode plasma is established on the basis of the condition of equality of the saturation ion current from the anode plasma and the ion current in the space-charge layer determined from the 3/2 power law

$$l\approx\sqrt{\varepsilon_0/n_e}U^{3/4}/(ekT_e)^{1/4}, \tag{1}$$

where l is the width of the space-charge layer, ε_0 is the dielectric constant, U is the accelerating voltage, e is the charge of an electron, k is Boltzmann's constant, and T_e and n_e are the electron temperature and density in the plasma.

If it is taken into account that, according to Refs. 7 and 8, the plasma density n_e created by a beam under similar conditions can be one to two orders of magnitude greater than the electron density in the beam n_b, then for $n_b=10^{10}$ cm^{-3}, $T_e=5-10$ eV, and $n_e=10^{11}$ cm^{-3} the width of the space-charge layer $l=7$ mm, if $U=10$ kV, and 4 mm, if $U=5$ kV. At pressures above $1-10^{-1}$ Pa the electron extraction efficiency (the ratio of the emission current to the discharge current $\alpha=I_{\rm em}/I_0$) is $\alpha=0.5-0.7$, and the electron beam entering the drift tube is transported in the plasma to the collector with small losses when the drift distance is approximately equal to 20 cm (Fig. 2a). The high transport efficiency attests to compensation of the beam space charge. Even without an external magnetic field, the collector current greatly exceeds the limiting current calculated from the Bogdankevich-Rukhadze formula for the vacuum case

$$I_m=\frac{4\pi\varepsilon_0 m_0c^3[(1+eU/m_0c^2)^{2/3}-1]^{3/2}}{e[1+2\ln(R/r)]}, \tag{2}$$

where m_0 is the electron rest mass, c is the velocity of light, and R and r are the radii of the drift and beam tubes.

The compensated beam is compressed by its self-magnetic field, and the beam diameter d measured by burning thin aluminum foils located at various distances L from the entrance opening of the tube decreases along the following sequence $d_1=7$ cm for $L=2$ cm, $d_2=5$ cm for $L=10$ cm, and $d_3=3$ cm for $L=20$ cm.

Figure 2a presents current oscillograms. At the beginning of a pulse an electron current flows in the circuit of the drift tube. It is formed by accelerated electrons scattered on the tube and plasma electrons. After a certain time, which

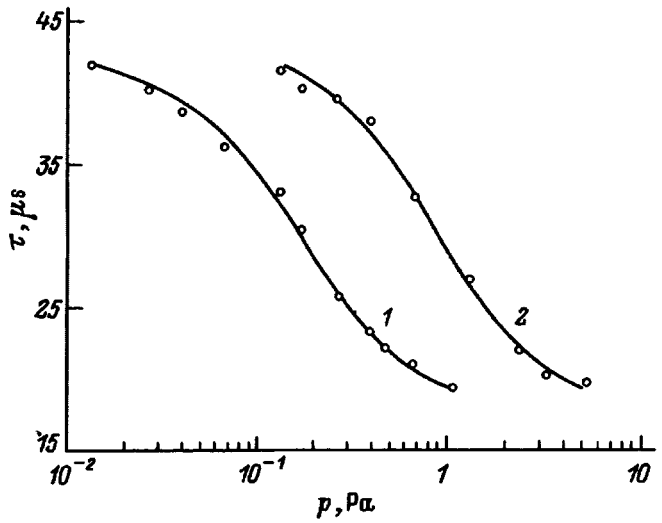

FIG. 4. Dependence of the compensation time of the beam space charge on the pressure of the working gas: *1*—air; *2*—helium. $I_0=600$ A, and the accelerating voltage $U=10$ kV.

depends on the pressure, the current drops to zero, and even the sign sometimes changes. It can clearly be assumed that at the moment when the current on the drift tube equals zero, the beam space charge is completely compensated.

The oscillogram-determined pressure dependence of the time τ, at which the current in the drift tube circuit becomes equal to zero, is presented in Fig. 4. In this case, if helium, whose ionization cross section is smaller than that of nitrogen, is employed as the working gas, a higher pressure is required to compensate the beam space charge. The time τ_0, after which the concentration created by the beam in the drift space is comparable to the electron concentration in the beam, can be determined from the expression

$$\tau_0=\frac{1}{n_0\sigma_i\sqrt{2eU/m}}, \tag{3}$$

where n_0 is the concentration of neutrals and σ_i is the ionization cross section.

Under the conditions of an experiment with a beam electron density $n_b=10^{10}$ cm^{-3}, $n_0=2.3\times10^{13}$ cm^{-3}, $U=10$ kV, and $\sigma_i=10^{-17}$ cm^{-3}, we find $\tau_0=3$ μs. The disparity between the calculated and experimental values of the compensation time is attributable either to the finite rise times of the electron current or to the fact that the escape of plasma electrons is hampered in the transverse direction to the beam's self-magnetic field and they can leave the ionization zone only along the beam toward the collector, while ions escape into the accelerating gap and onto the lateral surface of the drift tube.

The character of the extraction and transport of the beam changes drastically (Fig. 2b), if the pressure in the system is lowered. At a fixed discharge current, the emission and collector currents decrease by a factor of 1.5–2, considerable current fluctuations appear, and most of the beam current reaches the drift tube. The fluctuations are associated mainly with the instability of the current flow in the accelerating gap. Under the conditions of an increase in the discharge current and the emission capacity of the plasma emitter at low pressure, the anode plasma cannot ensure the stable existence of the space-charge layer in the accelerating gap due to the decrease in the ionization frequency $[I_e\approx I_i(M/m)^{1/2}]$. At pressures below $p=10^{-2}$ Pa the beam transport efficiency also decreases, since even the calculated beam space-charge compensation time exceeds 10 μs. This is confirmed by the fact that replacement of the accelerating electrode in the form of a diaphragm by a grid electrode for the purpose of stabilizing the boundary of the anode plasma at low pressure permits an $\sim$1.5-fold increase in the emission and collector currents and significantly reduces their fluctuations (Fig. 2c).

At intermediate pressures of $(1–2)\times10^{-2}$ Pa (Fig. 2d), if the anode plasma provides a number of ions sufficient for compensating the beam space charge and for stable existence of a space-charge layer during a time comparable to or less than the duration of the discharge current pulse, the fluctuations of the emission and collector currents decrease, and the currents reach values corresponding to the those measured at elevated pressures.

When the accelerating voltage is lowered, the compensation time decreases in accordance with (3) due to the increase in the ionization cross section, and the time of the transition to a fluctuation-free current-flow regime shifts to the beginning of the pulse. For example, switching the accelerating voltage from 10 to 5 kV causes a 10 μs shift in the time of the transition to the fluctuation-free regime. Under the conditions of our experiment at pressures exceeding 10^{-1} Pa, we, unlike Vlasov *et al.*,[6] did not observe the development of instabilities or a decrease in the current at the end of a pulse. However, no special investigations of the microwave radiation or plasma diagnostics were performed in the experiments.

Calorimetric measurements established that at elevated pressures the energy transferred by the beam to the collector during a pulse is approximately 20% less than the energy calculated from the oscillograms of the voltage on the gap and the current in the collector. A high current density, reaching $60-100$ A/cm^2, causes partial melting of a copper collector located at a distance of 20 cm from the emitter electrode. The electron beam obtained was used in experiments on the pulsed modification of metal surfaces by quenching from a melt.

CONCLUSIONS

Investigations of a gas-filled diode with a plasma emitter based on a low-pressure arc have been carried out, and a high-current electron beam with a perveance significantly exceeding the perveance of a vacuum diode beam calculated for a fixed gap has been transported. At 10^{-1} Pa the relatively high ionization efficiency of the low-energy beam results in compensation of its space charge in the drift space. The compensated beam is compressed by its self-magnetic field and transported in the drift tube in the absence of an external magnetic field over a distance of 20 cm. Due to its

compression, the current density of the beam at the collector increases by 6–8 fold, reaching ~ 100 A/cm^2 and thereby permitting its use in experiments on the pulsed modification of metal surfaces by quenching from a melt.

[1] G. E. Remnev and V. A. Shulov, Laser Part. Beams **11**, 707 (1993).
[2] D. S. Nazarov, G. E. Ozur, and D. I. Proskurovskiĭ, Izv. Vyssh. Uchebn. Zaved. Fiz. **3**, 100 (1994).
[3] N. N. Koval', Yu. E. Kreĭndel', and P. M. Shchanin, Zh. Tekh. Fiz. **53**, 1846 (1983) [Sov. Phys. Tech. Phys. **28**, 1133 (1983)].
[4] A. V. Zharinov, Yu. A. Kovalenko, I. S. Roganov, and P. M. Tyuryukanov, Zh. Tekh. Fiz. **56**, 687 (1986) [Sov. Phys. Tech. Phys. **31**, 413 (1986)].
[5] A. V. Lazarenko, E. S. Chebukov, and V. I. Énge l'ko, Zh. Tekh. Fiz. **59**(7), 159 (1989) [Sov. Phys. Tech. Phys. **34**, 806 (1989)].
[6] M. A. Vlasov, Yu. A. El'chin, A. V. Zharinov, and V. A. Safonov, Fiz. Plazmy **18**, 1617 (1992) [Sov. J. Plasma Phys. **18**, 840 (1992)].
[7] Yu. R. Alanakyan and N. P. Shternov, Fiz. Plazmy **18**, 355 (1992) [Sov. J. Plasma Phys. **18**, 186 (1992)].
[8] Yu. R. Alanakyan and N. P. Shternov, Fiz. Plazmy **19**, 134 (1993) [Plasma Phys. Rep. **19**, 70 (1993)].

Translated by P. Shelnitz

Consideration of vacancies in the interaction between a liquid phase and a solid phase

A. A. Veksler and A. P. Savitskiĭ

Institute of the Physics of Strength and Materials Science, Siberian Branch of the Russian Academy of Sciences, 634021 Tomsk, Russia

(Submitted July 16, 1996)
Zh. Tekh. Fiz. **68**, 49–52 (January 1998)

An analytical expression is obtained for the time dependence of the concentration of the second component in the diffusion zone of a solid phase during its interaction with a liquid phase. A relation describing the deformation of a powdered body with time is found on the basis of the equations derived. Analytical and experimental curves are compared. © *1998 American Institute of Physics.* [S1063-7842(98)00901-5]

It has been established that the passage of a solid phase into a melt during dissolution is preceded and then accompanied by the diffusion of atoms of the second component from the liquid phase into the solid, which results in the formation of solid solutions or intermetallic compounds in the surface layer of the solid phase adjoining the melt, whose concentration enables this layer to pass into the liquid phase by melting.[1] The law discovered enables us to understand the essence of such phenomena as contact melting[2] and the variation of the mechanical properties of metals and alloys stretched in the presence of melts on their surface (the Rebinder effect[3]). In addition, expressions relating the relative change of the linear dimensions of powdered bodies to the concentrations of the components established in the solid and liquid phases as a result of liquid-phase sintering have been obtained on the basis of new theories regarding the diffusive interaction between a liquid phase and a solid phase.[1] The relations obtained not only agree well with the experimental data, but can also be used very effectively to analyze the physical processes occurring during the liquid-phase sintering of concrete systems. Unfortunately, there is still no mathematical model of the diffusive interaction between a liquid phase and a solid phase that takes into account the kinetics of the real process.[4–6] The theoretical analysis of diffusion processes for regions with mobile phase boundaries in a nonstationary formulation has shown that the analytical methods for solving such problems are unacceptable in the general case, because the regions where they can be solved vary as the phase boundary moves.[6,7] The method used to numerically solve the Stefan problem for a system consisting of interacting solid and liquid phases with a mobile phase boundary is quite tedious.[6] In the present work an attempt is made to obtain an analytical time and temperature dependence of the concentration of a component of a liquid phase in a solid phase due to the diffusive mixing of these phases in the first stage of the interaction on the basis of the vacancy mechanism of diffusion in metals.

RECURRENCE RELATIONS

We start out from the assumption that two interrelated processes take place in the diffusion zone. One process is the filling of vacancies by atoms from the liquid phase, as a result of which the concentration of vacancies tends to decrease. The second process is the creation of new vacancies, as a result of which the concentration of vacancies tends to increase to the equilibrium value. The equilibrium concentration, in turn, continuously increases, because the chemical composition of the diffusion zone varies during diffusion, and the energy for the formation of vacancies decreases. Let us consider a certain region of the crystal lattice of a solid phase G, which consists of an ensemble of N_A of atoms of A and adjoins the liquid component B. Let there be N atoms of B per square centimeter of interface, the total contact area being equal to S. Atoms of B diffuse through the phase boundary into region G by the vacancy mechanism, since they can be accommodated only in lattice sites. At the initial moment in time region G contains an equilibrium number of vacancies[8]

$$b(0)=(N_A+b(0))\exp\left\{-\frac{u}{kT}\right\}.$$

According to the definition of the diffusion coefficient,[5] during the time

$$R=\frac{ba^2}{D_{BA}SN} \tag{1}$$

b atoms of B diffuse into region G. In (1) a is the interatomic distance, and D_{BA} is the diffusion coefficient. In the subsequent time intervals $t(i)$ the number of equilibrium vacancies in region G is determined from the recurrence formula[8]

$$t(i)=\sum_{j=1}^{i} R(j), \quad R(i)=\frac{b(t(i-1))a^2}{D_{BA}SN},$$

$$b(t(i))=\left(N_A+\sum_{j=0}^{i} b(t(j))\right)\exp\left\{-\frac{u}{kT}+\frac{\rho}{N_AkT}\times\sum_{j=0}^{i} b(t(j))\right\}\Bigg/\left(1-\exp\left\{-\frac{u}{kT}+\frac{\rho}{N_AkT}\times\sum_{j=0}^{i} b(t(j))\right\}\right); \quad i\geqslant 1, \quad t(0)=0, \tag{2}$$

where ρ is a constant, which takes into account the additional formation of vacancies as a consequence of the replacement of a certain fraction of the stronger $A-A$ atomic bonds by weaker $A-B$ bonds.

Under the assumption that most of the equilibrium vacancies are replaced by atoms of the second component, the concentration of atoms of B in region G is described by an equality, which has the form

$$C_B(i)\approx\left(\sum_{j=0}^{i} b(j)+N_A\right)^{-1}\sum_{j=0}^{i} d(j). \tag{3}$$

Equations (2) and (3) are valid for the values of the time at which the limiting saturation concentration of atoms of B in the solid phase is not achieved in region G and correspond to averaging over all the possible configurations of the system.[8]

We write equalities (2) for continuous division of the time. For this purpose we assume that $N_A=N_B(t)$ and that, in accordance with some functional measure, the functions $b(t)$ and $t(u)$ have a limit:

$$\lim_{h\to 0}[b(t(u+h))-b(t(u))]/h=db(t)/dt.$$

Then the following system of integrodifferential equations is valid:

$$\begin{aligned}\frac{db(t)}{dt}=&\exp\left\{-\frac{u}{kT}+\frac{\rho}{N_A(t)kT}\int_0^t b(x)dx\right\}b(t)\left(1+\frac{\rho}{kT}\right.\\&\left.+\frac{\rho}{N_AkT}\int_0^t b(x)dx\right)+\frac{db(t)}{dt}+\frac{d(t)^2}{N_AkT}\\&+\frac{dN_A(t)}{dt}+\left\{b(t)+\int_0^t b(x)dx+N_A(t)\right\}\\&\times\left(\frac{dN_A(t)}{dt}\,\frac{\rho}{N_A^2(t)kT}\int_0^t b(x)dx\right),\end{aligned}$$

$$t(z)=\int_0^z\frac{b(t(u))a^2}{D_{BA}SN}du. \tag{4}$$

The proof for this statement is based on the use of expansions of the functions in Taylor series.

The distribution function of the vacancies $F(x,y,z,t)$ in region G is needed to describe the concentration distribution of the atoms of B in it. Then the concentration distribution of the atoms of B in region G has the form

$$\begin{aligned}C_B(x(t),y(t),z(t),t)=&\left(\int_0^t b(u)dF(x,y,z,u)+N_A\right)^{-1}\\&\times\int_0^t b(u)dF(x,y,z,u),\end{aligned}$$

where the integration is construed in the Lebesgue sense. The total concentration of B in region G in the first stage of the diffusion process is determined from the following formula [$N_A(t)=N_A$ is a constant]:

$$C_B(t)=(N_B(t)+N_A)^{-1}N_B(t),\quad N_B(t)=\int_0^t b(u)du, \tag{5}$$

where $N_B(t)$ is the number of atoms of B in region G at the time t.

According to (4), at values of t that are so small that $N_A(t)$ is constant and

$$b(t)^2/N_AT\ll 1,\quad b(t)^2/N_AT\gg 1,$$

we have

$$b(t)=b(0)\exp\{\alpha t\}, \tag{6}$$

where

$$\alpha=\exp\left\{-\frac{u}{kT}\right\}b(0)\left(1+\frac{\rho}{kT}\right)\Big/\left(1-\exp\left\{-\frac{u}{kT}\right\}\right).$$

Therefore,

$$\begin{aligned}&C_B(t)=(N_B(t)+N_A)^{-1}N_B(t),\\&N_B(t)=b(0)\exp\{\alpha t\}/\alpha.\end{aligned} \tag{7}$$

The fact that Eq. (7) assigns an exponential growth law for the number of vacancies as a function of time and that the distribution function of the vacancies can have an unlimited ''tail'' partially accounts for the experimentally observed, fairly rapid formation of a solid solution based on the solid phase due to the diffusion of atoms of the second component in it.

Since $C_B(t)\leqslant C_S^*$, where C_S^* is the saturation concentration of atoms from the liquid phase in the solid phase according to the phase diagram, from (2) we have

$$\rho\leqslant u(1-C_S^*)/C_S^*. \tag{8}$$

DEFORMATION OF THE DIFFUSION ZONE

The diffusion of atoms from the liquid phase into the solid phase in the first stage, in which supersaturation has not yet been achieved in the surface layer, causes displacement of the phase boundary in the direction of the melt (the Kirkendall effect), since new sites occupied by foreign atoms form in the diffusion zone. Therefore, the magnitude of the displacement of the boundary in this stage of the process is proportional to the number of atoms passing into the lattice. The formation of the first portion of liquid in the surface layer of the solid phase after it has been supersaturated signifies that the phase boundary has reversed its direction of motion and traversed a distance equal to the thickness of the layer of the liquid phase formed in the direction of the solid phase.

If particles of a metallic powder that are thoroughly wetted by the melt act as the solid phase interacting with the liquid phase, diffusion from the liquid phase into the solid phase should cause an increase in the size of the particles in the bulk. Conversely, the passage of the surface layer of the particles of the solid phase into the melt by melting should decrease the dimensions of the particles. This obvious relationship between the changes in the volume of the particles and the direction of predominant mass transfer on the phase boundary provides researchers with a unique possibility for

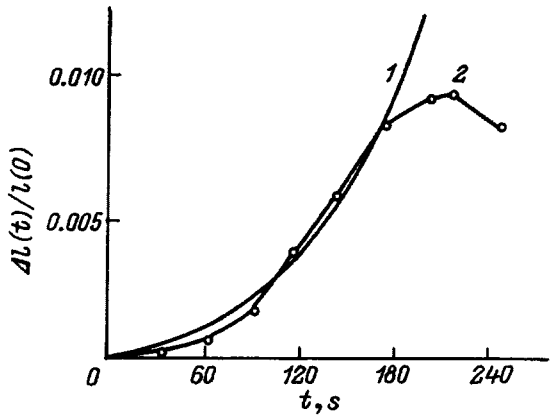

FIG. 1. Dependence of the relative change in the linear dimension of a powered body of the Al–Cu system during sintering.

studying the nature of the interaction of a solid phase with a liquid phase by measuring the volume changes in powdered bodies directly during liquid-phase sintering. In this case experimental results of dilatometric measurements of the growth of powdered bodies in the first stage of liquid-phase sintering can be used to evaluate the efficiency of the relation obtained (7). The relative change in the linear dimension of a powdered body upon sintering in this stage (the displacement of the phase boundary in the direction of the melt) is described by the relation[1]

$$\frac{L(t)-L(0)}{L(0)}=\frac{C_B(t)}{3(1-C_B(t))}, \tag{9}$$

where the function $L(t)$ specifies the linear dimension of a powdered body at the time t.

Substituting (7) into (9), we have

$$\frac{L(t)-L(0)}{L(0)}=\frac{b(0)}{3(\alpha N_A(t))}\ \exp\{\alpha t\},$$

$$\rho\cong u(1-C_S^*)/C_S^*. \tag{10}$$

Let us calculate (10) for a concrete aluminum–copper system, for which $C_S^*=0.033$,[1] $u\cong 1.2176644\times 10^{-19}$ J, and $k=1.38\times 10^{-23}$ J/K at $T=833$ K.[8] The results of the calculation and the experiment are presented in Fig. 1. Curve *1* was constructed according to (10), while curve *2* corresponds to experimental dilatometric data.[1] As can be seen from the figure, curves *1* and *2* nearly coincide at times corresponding to the first stage of the process. The difference between the curves becomes significant only after the third minute of sintering, when the shrinking of the powdered body due to displacement of the phase boundary toward the solid phase begins to be superimposed on the growth process.

We note that Eq. (6) is valid for small values of the time t, since it was assumed during its derivation that the entire region G is in the solid state. However, with the passage of time, some of the atoms of A, together with a certain number of atoms of B, appear in the composition of the liquid phase. This occurs, because along with the increase in the amount of the solid phase in the second stage of the interaction, part of it is lost due to melting on the boundary of G after the critical concentration is achieved in the surface layer. Thus, the increase in the dimension of a powdered body of the aluminum-copper system after the third minute of sintering lags behind an exponential increase (see Fig. 1). The same fact follows directly from Eq. (4), if it is noted that in the second stage of the process $N_A(t)$ is a nonincreasing function, i.e.,

$$\frac{dN_A(t)}{dt}<0.$$

CONCLUSIONS

Recurrence relations (2)–(4) enable us to predict the variation of the dimensions of powdered bodies in the growth stage during liquid-phase sintering as a function of the time and temperature. Their derivation is based on a molecular model of the diffusive interaction between a liquid phase and a solid phase. The expressions obtained, which relate the concentration of equilibrium vacancies in the solid phase to the diffusion time, are of more general interest from the standpoint of their use as tools in the theory of diffusion processes.

The research described in this paper was made possible in part by Grant NY 8000 from the International Science Foundation. We are sincerely grateful to George Soros for the financial support.

[1] A. P. Savitskiĭ, *Liquid-Phase Sintering of Systems with Interacting Components* [in Russian], Nauka, Novosibirsk (1991).
[2] L. K. Savitskaya and A. P. Savitskiĭ, in *Surface Phenomena in Melts and Solid Phases Formed from Them* [in Russian], Kabardino-Balkarskoe Knizh. Izd., Nal'chik (1965), pp. 454–460.
[3] A. P. Savitskiĭ, Izv. Vyssh. Uchebn. Zaved. Fiz. **6**, 56 (1972).
[4] A. N. Tikhonov and A. A. Samarskii, *Equations of Mathematical Physics*, Pergamon Press, Oxford (1964).
[5] A. I. Raĭchenko, *Mathematical Theory of Diffusion in Applications* [in Russian], Naukova Dumka, Kiev (1981).
[6] A. I. Nesterenko and N. G. Nesterenko, Zh. Vychisl. Mat. Mat. Fiz. **24**, 374 (1984).
[7] B. Ya. Lyubov, *Theory of Crystallization in Large Volumes* [in Russian], Nauka, Moscow (1975).
[8] Ya. E. Geguzin, *The Diffusion Zone* [in Russian], Nauka, Moscow (1979).

Translated by P. Shelnitz

Kinetics of the irreversible propagation of a thermal instability in the presence of a nonuniform temperature distribution over the cross section of a superconducting composite

V. R. Romanovskiĭ

Kurchatov Institute Russian Science Center, 123182 Moscow, Russia
(Submitted July 27, 1997)
Zh. Tekh. Fiz. **68**, 53–57 (January 1998)

The conditions for the irreversible propagation of a normal zone along a composite superconductor are investigated within the model of a continuous medium with consideration of its transverse thermal conductivity under the assumption of a uniform distribution of the current over the cross section of the wire. The numerical experiments performed for a current-carrying element of circular cross section with variation of the cooling rate and the transverse dimensions are compared with known results of the one-dimensional theory. It is shown that the one-dimensional theory, as opposed to the two-dimensional theory, leads to underestimated values of the velocity of a thermal instability. The size effect modifies the propagation conditions of a normal zone to the greatest extent as the heat-transfer coefficient increases. This law is based on an increase in temperature at the center of the wire with a simultaneous decrease on its surface, as a result of which the mean temperature over the cross section of the composite increases as its radius increases. © *1998 American Institute of Physics.* [S1063-7842(98)01001-0]

Providing conditions for the stable operation of the current-carrying elements of superconducting magnetic systems is one of the main problems in technical superconductivity. The formulation of these conditions is based on investigations of the processes taking place in the current-carrying elements as a result of the appearance and development of instabilities of various nature. A significant role in understanding the physical features of these phenomena is played by the theory of thermal stabilization.[1,2] Its main assumptions were formulated within very simple one-dimensional models. This made it possible to obtain the principal criteria for the stability of a superconducting state, such as, in particular, the stationary-stabilization condition[3] and the equal-area theorem.[4] In addition, analytical expressions have been derived in the one-dimensional approximation to calculate the velocity of a normal zone in the case of irreversible propagation along a superconductor,[5–7] the stability of the superconducting state of current-carrying elements toward thermal perturbations has been analyzed as a function of the duration and spatial extent of the sources of external heat release,[8,9] and the critical energies of the permissible perturbations have been determined with variation of the percentage of the superconductor in the composite,[10] its properties, the induction of the external magnetic field, and the temperature of the coolant.[11,12] However, the details of the development of a normal zone in a composite stipulated by the presence of transverse heat flow are disregarded in one-dimensional models.[13,14] Another factor that is not so obvious at first glance, on which the conditions for the thermal stabilization of current-carrying elements depend, is the variation of the temperature of the surface of the composite.[15] Therefore, further development of the theory of thermal stabilization should be based on solving problems which take into account the multidimensional character of the processes taking place. In this context, in the present work we investigated the features of the irreversible propagation of a thermal instability along a composite.

Let us consider a cooled superconducting composite of circular cross section with a current uniformly distributed over its cross section, which is immersed in a constant external magnetic field and placed in a coolant with an assigned temperature T_0. We describe the variation of its thermal state within the model of a continuous medium with temperature-averaged parameters using a two-dimensional equation of the form[2]

$$c\frac{\partial T}{\partial t}=\lambda_x\frac{\partial^2 T}{\partial x^2}+\lambda_r\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial T}{\partial r}\right)+\frac{I^2}{S^2}\rho(T),$$
$$t>0,\quad 0<x<l,\quad 0<r<r_0. \tag{1}$$

Here c is the volumetric specific heat of the composite; λ_x and λ_r are the thermal conductivities in the respective directions; I is the current; S is the cross-sectional area; $\rho(T)$ is the effective resistance of the composite

$$\rho(T)=\rho_0\begin{cases}1, & T>T_{CB},\\ \dfrac{T-T_C}{T_{CB}-T_C}, & T_C\leqslant T\leqslant T_{CB},\\ 0, & T<T_C=T_{CB}-(T_{CB}-T_0)\dfrac{I}{I_C},\end{cases}$$

where ρ_0 is the resistivity of the matrix, T_{CB} is the critical temperature of the superconductor at zero current in the assigned magnetic field, and I_C is the critical current of the composite at the temperature of the coolant.

In assigning the necessary initial and boundary conditions we start out from the following assumptions. Let a local site in the composite be heated instantaneously to a tempera-

ture above the critical temperature T_C in the initial moment in time as a result of heat release that is symmetric with respect to the coordinate origin and uniformly distributed over the cross section. Such a change in the thermal state can occur under the action of numerous irreversible mechanical loads, to which current-carrying elements are subjected. In addition, we take into account that the thermal state of the terminal regions will not have an influence on shaping the thermal wave in the composite, if the total spatial extent of the composite is many times greater than the so-called MPZ length.[2] We also assume that heat transfer with a constant heat-transfer coefficient h takes place on the surface of the composite and that the end surfaces of the cylinder are maintained at the temperature of the coolant.

According to the assumptions presented above, the initial boundary conditions have the form

$$T(x,r,0)=\begin{cases} T_1=\text{const}, & 0\leqslant x\leqslant x_1, 0\leqslant r\leqslant r_0, \\ T_0, & x>x_1, \end{cases}$$

$$\left.\frac{\partial T}{\partial x}\right|_{x=0}=0, \quad T|_{x=l}=T_0, \quad \lambda_r\left.\frac{\partial T}{\partial r}+h(T-T_0)\right|_{r=r_0}=0. \quad (2)$$

We introduce the dimensionless variables

$$X=x/L_x, \quad R=r/L_r, \quad i=I/I_C,$$

$$\tau=\lambda_x t/(cL_x^2), \quad \Theta=(T-T_0)/(T_{CB}-T_0),$$

where

$$L_{x,r}=(\lambda_{x,r}S^2(T_{CB}-T_0)/I_C^2\rho_0)^{1/2},$$

which permit simplification of the analysis being performed by using generalized variables that do not depend on the heat-transfer coefficient.[16] After the new variables are substituted into Eq. (1) and condition (2), we obtain

$$\frac{\partial\Theta}{\partial\tau}=\frac{\partial^2\Theta}{\partial X^2}+\frac{1}{R}\frac{\partial}{\partial R}\left(R\frac{\partial\Theta}{\partial R}\right)+i^2\begin{cases} 1, & \Theta>1, \\ \dfrac{\Theta-1+i}{i}, & 1-i\leqslant\Theta\leqslant 1, \\ 0, & \Theta<1-i, \end{cases}$$

$$\Theta(X,R,0)=\begin{cases} \Theta_1, & 0\leqslant X\leqslant X_1, \quad 0\leqslant R\leqslant R_0, \\ 0, & X>X_1, \end{cases} \quad (3)$$

$$\left.\frac{\partial\Theta}{\partial X}\right|_{X=0}=0, \quad \Theta|_{X=L}=0, \quad \left.\frac{\partial\Theta}{\partial R}+\beta\Theta\right|_{R=R_0}=0,$$

where $L=l/L_x$, $R_0=r_0/L_r$, $X_1=x_1/L_x$, and

$$\beta=\frac{hS}{I_C}\sqrt{\frac{T_{CB}-T_0}{\lambda_r\rho_0}}$$

is the dimensionless heat-transfer coefficient. It is related to the so-called thermal-stabilization parameter α (the Steckl parameter)

$$\alpha=\frac{I_C^2\rho_0}{hpS(T_{CB}-T_0)}$$

by the expression

$$\alpha=\frac{1}{\beta}\frac{S}{pL_r}$$

(p is the perimeter being cooled).

Problem (3) enables us to investigate the principal physical features of the kinetics of a normal zone with consideration of the size effect. A finite-difference method was used to solve it because of the piecewise-linear dependence of the effective resistivity of the composite on the temperature. The propagation rate of a thermal instability along a composite in the form of a typical temperature wave was determined numerically with consideration of the qualitative features of its formation.[16]

First, it was taken into consideration that a temperature wave forms under the action of a thermal perturbation, whose energy exceeds a certain threshold value, which describes the upper limit of permissible perturbations. To illustrate this, Fig. 1 presents curves which show the temporal variation of the spatial extent of a normal zone in the central part of a composite and on its surface, which is determined from the equation $\Theta(X_n,R,\tau)=1-i$, for various initial temperatures of the region of the composite that is in the normal state. The calculation was performed for the following parameters: $L=100$, $\beta=0.1$, $i=0.9$, $R_0=2$, and $X_1=1$. It is seen that in the case under consideration the thermal instability acquires an irreversible character when $\Theta_1>1.7$.

Second, since the velocity of a temperature wave does not depend on the character of the initial perturbation, the set of the transcritical parameters X_1 and Θ_1 can take arbitrary values when the velocity is determined numerically. The corresponding plots of the temporal variation of the spatial extent of a normal zone are constructed in Fig. 2 for the case of irreversible propagation along the composite for $L=100$, $\beta=0.1$, $i=0.9$, $R_0=2$, $\Theta_1=1$, and various values of the initial spatial extent of the normal zone. This figure graphically demonstrates the formation of a thermal wave and its propagation with the same velocity for various values of X_1 when the latter is virtually independent of the location of the point in the radial direction, in which the velocity is determined.

Third, to completely eliminate the influence of the transient process preceding the formation of the thermal wave, its velocity was determined only after the distribution of the temperature over a cross section of the composite in the most strongly heated part scarcely varied with time (Fig. 3). In other words, we took into account states that are sufficiently distant from the initial moment in time, which are asymptotically close to the steady-state distribution of the temperature of the thermal wave.

The features noted enable us not only to reduce the error in the numerical determination of the velocity of a thermal instability on the basis of the solution of the nonstationary

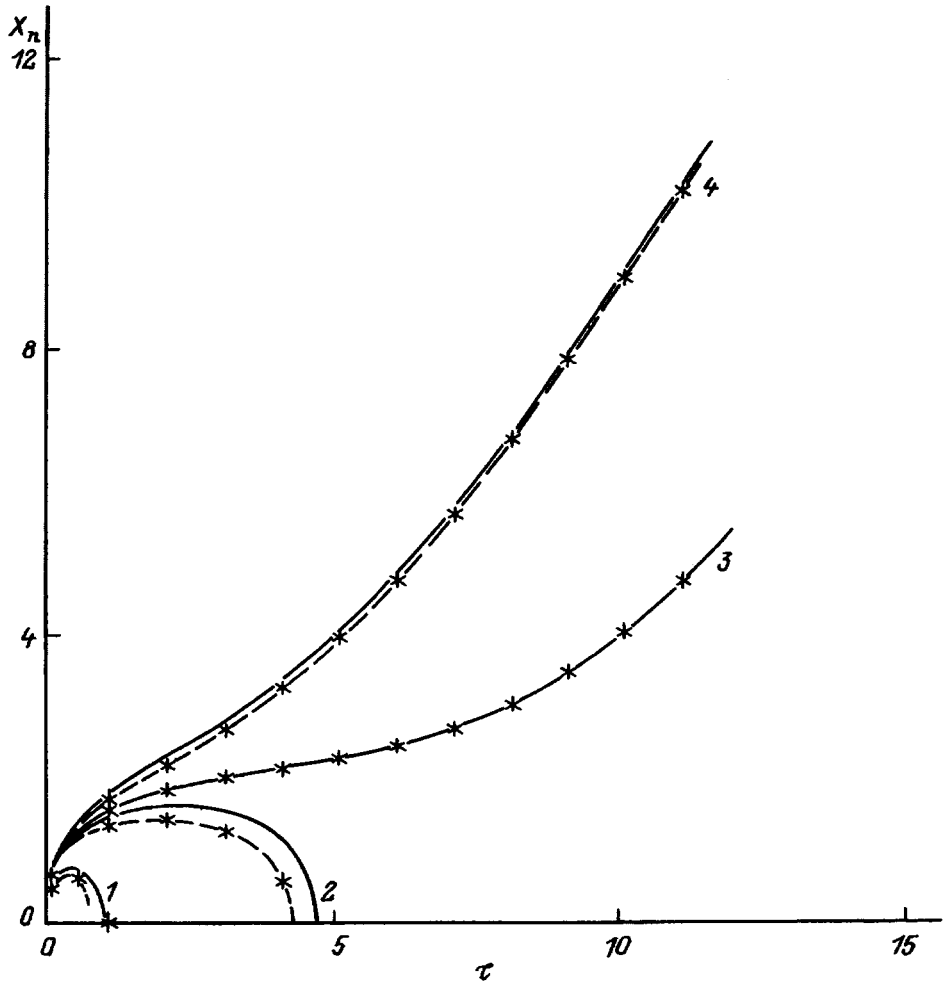

FIG. 1. Propagation of a normal zone at $R=0$ (—) and $R=R_0$ (——*——) under the influence of perturbations with an energy close to the critical value. Θ_i: *1*—1, *2*—1.6, *3*—1.7, *4*—2.

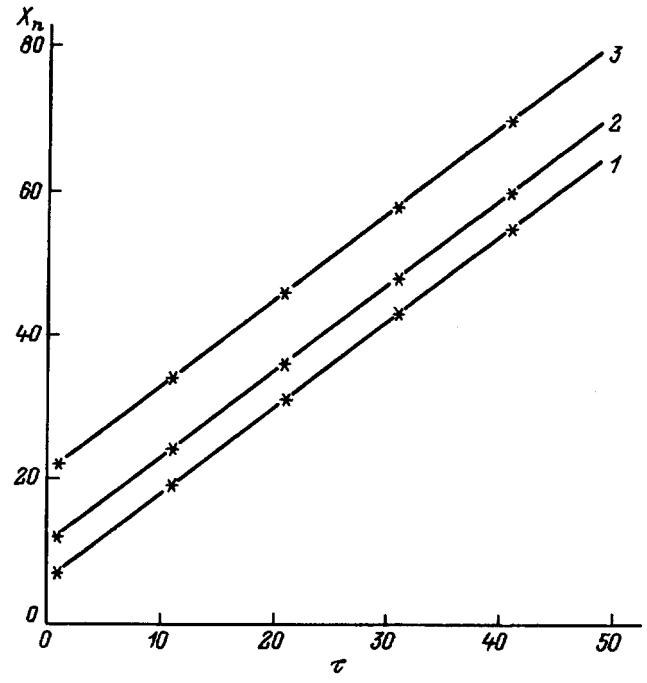

FIG. 2. Propagation of a normal zone at $R=0$ (—) and $R=R_0$ (——*——) under the influence of transcritical perturbations. X_1: *1*—5, *2*—10, *3*—20.

heat-conduction equation, but also to avoid a priori assignment of the temperature of the composite surface.

Figure 4 shows calculated plots of the dimensionless propagation rate of a thermal instability as a function of the transport current for various cooling conditions and wires with two characteristic radius values: a thermally thin composite ($R_0=1$, Fig. 4a) and a thermally thick composite ($R_0=10$, Fig. 4b). The dotted line shows the results of the corresponding calculations according to the one-dimensional model.[16]

As would be expected, the velocity of the thermal wave in a thermally isolated composite ($\beta=0$) does not depend on its transverse dimension, since the thermal state of the current-carrying element is uniform over a cross section. At the same time, the difference between the velocities of a normal zone calculated according to the one- and two-dimensional models for cooled composites depends on the value of β and the transverse dimensions. For example, the dependence of $V(i)$ for a thermally thin wire in the two-dimensional approximation slightly exceeds the corresponding values in the one-dimensional model over a broad range of variation of the current. However, thermally thick current-carrying elements display significant disparity between the one- and two-dimensional theories. This difference can have not only a quantitative character, but also a qualitative character because of the significant increase in the calculated

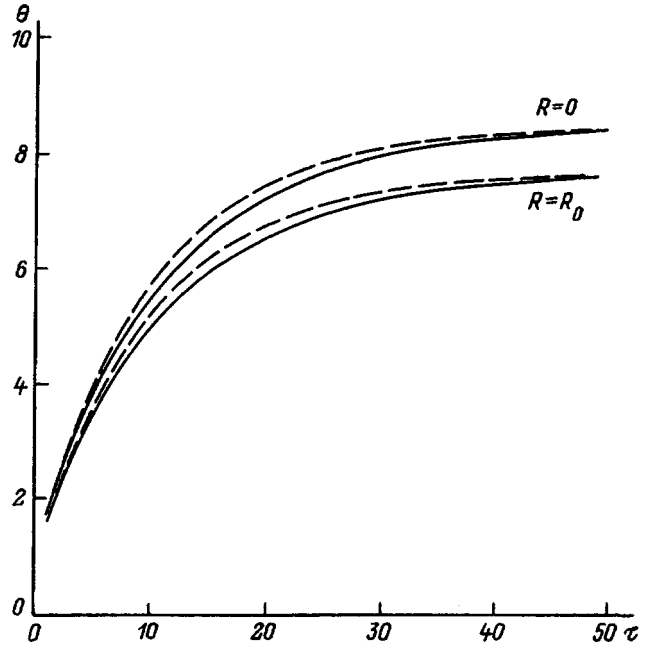

FIG. 3. Increase in the temperature of a composite in the cross sections at $X=0$ for $X_1=5$ (—) and $X_1=20$ (– – – –).

two-dimensional values of the velocity of a normal zone. Switching from the one-dimensional model to the two-dimensional model results in a significant decrease in the range of currents with negative velocity values, which correspond to stable superconducting states. Therefore, if the transient processes in a thick current-carrying element are investigated on the basis of one-dimensional models, and, in particular, the states which are stable toward arbitrary perturbations (at $V<0$) are determined, under real conditions with the parameters under consideration it can be in a metastable state ($V>0$).

The laws noted are based on the corresponding variation of the thermal state of a composite in a cross section, which is not taken into account in the one-dimensional theory. As β increases and, particularly, as heat transfer improves, the temperature distribution in a cross section becomes more nonuniform. As a result, the temperature increases in the central portion of a thick composite and decreases on the surface (Fig. 3). Therefore, despite the fact that in the two-dimensional model, in contrast to the one-dimensional model, there is a conductive heat flow in the transverse direction, the mean temperature in the most strongly heated part increases. To more rigorously prove this claim, let us consider the problem

$$\frac{1}{R}\frac{d}{dR}\left(R\frac{d\Theta_{\max}}{dR}\right)+i^2=0,$$

$$\frac{d\Theta_{\max}}{dR}+\beta\Theta_{\max}\Big|_{R=R_0}=0,$$

which describes the distribution of the temperature in a cross section of a composite that is in the normal state after a large time interval has passed. Its solution is a dependence of the form

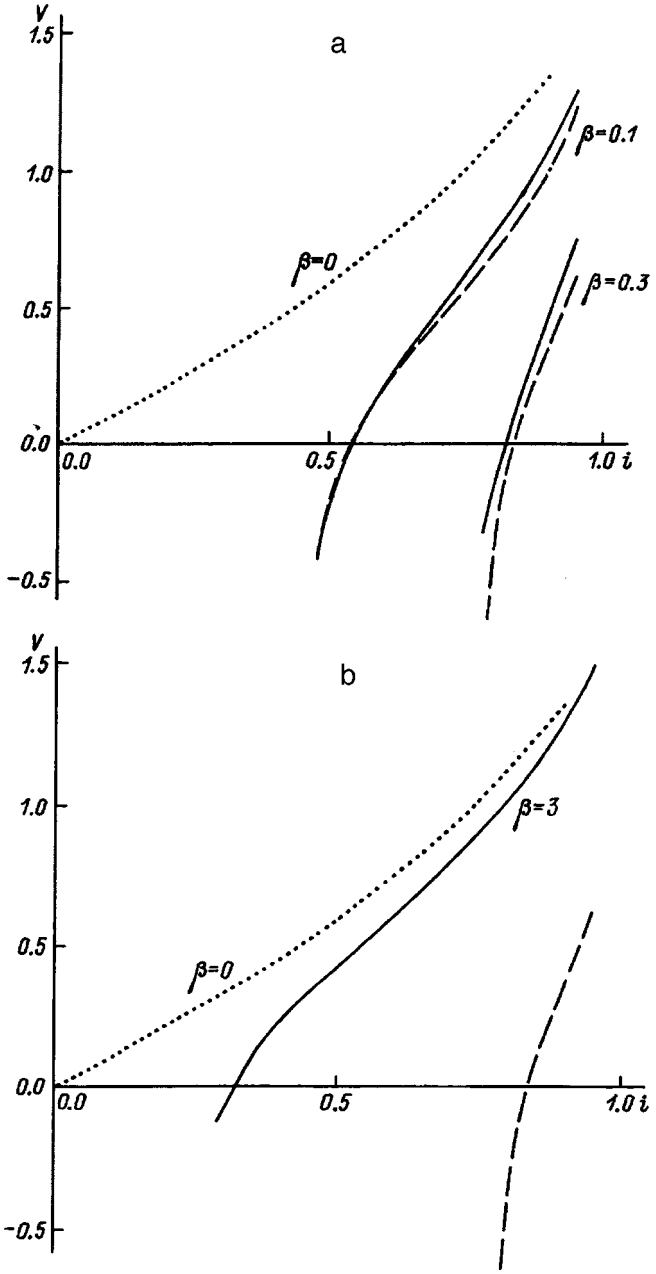

FIG. 4. Dependence of the velocity of a normal zone on the current calculated according to the one-dimensional (– – – –) and two-dimensional models.

$$\Theta_{\max}(R)=\frac{i^2R_0}{2\beta}(1+\beta R_{0/2})-\frac{i^2R^2}{4},$$

which can be used to easily determine the mean temperature over the cross section

$$\langle\Theta\rangle=\frac{1}{S}\int_S\Theta_{\max}ds=i^2(\alpha+R_{0/8}^2).$$

Furthermore, the expression shows that, as opposed to the corresponding one-dimensional value, which equals αi^2, consideration of the size effect leads to a monotonic increase in $\langle\Theta\rangle$ with increasing R_0. The dependence of $V(i)$ is modified accordingly.

Thus, a two-dimensional calculation of the kinetics of thermal instability along a composite shows that the one-dimensional theory underestimates the values of the rate of irreversible propagation of a normal zone. This difference is based on the increase in the mean temperature of the composite in its most strongly heated part. This difference will be observed to the greatest degree in well cooled thick current-carrying elements. It must be taken into account in developing large superconducting magnets.

This work was carried out with support from the Russian Fund for Fundamental Research (Project No. 95-02-03527).

[1] V. A. Al'tov, V. B. Zenkevich, M. G. Kremlev, and V. V. Sychev, *Stabilization of Superconducting Magnetic Systems*, Plenum Press, New York (1977).
[2] M. N. Wilson, *Superconducting Magnets* [Oxford University Press, London (1983); Mir, Moscow (1985)].
[3] Z. J. J. Stekly and J. L. Zar, IEEE Trans. Nucl. Sci. **12**, 367 (1965).
[4] B. J. Maddock, G. B. James, and W. T. Norris, Cryogenics **9**, 261 (1969).
[5] W. H. Cherry and J. I. Gittleman, Solid-State Electron. **1**, 287 (1960).
[6] V. E. Keilin, E. Yu. Klimenko, M. G. Kremlev *et al.*, in *Les Champs Magnetiques Intenses*, CNRS, Paris (1967), pp. 231–236.
[7] B. Turck, Cryogenics **20**, 146 (1980).
[8] V. E. Keilin and V. R. Romanovsky, Cryogenics **22**, 313 (1982).
[9] V. R. Romanovskiĭ, Izv. Akad. Nauk SSSR Energ. Transp. (4), 115 (1984).
[10] V. R. Romanovskii, Cryogenics **25**, 327 (1985).
[11] V. R. Romanovskii, Cryogenics **28**, 756 (1988).
[12] V. R. Romanovskiĭ, Zh. Tekh. Fiz. **63**(1), 41 (1993) [Tech. Phys. **38**, 10 (1993)].
[13] V. R. Romanovskiĭ, Zh. Tekh. Fiz. **60**(4), 31 (1990) [Sov. Phys. Tech. Phys. **35**, 419 (1990)].
[14] V. R. Romanovskiĭ, Izv. Akad. Nauk Energ., No. 2, 129 (1992).
[15] V. R. Romanovskiĭ, Dokl. Akad. Nauk SSSR **330**, 304 (1993) [Dokl. Akad. Nauk SSSR **38**, 216 (1993)].
[16] V. R. Romanovskiĭ, Izv. Akad. Nauk SSSR Energ. Transp., No. 1, 104 (1986).

Translated by P. Shelnitz

Surface plasmon–phonon polaritons of hexagonal zinc oxide

A. V. Mel'nichuk, L. Yu. Mel'nichuk, and Yu. A. Pasechnik

Ukrainian State Pedagogical University, 252030 Kiev, Ukraine
(Submitted May 21, 1996)
Zh. Tekh. Fiz. **68**, 58–62 (January 1998)

The anisotropy of surface plasmon–phonon polaritons of the first and second types in hexagonal zinc oxide is investigated at various electron concentrations and orientations of the C axis of the crystal relative to its surface. It is shown that surface plasmon–phonon polaritons of a third type are generated when the electron concentration in ZnO is greater than 2×10^{18} cm^{-3} and the orientation corresponds to $K\perp C$ and $xy\perp C$. The spectrum of the surface plasmon–phonon polaritons of the third type are calculated, and the conditions for the existence of surface plasmon–phonon polaritons of the third type in ZnO single crystals are determined. © *1998 American Institute of Physics.* [S1063-7842(98)01101-5]

The interaction of electromagnetic waves with the lattice vibrations of a film and a substrate alters the spatial structure of the fields, the existence regions, and the dispersion relations of various surface excitations of the solid-state system.[1,2] The surface plasmon–phonon polaritons of heavily doped anisotropic semiconductors were first investigated in Ref. 3. The number of dispersion relations for surface plasmon–phonon polaritons in a uniaxial semiconductor depends on the charge-carrier concentration and on the orientation of the C optical axis of the crystal relative to its xy surface. The experimental dispersion curves of the surface plasmon–phonon polaritons in heavily doped hexagonal silicon carbide (SiC 6H) were obtained in Ref. 4. However, their differences are determined predominantly by the anisotropy of the electron effective mass in SiC 6H. In the present work the anisotropy of the surface plasmon–phonon polaritons of a hexagonal zinc oxide single crystal was investigated with various orientations of the C axis of the crystal relative to its surface. The mutually consistent parameters of the model of ZnO obtained in Ref. 5 were used in the calculations.

Zinc oxide crystallizes in the wurtzite structure with space group $C_{6V}^4(P6_3mc)$.[6] Experimental modified attenuated total reflection (ATR) spectra of ZnO polaritons were obtained using an IKS-29M spectrometer and an NPVO-2 adapter. A KRS-5 semicylinder with a refractive index $n=2.38$ was selected as the ATR element. The dimensions of the ZnO single crystals ($10\times8\times8$ mm) allowed us to obtain spectra in polarized radiation with various orientations of the C axis relative to the xy surface.

Figure 1 (points) shows three experimental dispersion curves of $\nu_s(K)$ for the ZO2-3 sample, which correspond to the high-frequency ν^+ branches of the polaritons.[3] Curve *1* was obtained for $K\|C$ and $xy\|C$. The dimensionless wave vector $q=Kc/\omega_{T\|}$, where K is the wave vector of the surface plasmon–phonon polaritons, c is the speed of light, $\omega_{T\|}$ is the angular frequency of the transverse optical phonon when $E\|C$, and E is the electric vector of the infrared radiation. Curve *2* is a plot of $\nu_s(K)$ for $K\perp C$ and $xy\perp C$, and curve *3* is for $K\perp C$ and $xy\|C$.

Figure 2 presents the experimental spectra of the ZO2-3 sample for the orientation of the ZnO single crystal corresponding to $K\|C$ and $xy\|C$. The spectra were recorded with an air gap of thickness $d=26$ (curves *1* and *2*) and 3 μm (curves *3–5*) between the ATR element and the sample, and with the angles of incidence α of the IR radiation in the ATR element indicated in the captions to Fig. 2. The minima on the spectra correspond to the frequencies $\nu_{\min}=408$, 450, 496, 518, and 527 cm^{-1}, and the widths of the spectra are $\Gamma_s=32$, 27, 22, 17, and 15 cm^{-1}. Curve *8* was calculated using the data on the ZC1M sample (with an electron density $n_0=4.2\times10^{18}$ cm^{-3}) for polaritons of the third type.

The $\nu_s(K)$ curves (Fig. 1, solid curves) were obtained from the expression

$$K_x^2=[\varepsilon_\|(\nu)-\varepsilon_\perp(\nu)\varepsilon_\|(\nu)]/[1-\varepsilon_\perp(\nu)\varepsilon_\|(\nu)], \tag{1}$$

where $K_x=Kc/\omega$, ω is the angular frequency, ν is the frequency of the radiation, and $\varepsilon_\perp(\nu)$ and $\varepsilon_\|(\nu)$ are the dielectric functions of ZnO in the directions perpendicular and parallel to the C axis.

Equation (1) was written for $K\perp C$ and $xy\perp C$ (curve *2*). If $\varepsilon_\perp(\nu)$ and $\varepsilon_\|(\nu)$ are interchanged, the $\nu_s(K)$ curve (curve *1*) can be obtained using (1). Curves *1* and *2* in Fig. 1 correspond to extraordinary surface polaritons. Ordinary polaritons appear when $K\perp C$ and $xy\|C$. Equation (1) is transformed by replacing $\varepsilon_\perp(\nu)$ by $\varepsilon_\|(\nu)$ (curve *3*).

Agreement between the calculated and experimental data was achieved when the optical parameters of ZnO obtained on the basis of a dispersion analysis of the reflection spectra of the ZO2-3 sample with an electron density $n_0=9.3\times10^{16}$ cm^{-3} were used in the calculation and the anisotropy of the electron effective mass in ZnO was taken into account.[5] The frequencies of the plasmon–phonon modes (curves *2* and *3*) are 548 and 560 cm^{-1}, respectively, and the plasma resonant frequencies are $\nu_{p\perp}=90$ cm^{-1} and $\nu_{p\|}=100$ cm^{-1} (for $E\perp C$ and $E\|C$).

Equation (1) enables us to obtain two dispersion branches in analogy to the isotropic case. The high-frequency ν^+ branch begins at $\nu=\nu_{T\perp}$, and when $K\gg\omega/c$, ν_s asymptotically approaches the value corresponding to the solution of the dispersion equation with neglect of the lag, $\varepsilon_\perp=-1$:

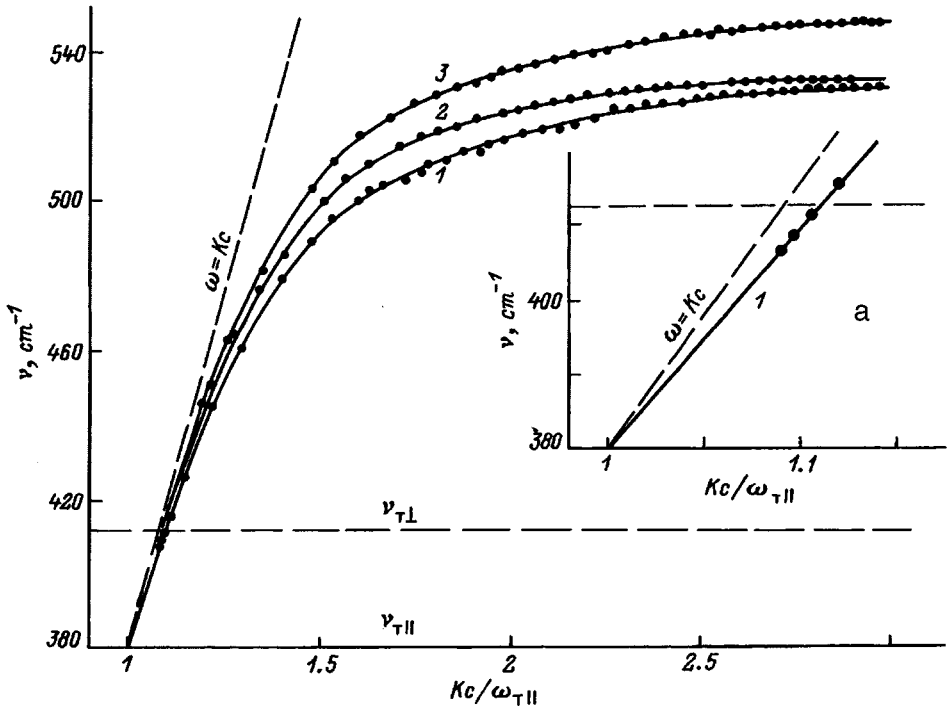

FIG. 1. Dispersion curves $\nu_s(K)$ for the surface plasmon–phonon polaritons of ZnO (the ZO2-3 sample, $n_0=9.3\times10^{16}$ cm^{-3}). *1*—$K\|C$, $xy\|C$; *2*—$K\perp C$, $xy\perp C$; *3*—$K\perp C$, $xy\|C$; a—$K\|C$, $xy\|C$.

$$\nu_{\pm}=(1/2)^{1/2}\{\varepsilon_{T\perp}^2+\xi_{p\perp}\pm[(\xi_{T\perp}^2+\xi_{p\perp}^2)^2-4\xi_{p\perp}^2\xi_{T\perp}^2]^{1/2}\}^{1/2},$$
$$\xi_{T\perp}^2=[(1+\varepsilon_{0\perp})/(1+\varepsilon_{00\perp})]\nu_{T\perp}^2,$$
$$\xi_{p\perp}^2=\varepsilon_{00\perp}\nu_{p\perp}^2/(1+\varepsilon_{00\perp}). \tag{2}$$

The low-frequency ν^- branch exists for all values of K. The polaritons investigated appear under two sets of conditions:

$$1)\,\varepsilon_\perp<0,\quad \varepsilon_\|<0;\quad 2)\,\varepsilon_\perp<0,\quad \varepsilon_\|>K_x^2. \tag{3}$$

In contrast to the isotropic case, new branches appear in ZnO when $K\|C$ and $xy\|C$. The number of existence regions for them depends on the electron density in the conduction band and on the relative positions of the $\nu_{T,\perp,\|}$, $\nu_{L\perp,\|}$, $\nu_{p\perp,\|}$, $\nu_{\perp,\|}^{+,-}$, and $\Omega_{\perp,\|}^{+,-}$ signals. The last eight quantities are defined by the relations[3]

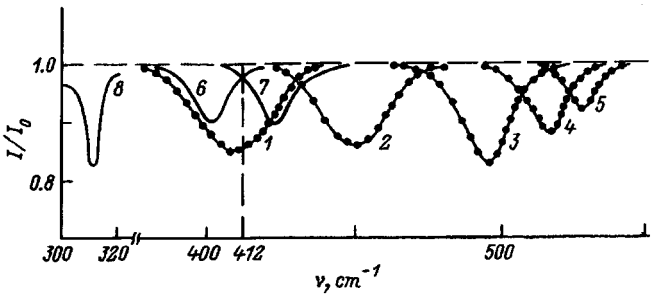

FIG. 2. ATR spectra of ZnO (the ZO2-3 sample, $K\|C$, $xy\|C$). *1,2*—$d=26$, *3–5*—3 μm; *1–5*—$\alpha=25.3$, 28, 34, 42, and 52°, respectively; $\Gamma_s=32$, 27, 22, 17, and 15 cm^{-1}; *6,7*—calculation: $d=72$ and 49 μm, $\gamma_{ph}=11$ cm^{-1}; $\alpha_6=25.2°$; $\alpha_7=25.7°$; $\nu_6=401$ cm^{-1}, $\nu_7=423$ cm^{-1}, $\Gamma_{s6.7}=12$ cm^{-1}; *8*—calculation for the ZC1M sample: $\gamma_{ph}=6$ cm^{-1}, $\gamma_{p\perp}=\gamma_{p\|}=1$ cm^{-1}, $\alpha=28°$, and $d=26$ μm.

$$\varepsilon_\perp(\nu_\perp^{+,-})=0,\quad \varepsilon_\|(\nu_\|^{+,-})=0,$$
$$\varepsilon_\perp(\Omega_\perp^{+,-})=1,\quad \varepsilon_\|(\Omega_\|^{+,-})=1,$$
$$\nu_{\perp,\|}^{+,-}=(1/2)^{1/2}\{(\nu_{L\perp,\|})^2+(\nu_{p\perp,\|})^2\pm[((\nu_{L\perp,\|})^2+(\nu_{p\perp,\|})^2)^2-4(\nu_{p\perp,\|})(\nu_{T\perp,\|})^2]^{1/2}\}^{1/2}. \tag{4}$$

The maximum number of branches can be 4. When $K\gg\omega/c$, their frequencies asymptotically approach the cutoff frequencies corresponding to the solutions of the equation $\varepsilon_\perp\varepsilon_\|=1$. Only the portions of the curves which lie in regions of the (ν,K) plane where conditions (3) hold correspond to surface polaritons. The relation $\varepsilon_\perp\varepsilon_\|=1$ is a fourth-order equation with respect to ν^2, but only one, two, or three solutions lie in regions where $\varepsilon_\perp<0$, i.e., they are confined to finite values of K.

In the isotropic case an ordinary polariton has one low-frequency ν^- branch, which begins at $\nu=0$ and increases up to ν^- as $K\to\infty$. The dispersion curves of the surface polaritons have two branches: $\nu_s^{+,-}(K)$. Figure 3 shows the $\nu_s(K)$ curves of ZnO for $K\|C$ and $xy\|C$. Ordinary polaritons appear in this orientation. Calculations of the dispersion curves were performed in reference to the ZO2-3 (curves *1* and *2*) and ZO1-3 samples (curves *3* and *4*). Curves *1* and *2* were obtained for $\nu_{p\perp}=90$ cm^{-1}, and curves *3* and *4* were obtained for $\nu_{p\perp}=240$ cm^{-1}. Thus, we have $\nu_{pf}^+=561$ cm^{-1} and $\nu_{pf}^-=59$ cm^{-1}, and when $Kc/\omega_{T\|}\to\infty$, we have $\nu_{pf}^+=578$ cm^{-1} and $\nu_{pf}^-=152$ cm^{-1}. The subscript pf is used for the cutoff values corresponding to the values of $\nu^{+,-}$ obtained according to Eq. (2). The calculations show that the frequencies of the polaritons are higher, the greater is the concentration of free charge carriers (electrons in ZnO). Curves *1–4* correspond to polaritons of the first type, which exist for any $K_x^2>1$ (Ref. 7).

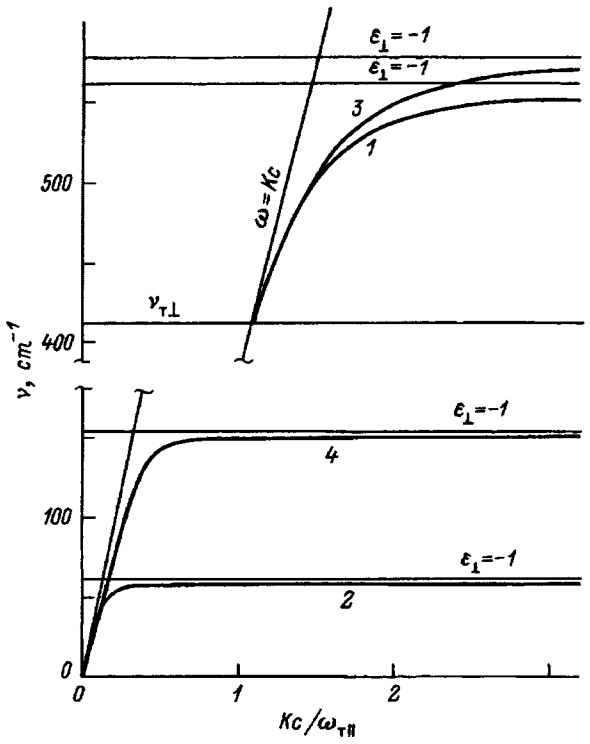

FIG. 3. Dispersion curves $\nu_s(K)$ for the surface plasmon–phonon polaritons of ZnO when $K\perp C$ and $xy\parallel C$.

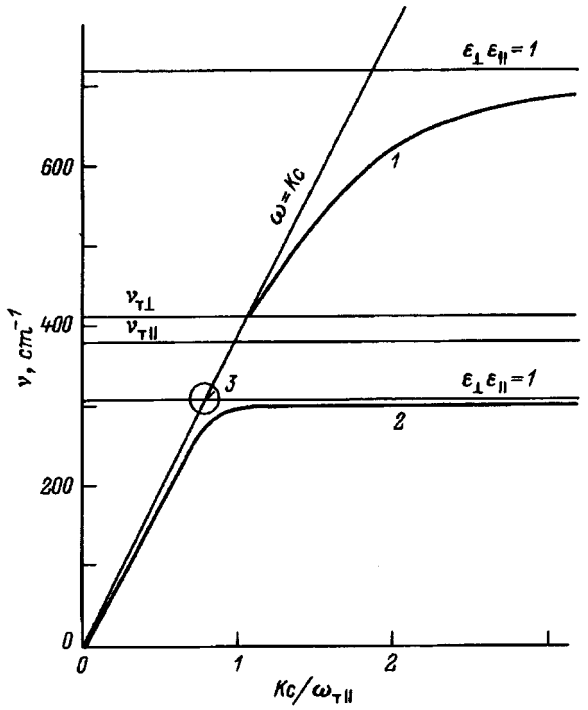

FIG. 4. Dispersion curves of $\nu_s(K)$ for the surface plasmon–phonon polaritons of ZnO when $K\perp C$ and $xy\perp C$.

For ZnO we have $\nu_{T\parallel}<\nu_{T\perp}<\nu_{L\parallel}<\nu_{L\perp}$, and the $\nu_s(K)$ dispersion curves begin at the frequencies on the optical line $\omega=Kc$:

$$\nu=0; \quad \nu=\nu_{T\perp}(\varepsilon_\perp=00) \quad \text{and}$$
$$\nu=\Omega_\parallel^+, \quad \nu=\Omega_\parallel^-(\varepsilon_\parallel=1). \tag{5}$$

For zinc oxide single crystals nonfulfillment of the condition $\Omega_\parallel^+<\nu<\nu_\perp^+$ leads to the presence of only three dispersion branches of $\nu_s(K)$. Figure 4 shows the $\nu_s(K)$ curves for the ZC1M sample, in which $\nu_{p\perp}=605$ cm^{-1} and $\nu_{p\parallel}=650$ cm^{-1}. The calculation reveals the existence of three dispersion curves (curves *1–3*). Here we have $\nu_{pf}^+=719$ cm^{-1} (curve *1*) and $\nu_{pf}^-=305$ cm^{-1} (curve *2*). Curve *3* begins at $\Omega_\parallel^-=309.9$ cm^{-1} for $Kc/\omega_{T\parallel}=0.815$ and ends at $\nu_\perp^-=318.4$ cm^{-1} for $Kc/\omega_{T\parallel}=1.632$. Curve *3* is presented on a magnified scale in Fig. 5a.

Figure 5 shows the dispersion curves of the surface polaritons of ZnO when $\nu_{p\perp}=1300$ cm^{-1} and $\nu_{p\parallel}=1430$ cm^{-1}. In this case $\nu_{pf}^+=1273.7$ cm^{-1}, and $\nu_{pf}^-=363.6$ cm^{-1} when $Kc/\omega_{T\parallel}\to 00$. Here $\nu_s(K)$ (curve *3*) begins at 363.8 cm^{-1} and tends to $\nu_3=390$ cm^{-1} as $Kc/\omega_{T\parallel}\to 00$. The calculation of $\nu_s(K)$ for ZnO with $\nu_{p\perp}=1500$ cm^{-1} and $\nu_{p\parallel}=1650$ cm^{-1} gives $\nu_{pf}^+=1450$ cm^{-1} and $\nu_{pf}^-=367.7$ cm^{-1}. The cutoff frequency of the third dispersion curve is $\nu_3=395.5$ cm^{-1}. The third branch of $\nu_s(K)$ for ZnO differs from the $\nu_s(K)$ curves previously investigated in anisotropic crystals. When the data for the ZC1M sample were used (Fig. 4), a spectrum of the third type was calculated (Fig. 2, curve *8*). The spectrum was obtained for $\gamma_{ph\perp}=\gamma_{ph\parallel}=6$ cm^{-1}, $\gamma_{p\perp}=\gamma_{p\parallel}=1$ cm^{-1}, an angle $\alpha=28°$, and a gap width $d=26$ μm. The minimum of the spectrum of the third type was at $\nu_3=312.5$ cm^{-1} when $Kc/\omega_{T\parallel}=0.92$.

The anisotropy of the polariton is manifested when $Kc/\omega_{T\parallel}>1.2$. For example, it is manifested when $Kc/\omega_{T\parallel}=2\delta\nu_s=18$ cm^{-1} (curves *1* and *3*). As $K\to 00$, curves *1* and *2* tend to 548 cm^{-1}. Figure 1a shows the part of the $\nu_s(K)$ curve in the frequency range 380–420 cm^{-1} on a magnified scale. Surface plasmon–phonon polaritons of the second type, whose existence is restricted by the conditions $\varepsilon_x(\nu)<0$ and $\varepsilon_z(\nu)>K_x^2$ (Ref. 7) are exhibited in this range up to 412 cm^{-1}. The points show the experimental data for the frequencies ν_s of surface plasmon–phonon polaritons of the second type, which are consistent with the calculation; the $\nu_s(K)$ curves of polaritons of the first type [$\varepsilon_x(\nu)<0$ and $\varepsilon_z(\nu)<0$] begin at 412 cm^{-1}, and the experimental $\nu_s(K)$ curve (curve *1*) is continuous. We simulated the spectra of polaritons of the first and second types on a computer at points near 412 cm^{-1} (the ''stopping point'').[8] The spectra were calculated for $K\parallel C$ and $xy\parallel C$ at the frequencies $\nu_{sp2}=411$ cm^{-1} and $\nu_{sp1}=413$ cm^{-1} with gaps equal to $d=72$ and 49 μm and the damping coefficient of the transverse-optical phonons $\gamma_{ph}=11$ cm^{-1} (Ref. 5). Here $K_{sp2}=Kc/\omega_{T\parallel}=1.10415$, and $K_{sp1}=Kc/\omega_{T\parallel}=1.11054$. Both spectra have the same intensity at the minimum and the same width $\Gamma_p=12$ cm^{-1}. The spectra practically overlap each other, indicating that surface polaritons of the first and second types are generated simultaneously in the presence of damping at an emission frequency of 412 cm^{-1}.

The experimental spectrum with a minimum at 408 cm^{-1} has $\Gamma_p=32$ cm^{-1}, which corresponds to the damping coefficient of surface plasmon–phonon polaritons $\Gamma_{sp}=4$ cm^{-1}. The cutoff frequencies are determined using Eq. (2).

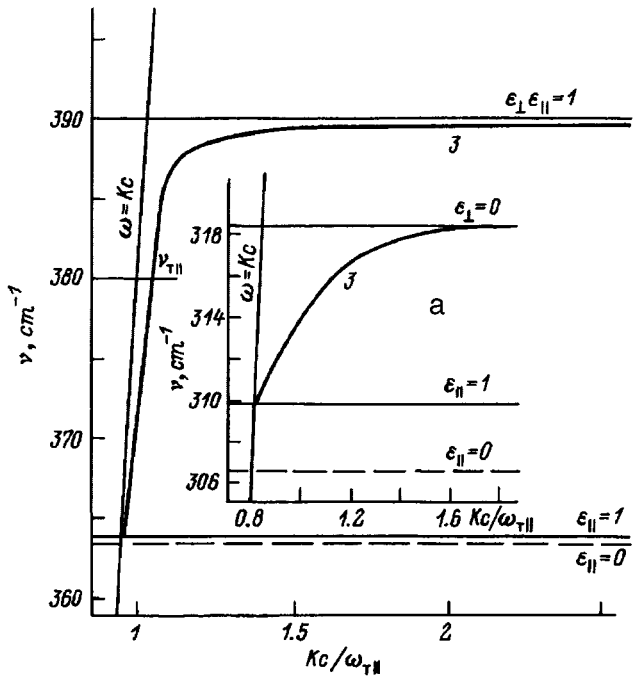

FIG. 5. Dispersion curves $\nu_s(K)$ for the surface plasmon–phonon polaritons of ZnO when $K\perp C$ and $xy\perp C$. *1*—$\nu_{pf}^{+}=1273.7$ cm^{-1}, *2*—$\nu_{pf}^{-}=363.5$ cm^{-1}, *3*—$\nu_3=390$ cm^{-1}; $\Omega_{\parallel}=363.8$ cm^{-1}; $\nu_{\parallel}^{-}=363.6$ cm^{-1}; a—$\nu_s(K)$ of the surface plasmon–phonon polaritons of ZnO for $K\perp C$ and $xy\perp C$; $\nu_{p\perp}=605$ cm^{-1}, $\nu_{p\parallel}=650$ cm^{-1} (the ZC1M sample), $\nu_{\perp}^{-}=318.4$ cm^{-1}; $Kc/\omega_{T\parallel}=1.632$ ($\varepsilon_1=0$), $\Omega_{\parallel}^{-}=309.9$ cm^{-1}; $Kc/\omega_{T\parallel}=0.815$ ($\varepsilon_{\parallel}=1$), $\nu_{\parallel}^{-}=306.5$ cm^{-1}; $Kc/\omega_{T\parallel}=0.017$ ($\varepsilon_{\parallel}=0$).

Curve *2* in Fig. 1 corresponds to the $\nu_s(K)$ curve for $K\perp C$ and $xy\perp C$. It begins at 412 cm^{-1} and tends to the frequency $\nu_{pf}^{+}=548$ cm^{-1} as $Kc/\omega_{T\parallel}\to\infty$. Polaritons of the second type can be displayed only when $\varepsilon_{\perp}<0$ and $\varepsilon_{\parallel}>K_x^2$3 . In the case of ZnO the existence region of polaritons of the second type is restricted to the range 380–412 cm^{-1} (Fig. 1a). They are exhibited when the orientation of the sample corresponds to $K\parallel C$ and $xy\parallel C$.

The $\nu_s(K)$ dispersion curves begin at $\nu=0$, $\nu=\nu_{T\perp}$ ($\varepsilon_{\perp}=00$) and at $\nu=\Omega_{\parallel}^{+}$, $\nu=\Omega_{\parallel}^{-}$ ($\varepsilon_{\parallel}=1$), which coincide with the straight line $\omega=Kc$. The dispersion curves of $\nu_s(K)$ for the polaritons of the third type, which are displayed only in anisotropic crystals at charge-carrier concentrations above a definite value, are of special interest. The existence region of $\nu_s(K)$ for polaritons of the third type is bounded by the straight line $\varepsilon_{\parallel}=1$ and by the straight lines $\omega=Kc$ and $\varepsilon_{\perp}=0$. If $\nu_{\perp}^{-}>\nu_{T\parallel}$, then $\nu_s(K)$ exists at all $K>\Omega_{\parallel}^{-}/c$. Since $\Omega_{\parallel}^{-}\geqslant 301.1$ cm^{-1}, polaritons of the third type begin to be exhibited at $\nu_{p\perp}\geqslant 550$ cm^{-1}. The condition for the existence of the polaritons of the third type of ZnO is 550 cm^{-1} $\leqslant\nu_{p\perp}<1077.3$ cm^{-1}. Figure 4 shows the $\nu_s(K)$ curve of polaritons of the third type (the ZC1M sample) at $\nu_{p\perp}=605$ cm^{-1}. In this case $\nu_{\parallel}^{-}=306.5$ cm^{-1}, $Kc/\omega_{T\parallel}=0.017$ ($\varepsilon_{\parallel}=0$), $\Omega_{\parallel}^{-}=309.9$ cm^{-1}, $Kc/\omega_{T\parallel}=0.815$ ($\varepsilon_{\parallel}=1$), and $\nu_{\parallel}^{-}=318.4$ cm^{-1}, $Kc/\omega_{T\parallel}=1.632$ ($\varepsilon_{\perp}=0$).

The polaritons of the third type of the ZC1M sample exist in the frequency range 306.5–318.4 cm^{-1}. The dispersion curves of $\nu_s(K)$ in Fig. 5 were obtained for ZnO in the orientation corresponding to $K\perp C$ and $xy\perp C$ at $\nu_0>\nu_1$ ($\nu_{\perp}^{-}>\nu_{T\parallel}$), where the cutoff frequency $\nu_3=390$ cm^{-1}. At $\nu_{\perp}^{-}>\nu_{T\parallel}$ polaritons exist when $Kc/\omega_{T\parallel}\to\infty$. The condition for their existence at frequencies from $\Omega_{\parallel}^{-}$ to $\nu_{T\parallel}$ is similar to the case of polaritons of the second type, and polaritons of the first type appear in the range from $\nu_{T\parallel}$ to $\nu_{\perp}^{-}$. The low-frequency $\nu_s^{-}(K)$ curves of the polaritons begin at $\nu>0$, where $K_x^2>1$. As the carrier concentration is increased, $\nu_s(K)$ varies from 0.54 to 9.5 cm^{-1}.

In summary, we have investigated surface plasmon–phonon polaritons of the first and second types in doped anisotropic single crystals of zinc oxide with various orientations of the wave vector relative to the surface and optical axis of the crystal. We have shown that a new dispersion curve for polaritons of a third type forms in ZnO when the electron concentration is greater than 2×10^{18} cm^{-3} and the orientation corresponds to $K\perp C$ and $xy\perp C$. We have determined the existence conditions for polaritons of the new type and have calculated their dispersion relations and spectrum.

[1] E. F. Venger, Yu. A. Pasechnik, O. V. Snitko *et al.*, Pis'ma Zh. Tekh. Fiz. **5**, 1128 (1979) [Sov. Radiophys. **5**, 471 (1979)].
[2] I. I. Burshta, Yu. A. Pasechnik, and O. V. Snitko, Zh. Tekh. Fiz. **57**, 423 (1987) [Sov. Phys. Tech. Phys. **32**, 256 (1987)].
[3] L. É. Gurevich and R. G. Tarkhanyan, Fiz. Tverd. Tela (Leningrad) **17**, 1944 (1975) [Sov. Phys. Solid State **17**, 1273 (1975)].
[4] Yu. A. Pasechnik and E. F. Venger, Poverkhnost' (8) 63 (1982).
[5] A. V. Mel'nichuk, L. Yu. Mel'nichuk, and Yu. A. Pasechnik, Fiz. Tverd. Tela (St. Petersburg) **36**, 2624 (1994) [Phys. Solid State **36**, 1430 (1994)].
[6] I. P. Kuz'mina and V. A. Nikitenko, *Zinc Oxide. Preparation and Optical Properties* [in Russian], Nauka, Moscow (1984).
[7] V. V. Bryksin, D. N. Mirlin, and I. I. Reshina, Fiz. Tverd. Tela (Leningrad) **15**, 1118 (1973) [Sov. Phys. Solid State **15**, 760 (1973)].
[8] H. J. Falge and A. Otto, Phys. Status Solidi B **56**, 523 (1973).

Translated by P. Shelnitz

Relationship between the electronic properties of the interface and the interphase interactions in NbN–GaAs heterostructures

A. A. Belyaev, E. F. Venger, V. G. Lyapin, R. V. Konakova, V. V. Milenin, and Yu. A. Tkhorik

Institute of Semiconductor Physics, Ukraine Academy of Sciences, 252650 Kiev, Ukraine

I. Hotovy

Slovakian Technical University, Bratislava, Slovakia

G. N. Kashin

Institute of Surface Chemistry, Ukraine Academy of Sciences, Kiev, Ukraine

(Submitted August 13, 1996)
Zh. Tekh. Fiz. **68**, 63–66 (January 1998)

Peculiarities of contact formation in the system NbN–GaAs upon a change in the structural–phase state of the deposited metal are investigated. It is shown that the role of the chemical factor in the processes of contact formation decreases as the degree of structural perfection of the NbN alloy increases. The causes of the corresponding changes in the electronic structure of the interface as a result of the phase transition NbN–Nb_4N_3 are discussed. The Auger spectrum and current–voltage and capacitance–voltage characteristics of NbN–GaAs surface-barrier structures are measured before and after annealing in vacuum at $T=850$ °C for 10 s. © *1998 American Institute of Physics.* [S1063-7842(98)01201-X]

INTRODUCTION

Despite active efforts directed towards clarifying the features of the interactions between metals and the semiconductors that are widely used in practical applications, at this time there is no universal model capable of qualitatively describing or predicting the parameters of the Schottky barriers that form when they come in contact. This state of the problem is due to the extreme complexity of the processes that take place when a metal is deposited onto a semiconductor. In general, these processes include diffusion-induced mixing and chemical interactions between the components of the heterojunction pair, which lead to chemical heterogeneity of the interphase boundary, and the formation and decomposition of solid solutions (alloys) in the transition layer of the contact. Neither the intensity and direction of these processes nor the structural states of the reaction products can be predicted from the bulk state diagrams, because the thermodynamic constants at the surface of the semiconductors differ from their corresponding values in bulk. Symmetry breaking, changes in the atomic force constants near the semiconductor surface, and the presence of electric and mechanical fields in the reaction zone of the contact all cause significant changes in the thermodynamic and kinetic parameters of the interacting phases.

These facts give us grounds to assert that the processes of interaction between a metal and a semiconductor depend not only on the physical and chemical state of the semiconductor surface but also to a considerable degree on the state of the thin-film condensate, its structural and morphological characteristics, and its ''reactivity,'' all of which can change the intensity of diffusion processes and phase formation in the contact. One consequence of this is the appearance of transition regions with structure and chemical contents that differ from those of the phasees in contact.[1–5] In order to experimentally justify these conclusions, in this paper we investigate how the structural–phase state of the metal contact affects the characteristics of a NbN–GaAs Schottky contact.

EXPERIMENTAL METHOD

Layers of NbN ($d\cong 100-150$ nm) were obtained by reactive sputtering of Nb (purity 99.95%) in a mixture of argon and nitrogen onto a substrate of n-type GaAs:Te ($n\cong 5\times 10^{16}$ cm^{-3}) with (100) surface orientation.[5] Before the sputtering, the substrate was etched in $H_3PO_4 : H_2O$ and chemically cleaned in a solution of (1:10) $NH_4OH : H_2O$.

The structures of the niobium nitride films were investigated by x-ray phase analysis. Data on the physical and chemical composition of the substrate surfaces prepared for sputtering were obtained using x-ray photoelectron spectroscopy. The character of the distribution of atomic components in the NbN–GaAs heterostructures was determined by layer-by-layer Auger analysis. The electrical parameters of the surface-barrier junctions were studied by current–voltage and capacitance–voltage characteristics. The diode and test structures were subjected to annealing for 10 s at $T=800-950$ °C in vacuum.

RESULTS AND DISCUSSION

Typical photoelectron spectra of the gallium arsenide after chemical processing and washing in deionized water indicate that the dominant phases present at the surface of GaAs prepared for sputtering are: Ga_2O_3 (the binding energy E_b of the $2p^{3/2}$ electrons of Ga is $\approx 1118.4-1118.9$ eV), arsenic (for which $E_b\approx 40.7-41.6$ eV), and As_2O_3

 1063-7842/98/010056-04$15.00

($E_b \approx 44.2$ eV). The Ga : As concentration ratio is substoichiometric (≈ 0.9), and the oxygen content reaches 80 at. % in the subsurface layer.

When magnetron sputtering is used to deposit layers of niobium nitride on such a chemically heterogeneous surface, the data from x-ray structural analysis depend on the content of nitrogen in the working gas mixture. At nitrogen concentrations less than 5% or more than 20%, diffraction phase analysis does not reveal crystallographic phases corresponding to stoichiometric modifications of niobium nitride, and the film structure is amorphous in this case. At intermediate nitrogen concentrations the deposited layers have a polycrystalline structure consisting of grains of the NbN phase, with a phase-centered cubic lattice whose lattice constant varies with the nitrogen concentration in the film. The facts that the ratio of atomic radii of nitrogen and niobium is 0.59 and that the solubility of nitrogen in niobium is small indicate that layers deposited under these conditions are typical interstitial phases with respect to their physical and chemical properties.

Annealing at a temperature of 850 °C leads to crystallization of the amorphous layers of niobium nitride into the face-centered cubic phase of NbN for films obtained with a nitrogen content of 5% in the working mixture, and into the tetragonal phase for layers obtained at high ($\approx 20\%$) nitrogen concentrations. Thus, depending on the content of nitrogen in the working mixture, layers of niobium nitride on GaAs undergo various structural transformations under the same annealing conditions, which indicates that the processes of nucleation and growth of the layers under these conditions are different.

When polycrystalline layers in which the dominant phase is NbN are annealed, a structural transition is observed with the formation of the phase Nb_4N_3.

Not a single case of an intermediate phase including components of the semiconductor substrate or impurities—oxygen and carbon—was reliably identified. This is a problem in need of further detailed investigation.

Thus the results of our structural–phase analysis of layers deposited at a fixed discharge power with various nitrogen contents in the working gas and subsequently annealed suggest that our samples can be separated into three groups. Group I includes samples grown at 5% N_2 content in the working gas. Annealing these samples at 850° for 10 s leads to a structural transformation of the amorphous layer to the NbN phase. Group II consists of structures grown at 15% N_2. These samples are characterized by a NbN–Nb_4N_3 structural phase transition. Group III consists of structures grown at 20% N_2 content; annealing them leads to crystallization of the amorphous niobium nitride film into the phase Nb_4N_3.

Figure 1 shows the results of electrophysical studies of niobium nitride surface-barrier structures subjected to high-temperature annealing. These curves imply that the Schottky barrier parameters and thermal stability depend on the conditions of preparation of the contacts. This is particularly obvious in the reverse-current characteristics of the diode structures, which can change by several orders of magnitude.

We can identify two factors that can modify the electronic structure of heterointerfaces both as they form and during subsequent anneals: a chemical factor associated with the formation of new products when the components that make up the heterojunction pair react with one another, and a structural factor due to peculiarities in the incorporation of the film condensate into the semiconductor lattice.

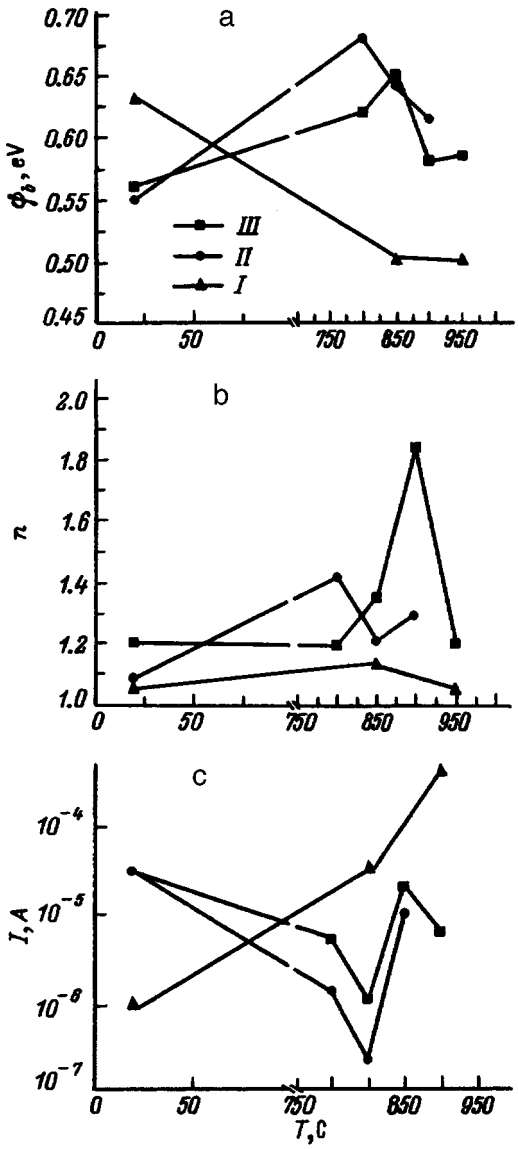

FIG. 1. Schottky barrier height (a), ideality factor (b), and magnitude of reverse current for $U_{rev} = 0.4$ B (c) versus the annealing temperature for NbN–GaAs samples of types I, II, III.

In order to estimate the influence of the first factor, we studied the distribution of elements in the heterojunctions by layer-by-layer Auger analysis (Fig. 2). Our data imply that there is no appreciable difference in the mixing layers in either the as-grown or annealed structures. No clearly expressed features that might indicate the creation of an additional phase were identified in the elemental distribution profiles within the junction region of the contact. The widths of the junction layers calculated from the data of Fig. 2 increased in the following sequence: for type I—amorphous layer–Nb_4N_3, 122–134 Å, for type II—NbN–Nb_4N_3, 210–257 Å, for type I—amorphous layer–NbN, 250–317 Å ; the widths were not correlated with changes in the electrophysical parameters of the barrier structures. Therefore, we

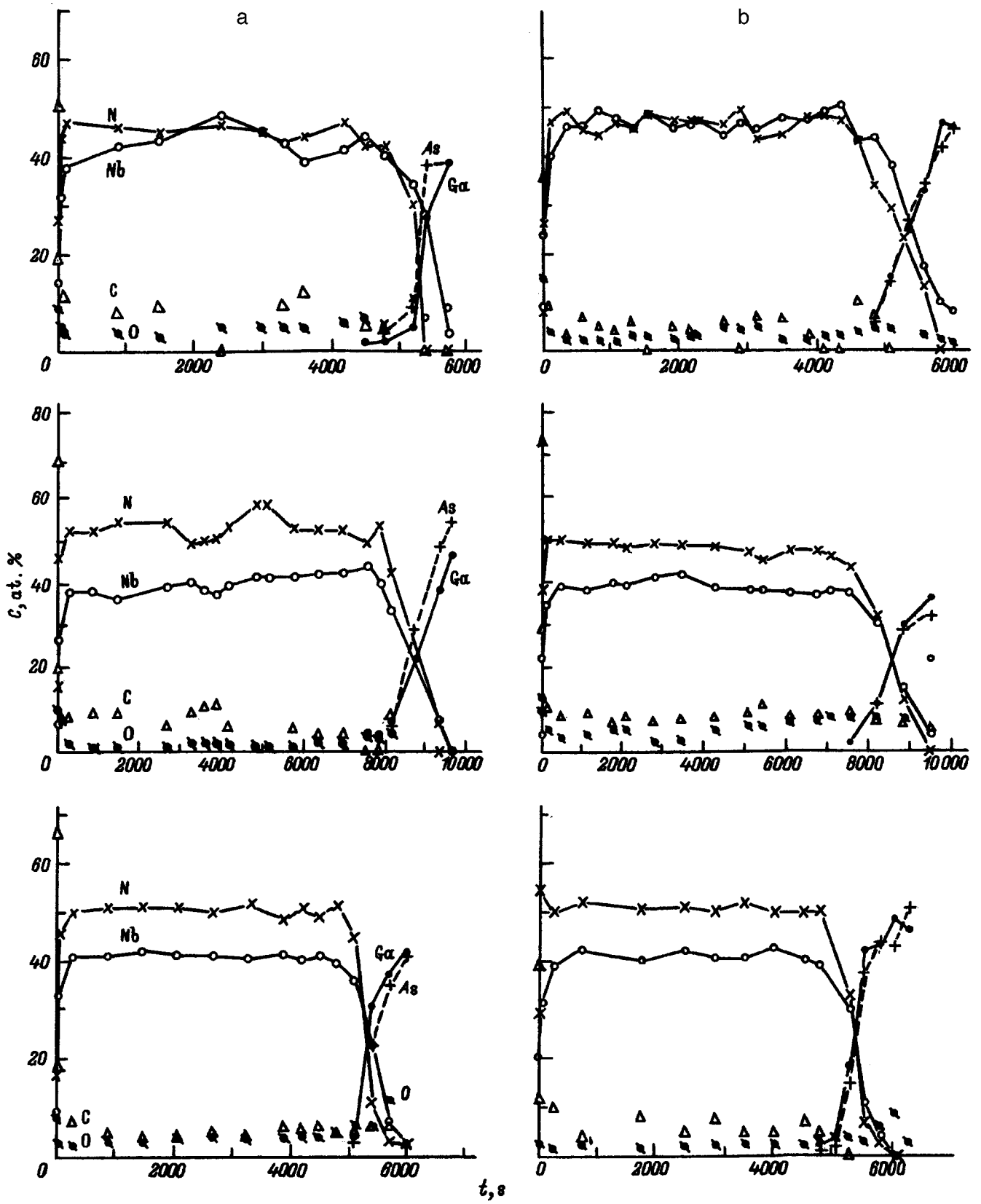

FIG. 2. Profiles of impurity distributions in NbN–GaAs contacts (type I—upper, type II—middle, type III—lower). a—as-grown contact, b—after annealing at $T=850$ °C (10 s).

will assume that the factor connected with the structural state of the deposited film plays the more important role in the processes that determine the electronic properties of Schottky barriers. In this case, not only is the structural compatibility between the metal film and the semiconductor important, but also the degree of disordering of the semicon-

ductor surface caused by the conditions under which the layers were deposited. The degree of disorder determines the intensity of mass exchange in the metal–semiconductor system, which in turn can change the degree of doping of the subsurface layer of GaAs; evidence for this (exhibited by structures of types I and II after annealing and type III before annealing) is the appearance of a voltage-independent segment that interrupts the linear trend of the inverse-square junction capacitance versus the applied bias. Moreover, in polycrystalline films an important role is played by the ratio of diffusion fluxes through the volume of a grain and via intergrain boundaries, which depends on the size of the grains and the layer thickness. Although a systematic and consistent inclusion of all the circumstances mentioned above will require additional studies, for now we can say that the best electrical parameters are exhibited by contacts between GaAs and crystalline Nb_4N_3 films, regardless of how the latter phase is formed, i.e., by crystallization of an amorphous layer or as a result of a recrystallization NbN $\rightarrow Nb_4N_3$.

Calculations based on the principle of three-dimensional matching, analogous to those of Ref. 4, show that direct incorporation of both Nb_3N_4 and NbN into the crystal lattice of the GaAs substrate is accompanied by the appearance of an appreciable concentration of electrically active defects at the boundary, while contacts with good electrophysical characteristics are possible only when the defect concentration at the boundary is small.

This latter difficulty probably could be addressed by putting an intermediate monolayer, e.g., of NbAs, between the GaAs substrate and Nb_3N_4 or NbN layer.

CONCLUSIONS

(1) The atomic structure of a niobium nitride layer depends on the growth conditions; (2) polycrystalline films of niobium nitride obtained by magnetron sputtering are, as a rule, single-phase; (3) heat treatment of NbN/GaAs junctions causes the structural state of the deposited layer to change, converting it from an amorphous layer to a crystalline or recrystallized layer, with the formation of a new phase; (4) the best parameters of niobium nitride–GaAs Schottky contacts are achieved by heat treatment at 850 °C and correspond to a film state whose dominant phase is Nb_4N_3.

In closing, the authors of this paper who are on the scientific staff of the Institute of Semiconductor Physics, National Academy of Sciences of the Ukraine, express our deep appreciation for joint funding by the government of the Ukraine and the International Science Foundation (Grant N R5R100), which provided us with support while we carried out these investigations.

[1] X. W. Wu, L. C. Zhang, P. Bradley *et al.*, Appl. Phys. Lett. **50**, 287 (1987).
[2] A. I. Akhin'ko, E. Ya. Gol'dberg, A. T. Grigor'ev *et al.*, Élektronnaya Tekhnika Ser. 3: Mikroélektronika **1**(135), 35 (1990).
[3] A. I. Akhin'ko, V. N. Inkin, E. P. Karamyshev *et al.*, Élektronnaya Tekhnika Ser. 3: Mikroélektronika **5**(139), 42 (1990).
[4] V. N. Inkin, G. G. Kirpilenko, and S. M. Portnov, Élektronnaya Tekhnika Ser. 3: Mikroélektronika **4**(143), 32 (1991).
[5] I. Hotovy, J. Brcka, and J. Huran, Fizika A **4**(2), 337 (1995).

Translated by Frank J. Crowne

Production of highly enriched mercury isotopes by a photochemical method

Yu. V. Vyazovetskiĭ and A. P. Senchenkov

Kurchatov Institute Russian Science Center, 123182 Moscow, Russia
(Submitted August 2, 1996)
Zh. Tekh. Fiz. **68**, 67–74 (January 1998)

The Foton-M automatic system for the production of mercury isotopes has been developed in Russia. The isotopes are obtained by means of a photochemical reaction between excited mercury atoms and oxygen in the presence of 1.3-butadiene. The possibilities for separating all mercury isotopes, including those with overlapping resonance line profiles, are examined. The Foton-M system can produce 98%, 99%, and 99% concentrations of the isotopes Hg-196, Hg-198, and Hg-202, respectively, in sufficient quantities to meet demand in Russia and for export abroad. © *1998 American Institute of Physics.* [S1063-7842(98)01301-4]

INTRODUCTION

There is a steady demand for mercury isotopes in Russian and in the world markets. One method of obtaining mercury isotopes is the photochemical method, which has recently seen intensive development. Technologies and automatic systems[1] designed to produce appreciable quantities of mercury isotopes have been developed in Russia. At the time, attention was focused mainly on studies aimed at the development of a technology to produce the rarest, most expensive mercury isotope Hg-196. This isotope attracted interest mainly because of its medical application for the diagnosis of various diseases[2] and also to investigate the possibility of using mercury up to 2–4% enriched in Hg-196 to enhance the luminous efficiency of light sources.[3]

The photochemical method of isotope production saw further progress with the development of a technology for Hg-202 enrichment of mercury to a concentration of 99.8% (Ref. 4). The methods and techniques developed can produce this high enrichment not only for Hg-202 but also for Hg-196 and Hg-200. The Foton-M photochemical system[5] based on the Foton system described earlier was used for these experiments.

In some previous studies it was shown that the photochemical method of obtaining Hg-196 was more economical than the electromagnetic technique.[1] Mercury enriched in Hg-196 was first obtained in 1968 by French researchers.[6] Their experimental separating system delivered a product enriched in Hg-196 with a yield of around 1 mg/d. Some of the main constants characterizing this process were measured. Slightly later, some German scientists published studies of various aspects of obtaining Hg-196 by using a photochemical mercury oxidation reaction in the presence of 1.3-butadiene.[7,8] The light source in these experiments was a low-pressure lamp filled with Hg-198-enriched mercury.

American researchers proceeded to study the photochemical reaction between excited mercury atoms and hydrogen chloride.[9] The reaction with HCl has a substantially higher quantum yield compared with the oxidation reaction, but the enrichment factor is lower.

The technology involved in producing mercury isotopes by a photochemical method has under development in Russia for some years. The main process used in these experiments is the photochemical reaction between excited mercury atoms and oxygen in the presence of 1.3-butadiene. The mercury atoms are excited by $\lambda = 253.7$ nm resonance radiation from a low-pressure lamp. This reaction has the maximum selectivity for the separation process, which is particularly important for the enrichment of mercury isotopes with overlapping 253.7 nm line profiles. In addition, the enrichment process can be automated by using various physical and technological means so that it can be carried out round the clock with minimum supervisory staff.[10]

POSSIBILITIES FOR OBTAINING MERCURY ISOTOPES BY THE PHOTOCHEMICAL METHOD

The isotopic hyperfine structure of the 253.7 nm mercury resonance line is shown in Fig. 1. It can be seen that under real enrichment conditions, even when the mercury atom is excited by a fairly narrow radiation line, only the isotopes Hg-202, Hg-200, and Hg-196 can be separated from the isotope mixture. The other isotopes will be excited and will therefore be separated in groups: Hg-199, Hg-201 and Hg-204; Hg-198 and Hg-201; Hg-199 and Hg-201.

Furthermore, the radiation line of the mercury lamp used to excite the mercury atoms is not monochromatic. Even if the lamp is filled with mercury enriched up to 99.8% in the target isotope, a small percentage of its radiation will excite other isotopes because of the nonlinearity of the luminous efficiency. The mutual overlap of the emission line profiles of the lamp and the absorption line profiles of the mercury atoms in the reaction cell has the result that the optical selectivity of the process is limited.

The mutual overlap of the resonance line profiles of the Hg-198, Hg-199, Hg-201 and Hg-204 isotopes means that direct excitation of the target isotopes by radiation from a monochromatic light source is inadequate to obtain high enrichment of these isotopes. However, this does not mean that isotopes with overlapping spectra cannot be separated by the photochemical method. Tested and proven procedures are available to separate all isotopes of mercury photochemically. Among these procedures, particular mention should be made of filtering the light source radiation to suppress that component which reduces the selectivity of the process. The selectivity of the photochemical reaction can be enhanced

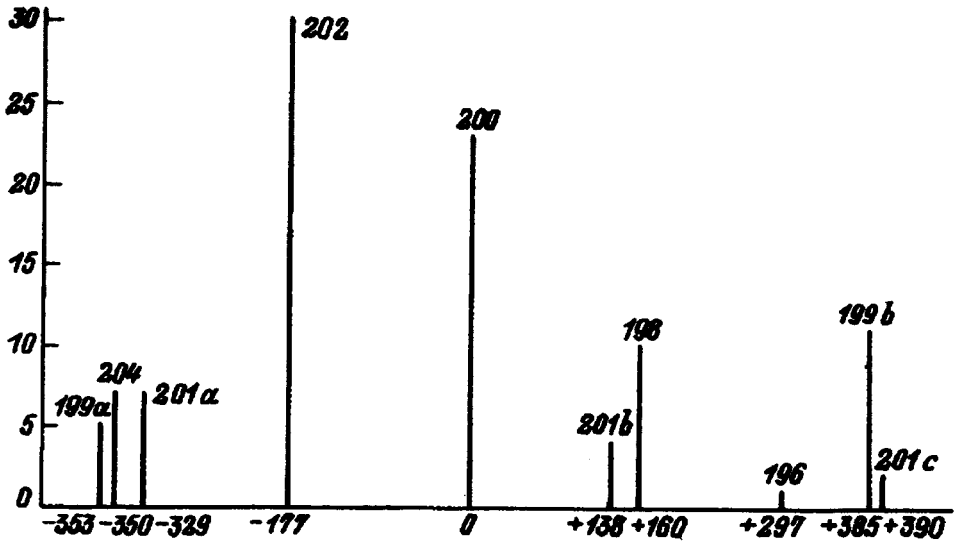

FIG. 1. Hyperfine structure of the 253.7 nm mercury resonance line.

considerably by optimizing the isotopic composition of the mercury in the filter, the type and pressure of the buffer and quencher gas, and the cold point temperature of the filter.

It is difficult and frequently impossible to obtain highly enriched mercury isotopes without using ''negative'' enrichment. Unlike the direct or ''positive'' enrichment process, where radiation from the light source excites and separates the target isotopes, in the reverse or negative enrichment process impurity isotopes are excited and undergo a photochemical reaction. In this case, the valuable product is not only the mercury enriched in the target isotope but often includes mercury enriched in impurity isotopes.

At least two procedures can be proposed to separate all the isotopes of mercury. In the first procedure,[11] mercury of mixed isotopic composition is irradiated by light from a mercury lamp containing Hg-199, Hg-204, or a mixture of these isotopes. The isotopes Hg-199, Hg-201, and Hg-204 are then excited and undergo a photochemical reaction. By exposing this isotope mixture to a lamp containing Hg-198, it is then possible to remove Hg-201. The mixture of Hg-199 and Hg-204 is then separated using a lamp filled with Hg-196 or Hg-204. The isotopes Hg-198, Hg-200, and Hg-202 can then easily be extracted from the mercury depleted in Hg-199, Hg-201, and Hg-204 by using suitable light sources.

In the second procedure[12] the mercury isotopes are separated in several stages. At the first stage, mercury of arbitrary isotopic composition is irradiated by a lamp containing Hg-198 and is separated into a mixture enriched in Hg-198 and Hg-201 and a mixture depleted in these isotopes. The isotopes Hg-198 and Hg-201 are separated using a lamp filled with Hg-199 or Hg-204 or a mixture of these. Using the same light source, a mixture of Hg-199 and Hg-204 isotopes is extracted from the mercury depleted in Hg-198 and Hg-201. The isotopes Hg-199 and Hg-204 may be separated by using light sources containing Hg-204 or Hg-196.

After significantly modernizing the separator system and modifying its process cycle, it was found to be possible to separate all the isotopes of mercury on an acceptable economic basis.

FOTON-M UNIVERSAL AUTOMATIC SEPARATOR AND ITS PROCESS CYCLE

A schematic of the Foton-M separator is shown in Fig. 2.

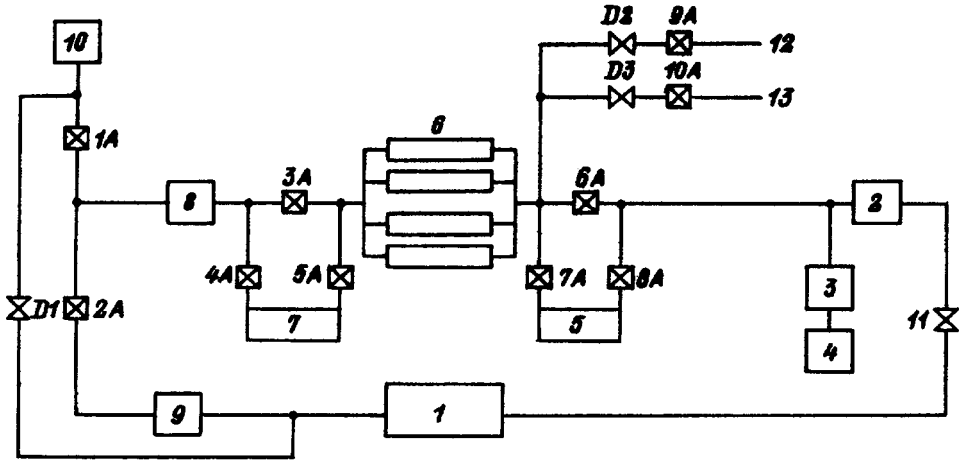

FIG. 2. Schematic of Foton-M separator: *1*—compressor, *2* — reactant flow rate meter, *3*—gas pressure reducer, *4*—cylinder containing reactants, *5*—mercury evaporator, *6*—reaction cells, *7*—isotope collector, *8*—mercury vapor density meter, *9*—mercury vapor trap, *10*—forepump, *1A–10A*—automatically controlled valves, *D1–D3*—valves for fine regulation of the flow, *11*—manual valve, *12*—hydrogen, *13*—oxygen.

In order to obtain highly enriched mercury isotopes, and especially isotopes with overlapping resonance line profiles, the separator must be capable of achieving efficient direct enrichment of an isotope or group of isotopes, negative enrichment, or a combination of these procedures, deep depletion of the mixture in one or several isotopes to as low as fractions of a percent, and there must also be provision for filtering of the lamp radiation if required.

The process cycle of the separator consists of six consecutive stages. At the first stage the circulation loop is filled with a reactant mixture from the cylinder *4* via a gas pressure reducer *3*. The mercury lamps, compressor *1*, and heater of the high-speed evaporator *5* are switched on simultaneously. After the mercury atoms in the reaction cells *6* have reached a given density, recorded by a density meter *8*, the system switches to the stage where enriched mercury oxide builds up on the inner surfaces of the reaction cells. For specific applications, the cells may be connected in parallel, in series, or a combination of both. Mercury depleted in the target isotopes is collected in a mercury vapor trap *9* cooled to between $-50-70\,^{\circ}\mathrm{C}$.

After this stage has been completed, the heater of the evaporator *5* is switched off and the density of the mercury atoms falls by more than an order of magnitude over several minutes. The incorporation of this stage in the process cycle substantially reduces the losses of raw material, which is frequently extremely expensive.

At the next stage, the mercury lamps and compressor are switched off and the reaction cells are purged with hydrogen to remove reactants and residual mercury vapor. The mercury oxide enriched in the target isotope, which has been deposited in the reaction cells, is reduced to the metal in a hydrogen atmosphere by exciting an rf glow discharge in the cells (this process was described in Ref. 10).

At the last stage of the process cycle, the inner surfaces of the reaction cells are cleaned of polymer products (derivatives of 1.3-butadiene) in an oxygen stream by exciting an rf discharge. After this stage has been completed, a new working cycle begins.

Given the technological capabilities of the system, which have already been demonstrated in the production of highly enriched Hg-196 and Hg-202 isotopes and are demonstrated here in the separation of Hg-198, we have every confidence that other mercury isotopes will also be obtained.

SEPARATION OF THE MERCURY ISOTOPE Hg-198 FROM A NATURAL MIXTURE

It is difficult to separate Hg-198 photochemically because the absorption line profiles of Hg-198 and one of the components of Hg-201 in the hyperfine structure of the 253.7 nm resonance line completely overlap under real enrichment process conditions. Nevertheless, a highly concentrated Hg-198 isotope can be obtained from an arbitrary mixture of mercury isotopes by two methods. The first involves carrying out the enrichment process in two stages. At the first stage, mercury of complex isotopic composition, such as a natural mixture, is irradiated by a lamp filled with Hg-198 enriched mercury, and mercury enriched in Hg-198 and Hg-201 is removed. At this stage the valuable products are both the enriched mercury and the mercury depleted in Hg-198 and Hg-201, which may be used as a raw material to obtain other mercury isotopes and also for filling lamps intended for use at the second enrichment stage.

At the second stage, mercury enriched in Hg-198 and Hg-201 is irradiated by a lamp filled with mercury depleted in these isotopes. At this stage the isotope Hg-201 and the other impurity isotopes are separated from the mixture of Hg-198 and Hg-201. The residual Hg-198 enriched product is collected in a trap.

In addition to Hg-198, this two-stage enrichment process yields Hg-201 and also mercury depleted in these isotopes, which is a valuable raw material for obtaining other mercury isotopes.

The second method—a single-stage process—is the simplest for the separation of Hg-198 from a natural mixture. Natural mercury vapor is irradiated by a lamp filled with mercury depleted in Hg-198 and Hg-201. In this case, all the isotopes, except for Hg-198 which is collected in a trap, are excited and undergo a photochemical oxidation reaction. By using several enrichment cycles, it is possible to obtain appreciable quantities of highly enriched Hg-198 at relatively low cost. The high efficiency of producing Hg-198 by this method can be attributed to the enhanced efficiency of the process in each successive cycle compared with the previous one. In addition, the yield increases and the losses of the target isotope decrease as the number of cycles increases.

DIRECT ENRICHMENT OF MERCURY IN ISOTOPES 198 AND 201

The method of direct enrichment of mercury isotopes has been fairly well studied and has so far been used to obtain Hg-196 and Hg-202. However, because of their overlapping absorption lines, enrichment of mercury in the isotopes Hg-198 and Hg-201 has various technological characteristics and dependences.

A lamp filled with mercury enriched in Hg-198 to a concentration of 97.7% was used in experiments to study the direct enrichment of mercury in the isotopes Hg-198 and Hg-201. The body of the lamp was made of a chemically pure, transparent quartz tube with an inner diameter of 14 mm. A water-cooling jacket, also made of quartz glass, was fused onto the tube over its entire length. Leads with oxide cathodes attached to their inner ends were fused onto the ends of the lamp. The operating mode of the lamp was optimized in terms of the main parameters of the enrichment process—the yield and the concentration of target isotopes in the product. The temperature of the cooling jacket was maintained at $22-24\,^{\circ}\mathrm{C}$ and the electrical power consumption of the lamp was around 100 W.

Mercury of natural isotopic composition was used as the raw material in the first enrichment cycle. In this cycle, and particularly in the following cycles, all the experiments to separate Hg-198 and Hg-201 were carried out with the highest possible degree of extraction β of the target isotopes from the raw material. The degree of extraction of the target isotope from the raw material is taken to be the ratio of the yield of this isotope to the amount supplied to the reaction cell during this time.

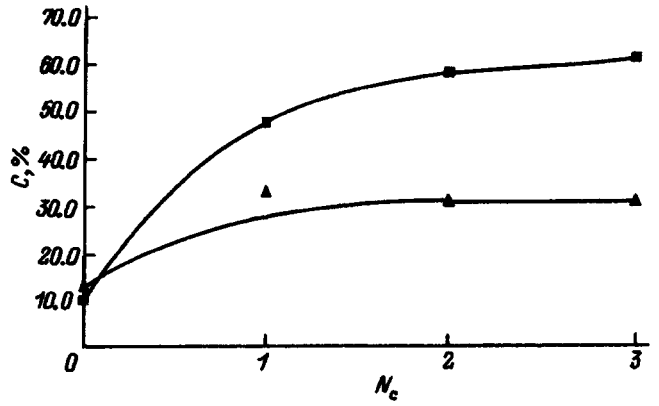

FIG. 3. Isotope concentration versus number of enrichment cycles: ■—Hg-198, ▲—Hg-201, C—isotope concentration, and N_c—cycle number.

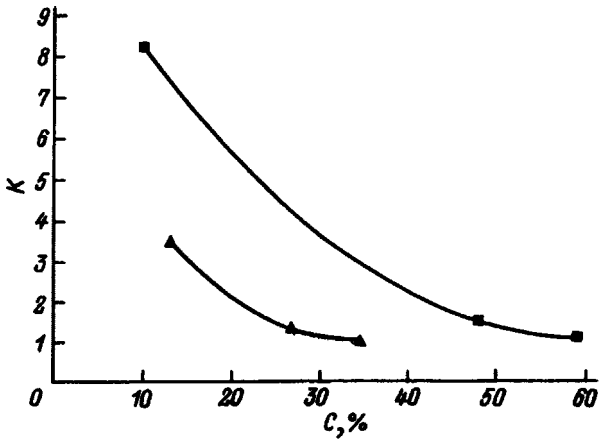

FIG. 4. Separation factor K versus initial isotope concentration: ■—Hg-198 and ▲—Hg-201.

Experience in obtaining the isotopes Hg-196 and Hg-202 by direct enrichment shows that the separation factor α decreases with each successive cycle, i.e., with increasing initial concentration of target isotopes in the raw material. The separation factor α is defined by the expression

$$\alpha=\frac{C_k(1-C_0)}{(1-C_k)C_0},$$

where C_0 is the initial concentration of the target isotope (expressed as a fraction) and C_k is the final concentration of the target isotope (also expressed as a fraction).

Cycling of the process of mercury enrichment in the target isotopes is used to improve their concentration in the product. However, very appreciable losses of target isotopes, as high as 25%, are observed for each enrichment cycle. In experiments which were undertaken to increase the concentration of target isotopes in the final product the losses for each enrichment cycle did not exceed 20%.

The results of these experiments to study the direct enrichment of mercury in the isotopes Hg-198 and Hg-201 were used to plot the concentrations of these isotopes as a function of the number of the enrichment cycle (Fig. 3), and the separation factors α for Hg-198 and Hg-201 as a function of their concentrations C_0 in the raw material (Fig. 4).

An analysis of these results showed that whereas the first and second enrichment cycles are justifiable, the third and subsequent cycles do not produce any appreciable increase in the concentrations of Hg-198 and Hg-201 and thus, if we take into account the losses, are not economically viable.

The concentration of Hg-198 and Hg-201 may be improved by using a mercury-filled lamp with a total concentration of Hg-198, Hg-201, and Hg-204 isotopes of 0.3–0.4% or less in the direct enrichment process or by using an optical filter containing mercury depleted in Hg-198 and Hg-201.

PREPARATION OF MERCURY DEPLETED IN Hg-198 AND Hg-201

A light source containing mercury depleted in Hg-198 and Hg-201 should be used to separate the mixture of these isotopes obtained by the direct enrichment method. The efficiency of the separation process depends to a considerable extent on the ratio of the fraction of light emitted by the impurity isotopes to the fraction in the Hg-198 absorption line. The technology used to separate Hg-198 may be considered to be efficient when the amount of impurity isotopes extracted from the raw material substantially exceeds the extraction of Hg-198. Thus, the preparation of mercury depleted in Hg-198 and Hg-201 to fractions of a percent was identified as a separate problem.

Figure 5 gives the concentrations of Hg-198 and Hg-201 in the depleted mercury as a function of the cycle number N_C. This shows that the rate of decrease in Hg-198 of the raw material exceeds the rate of loss of Hg-201. The most probable reason for this phenomenon is the incomplete matching between the emission line profile of the lamp and the absorption line profile of the Hg-201 component, i.e., the center of the absorption line of the Hg-201 component is irradiated by the wing of the lamp emission line.

Figure 6 gives the yield of the system as a function of the cycle number.

The decrease in the concentration of Hg-198 and Hg-201 isotopes in each cycle means that the depletion process can

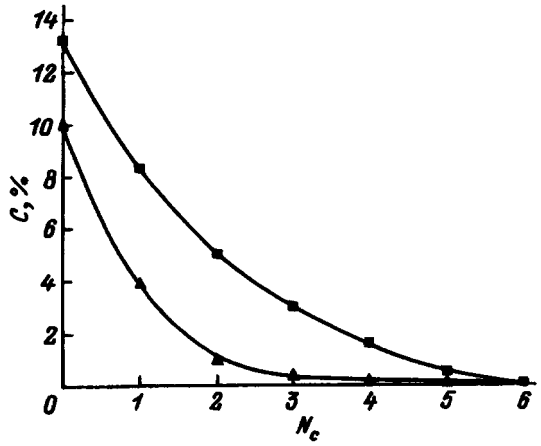

FIG. 5. Isotope concentration C versus number of cycles (N_c is the cycle number): ▲—Hg-198 and ■—Hg-201.

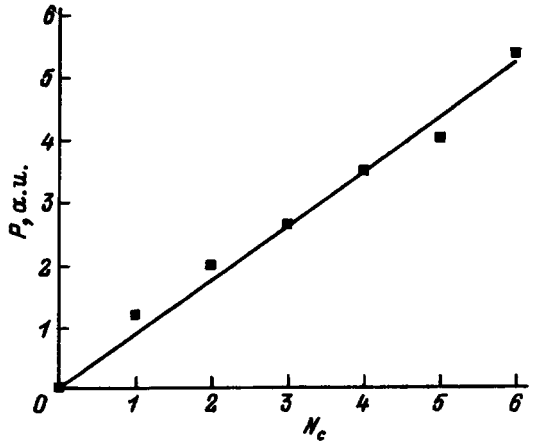

FIG. 6. Separator yield P versus number of cycles.

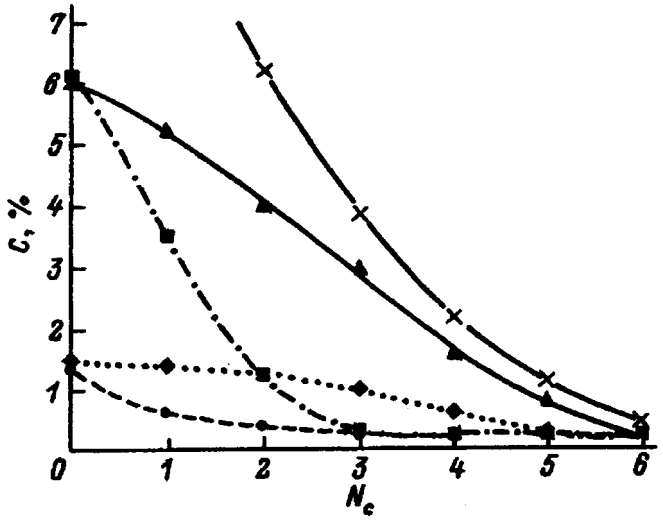

FIG. 7. Concentration C of impurity isotopes versus number of cycles (N_c is the cycle number): ▲—Hg-199, ■—Hg-200, ×—Hg-201, ●—Hg-202, and ♦—Hg-204.

be carried out with an increasing rate of evaporation of raw material and therefore with increasing yield. These data can be used to predict the parameters of the preparation process for mercury depleted in Hg-198 and Hg-201 which may be used as a valuable raw material for the photochemical preparation of Hg-199 and Hg-204 as well as being used in light sources and optical filters designed to separate Hg-198.

Mercury containing less than 0.05% Hg-198 and less than 0.15% Hg-201 was obtained in this experiment. A lamp filled with this mercury was then used to study the negative enrichment of mercury in Hg-198.

SEPARATION OF A MIXTURE OF Hg-198 AND Hg-201 ISOTOPES

The high concentration of impurity isotopes (Hg-199, Hg-200, Hg-202 and Hg-204) in the product obtained by direct enrichment makes it very difficult to obtain highly enriched Hg-201 from this mixture. Thus, only highly enriched Hg-198 was separated at the second stage. The initial raw material for the second stage was mercury having the isotopic composition shown in Table I.

The raw material was irradiated by a lamp filled with mercury depleted in Hg-198 and Hg-201.

At this stage, all the isotopes except for Hg-198 are excited by the mercury lamp and undergo a photochemical oxidation reaction.

Figure 7 shows how the concentrations (C) of impurity isotopes decrease as a function of the cycle number N_c.

The high concentration of Hg-201 relative to the other impurity isotopes in the raw material imposes constraints on the choice of various process parameters such as the density of mercury atoms in the reaction cell. This constraint arises from the need to maximize the extraction of impurity isotopes from the raw material in each cycle.

An analysis of these results reveals that the rates of extraction of impurity isotopes differ and are mainly determined by the light source intensity. The relatively low rate of extraction of the Hg-199 and Hg-204 isotopes is attributed to the insufficiently high concentration of these isotopes in the mercury used to fill the light source.

The yield of the system at this stage of obtaining Hg-198 may be improved after optimizing the isotopic composition of the mercury intended for the light source. In order to equalize the rates of extraction of the impurity isotopes from the raw material, the concentrations of Hg-199 and Hg-204 in the lamp should be increased by reducing the concentration of Hg-200 and especially Hg-202. A typical isotopic composition for this mixture is given in Table II. The preparation of mercury intended for use in a light source to enrich the product of the first stage in Hg-198 does not present any particular difficulties. By using a lamp filled with mercury containing Hg-200 and Hg-202, mercury depleted in Hg-198 and Hg-201 can be depleted in Hg-200 and Hg-202 in a single cycle, with 65% and 80% of these isotopes being extracted from the raw material, respectively.

TABLE I.

Isotope	Concentration, %
196	0.2
198	48.1
199	6.0
200	6.1
201	36.7
202	1.4
204	1.5

TABLE II.

Isotope	Concentration, %
196	0.2
198	<0.05
199	42.0
200	20.00
201	<0.15
202	18.0
204	19.6

TABLE III.

Isotope	Concentration, %
196	0.11
198	0.05
199	21.33
200	26.87
201	0.15
202	41.93
204	9.58

PREPARATION OF Hg-198 BY A SINGLE-STAGE PROCESS

The single-stage process for the production of Hg-198 is also highly promising because it can produce highly enriched Hg-198 and also a mixture of isotopes depleted in Hg-198 which can then be used to obtain mercury enriched in Hg-201.

In the single-stage process a mixture of mercury isotopes of arbitrary isotopic composition, such as a natural mixture, is exposed to light from a lamp containing mercury depleted in Hg-198 and Hg-201. All the isotopes except for Hg-198 are excited and undergo a photochemical oxidation reaction. A lamp filled with mercury having the isotopic composition given in Table III was used in the experiments to separate Hg-198 by this process.

Since the absorption and emission lines of the Hg-199, Hg-201 and Hg-204 isotopes in the hyperfine structure of the 253.7 nm resonance line overlap, the emission from the Hg-199 and Hg-204 isotope components will excite and convert to the oxide not only these isotopes but also Hg-201. In addition to these isotopes, the lamp radiation also excites Hg-200 and Hg-202, and these isotopes should be excited considerably more efficiently than the others because of their high content in the lamp.

One of the main conditions for a high enrichment efficiency is the efficient removal of impurity isotopes from the raw material at each stage. This primarily involves optimizing the operating mode of the lamp. The width of the lamp emission lines should be sufficient to excite two components of the Hg-201 isotope line, the stronger one separated from the emission line of the Hg-204 isotope by 0.03 cm^{-1} and the other weaker one separated from one of the components of the Hg-199 emission line by 0.011 cm^{-1}.

It was established experimentally that optimum conditions are achieved when the lamp temperature is 28 °C and the electrical power consumption of the lamp is around 110 W. Graphs showing the rate of decrease in the impurity isotope concentration (C) as a function of the cycle number are plotted in Fig. 8.

An analysis of these experimental results showed that the rates of decrease in the concentrations of Hg-200 and Hg-202 isotopes in the enriched product considerably exceed the rates of decrease in the concentrations of Hg-199, Hg-204, and especially Hg-201 isotopes. By equalizing the rates of extraction of the impurity isotopes, it will clearly be possible to reduce the number of enrichment cycles and enhance the concentration of Hg-198 in the final product. The degree of extraction of impurity isotopes from the raw material is

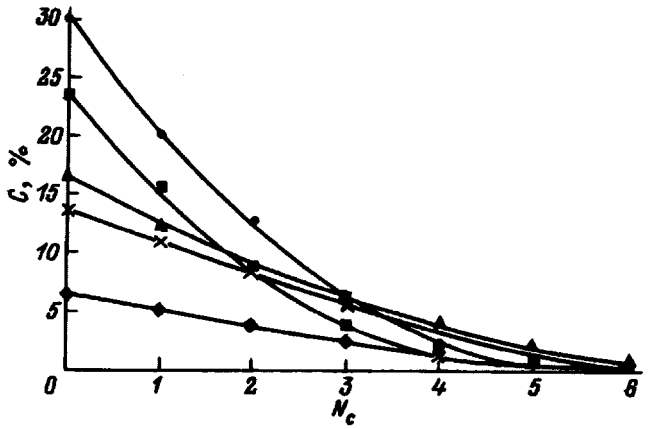

FIG. 8. Concentration C of impurity isotopes versus number of cycles (N_c is the cycle number): ▲—Hg-199, ■—Hg-200, ×—Hg-201, ●—Hg-202, and ♦—Hg-204.

most strongly influenced by their concentration in the lamp. It has been shown that the mercury used to fill the lamp should be depleted in Hg-200 and Hg-202.

Mercury having a low concentration of Hg-198 and Hg-201 was depleted in Hg-200 and Hg-202 in a single cycle using a lamp containing a mixture of Hg-200 and Hg-202 isotopes. The isotopic composition of the mercury obtained as a result of this process is shown in Table IV.

The mercury obtained was used to fill a lamp intended for a new series of experiments to study the enrichment of mercury in Hg-198.

As was to be expected, lamps containing mercury of nearly optimum isotopic composition yielded mercury up to 99.2% enriched in Hg-198 within four enrichment cycles.

The operating efficiency of the system in the negative enrichment regime may be improved substantially by various combinations of the separator cells, for instance, all in series, two in parallel and the others in series, and so on.

CONCLUSIONS

An automatic system has been developed to obtain highly concentrated isotopes by a photochemical method, including isotopes with overlapping resonance line profiles. For the first time this system has produced considerable quantities of the isotopes Hg-196, Hg-202, and Hg-198 with concentrations higher than 96%, 99%, and 99%, respectively. These isotopes were obtained with a high economic efficiency and exceed the demand in Russia.

TABLE IV.

Isotope	Concentration, %
196	0.3
198	0.1
199	42.1
200	15.1
201	0.3
202	23.7
204	18.4

The Foton-M separator may be used to obtain other mercury isotopes with an acceptable economic efficiency.

[1] Yu. V. Vyazovetskiĭ, A. P. Senchenkov, A. M. Shestopalov, and Yu. Yu. Chuvilin, Zh. Tekh. Fiz. **57**, 1643 (1987) [Sov. Phys. Tech. Phys. **32**, 979 (1987)].
[2] B. A. Shlevkov, Med. Radiolog., No. 11, 36 (1968).
[3] J. Maya, M. W. Grossman, R. Lagychenko, and J. F. Waymouth, Science **226**, 435 (1984).
[4] Yu. V. Vyazovetskiĭ, A. P. Senchenkov, and N. R. Stankov, Russian Patent No. 1594749.
[5] N. V. Vyazovetskiĭ, Yu. V. Vyazovetskiĭ, A. P. Senchenkov, and N. R. Stankov, Russian Patent No. 1821994.
[6] J. P. Morand, M. Wacogne, and E. Poth, Energie Nuclaire **10**, 362 (1968).
[7] G. Müller *et al.*, J. Photochem. **13**, 109 (1980).
[8] G. Müller *et al.*, Isotopenpraxis **17**, 200 (1981).
[9] M. Grossman and J. Maya, in *Proceedings of the 13th International Conference on Quantum Electronics*, Anaheim, CA (1984).
[10] A. P. Aenchenkov and V. I. Boloshin, Inventor's Certificate (USSR) No. 714697 [in Russian].
[11] Yu. V. Vyazovetskiĭ, Patent application No. 93035083 (1993); granted July 11, 1995.
[12] Yu. V. Vyazovetskiĭ, Patent application No. 93935095 (1993); granted July 11, 1995.

Translated by R. M. Durham

Potential distribution in a bounded piezoelectric in the presence of the photothermoacoustic effect

D. A. Andrusenko and I. Ya. Kucherov

Taras Shevchenko Kiev University, 252017 Kiev, Ukraine
(Submitted June 17, 1996)
Zh. Tekh. Fiz. **68**, 75–79 (January 1998)

The photothermoacoustic effect in a bounded piezoelectric is studied theoretically and experimentally. Calculations are made of the potential distribution over the thickness and the potential difference in a thin piezoelectric layer as a function of the coordinate of this layer. The amplitude distribution of the potential oscillations over the thickness of the piezoelectric reveals two maxima. The phase of the oscillations at these maxima is shifted by approximately 180°. It is shown that the potential difference on the opposite surfaces of the plate is zero regardless of the propagation constant of the thermal wave. An experiment was carried out using a layered disk formed by seven identical piezoelements. The amplitude and phase shift of the potential difference were investigated for each element at different light modulation frequencies. The experimental results show good agreement with the theory.
 [S1063-7842(98)01401-9]

The photothermoacoustic effect is now widely used to study materials in various aggregate states.[1,2] In this effect, an object is exposed to modulated light, is heated periodically by the absorption of optical energy and generates a thermal wave, which is recorded by some means. The amplitude and phase of the thermal wave carry information on the thermal and optical parameters of the object. A sensitive method of recording this thermal wave involves using a piezoelectric transducer. In this method the piezoelectric transducer records elastic vibrations, which are created in the sample as a result of the thermoelastic effect.[1] Various studies have been devoted to the piezoelectric recording method.[3–5] However, this topic has not been fully analyzed and solutions have only been obtained for various particular cases because of the complexity of the problem. At the same time, the piezoelectric transducer method has various advantages for studying physical processes in solids,[6,7] so that its further development is a matter of some urgency.

It was shown in Refs. 8 and 9 that when a bounded piezoelectric is heated nonuniformly, a complex polarization distribution (with inversion of the sign) is observed. This implies that the magnitude and phase of the signal recorded from the piezoelectric transducer must depend on the geometry of the sample–transducer system, and especially on the ratio of their thicknesses. The potential distribution over the thickness of the piezoelectric transducer in the presence of the photothermoacoustic effect must be known to optimize the thicknesses of the components of the sample–transducer system. This aspect is also of independent interest from the point of view of identifying the characteristics of the photoacoustic effect in piezoelectrics. Here we report theoretical and experimental studies of this topic.

We consider the following geometry (Fig. 1). A piezoelectric transducer in the form of a plate of thickness h is made of a class c_{6v} crystal. The C_6 axis coincides with the z axis of the coordinate system, The surface $z=0$ is uniformly illuminated by light modulated at the frequency ω. The coefficient of absorption of light by the crystal is α. We consider the case $\alpha h \gg 1$, which is closest to the real situation when a thermal wave is recorded via the elastic stresses by means of a piezoelectric transducer. To simplify the problem, we shall assume that the plate is isotropic in terms of elastic and thermal parameters. The thermal wave generated as a result of heating of the piezoelectric by the absorbed light propagates in the direction of the z axis. The plate is free, with vacuum on both sides. We find the distribution of the variable component of the temperature in the plate, neglecting the thermal radiation into the surrounding space. This involves solving the heat conduction equation in the presence of bulk heat sources produced by the absorbed light

$$c\rho\frac{\partial\Theta}{\partial t}-\varkappa\frac{\partial^2\Theta}{\partial z^2}=\alpha P\ \exp(-\alpha z) \tag{1}$$

together with the boundary conditions: the heat fluxes at the sample–vacuum interfaces are zero:

$$\varkappa\frac{\partial\Theta}{\partial z}=0\Big|_{z=0,h}, \tag{2}$$

where Θ is the variable component of the temperature, $P=(P_0/2)(1+\cos\omega t)$ is the optical intensity, c, ρ, and $\varkappa$ are the specific heat, density, and thermal conductivity of the plate material, respectively.

We assume that all the absorbed thermal energy is converted into heat. As in Ref. 1, Eqs. (1) and (2) yield the following expression for the variable component of the temperature

$$\begin{aligned}&\Theta=\Theta_0\{(\alpha/\gamma)[\exp(\gamma z)+\exp(2\gamma h)\exp(-\gamma z)]/\\&\qquad[1-\exp(2\gamma h)]+\exp(-\alpha z)\}\exp(i\omega t),\\&\Theta_0=-\alpha P_0/[2\varkappa(\alpha^2-\gamma^2)];\\&\gamma=l_T^{-1}(1+i);\quad l_T=(2\varkappa/c\rho\omega)^{+1/2}.\end{aligned} \tag{3}$$

Assuming $\alpha h\gg 1$ and $\alpha\gg\gamma$, and omitting the time term, we simplify expression (3) to the form

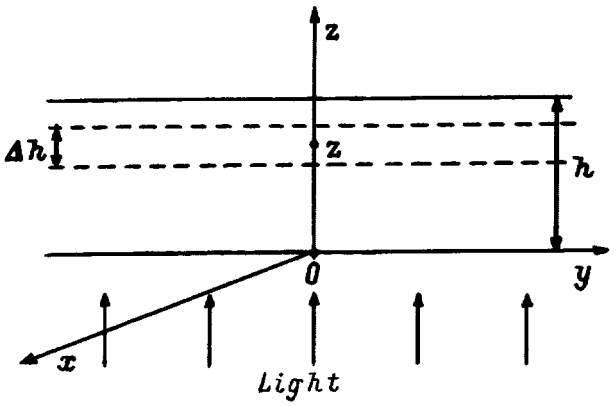

FIG. 1. Geometry of the problem.

$$\Theta = P_0/(2\varkappa\gamma)\{\Theta_1 \exp(-\gamma z) - \Theta_2 \exp(\gamma z)\},$$
$$\Theta_1 = (1-\exp(-2\gamma h))^{-1}, \quad \Theta_2 = (1-\exp(2\gamma h))^{-1}, \tag{4}$$

γ is the propagation constant of the thermal wave, and l_T is the thermal diffusion length.

As a result of the appearance of thermoelastic stresses, this heating leads to electric polarization of the plate material caused by the piezoelectric effect. We consider the thermoelastic part of the problem in the quasistatic approximation, i.e., for the case of relatively low modulation frequencies, when the length of the acoustic waves is much greater than the characteristic dimensions of the sample. We shall also assume that the plate is thin, i.e., its thickness h is much smaller than its other dimensions. In this case, under the action of the thermoelastic forces, the plate undergoes forced oscillations which may be described by the quasistatic theory of thermoelasticity. The deformations are then a certain superposition of the stretching and pure bending of the plate, while the elastic stresses may be expressed in the form[10]

$$T_{11} = T_{22} = [(\alpha_T E)/(1-\nu)]\left\{(1/h)\int_0^h \Theta\, dz + (z-h/2) \times (12/h^3)\int_0^h \Theta(z-h/2)\,dz - \Theta\right\}, \tag{5}$$

where α_T is the coefficient of thermal expansion, E is Young's modulus, and ν is Poisson's ratio.

The temperature Θ is defined by expression (4). Using this expression for T_{11}, we obtain

$$T_{11} = [\alpha_T E/(1-\nu)]\{a_1/h + (12/h^3)(z-h/2)a_2 - \Theta\},$$
$$a_1 = P_0/(2\varkappa\gamma^2), \tag{6}$$
$$a_2 = a_1 h\{(\cosh(\gamma h)-1)/(\gamma h \sinh(\gamma h)) - 1/2\}.$$

The electric field induction D_z in the piezoelectric is given by

$$D_z = 2d_{31}T_{11} + \varepsilon_{33}E_2, \tag{7}$$

where d_{31} and ε_{33} are the piezomodulus and the permittivity, and E_z is the electric field strength.

Since $D_z = 0$ outside the plate, we find E_z from the condition that D_z be continuous using expression (7):

$$E_z = -\frac{2d_{31}}{\varepsilon_{33}} T_{11}. \tag{8}$$

The potential distribution over the thickness of the plate is given by

$$\int_0^z dU = -\int_0^z E_z dz. \tag{9}$$

Integrating expression (9) with allowance for Eqs. (8), (6), and (4), we find the potential distribution $U(z)$ in the plate:

$$U(z) - U(0) = [(d_{31}\alpha_T E P_0)/(\varepsilon_{33}\varkappa\gamma^2(1-\nu))] \times \{z/h + 6((\cosh(\gamma h)-1)/(\gamma h \sinh(\gamma h)) - 1/2)(z^2/h^2 - z/h) + \sinh(\gamma(h-z))/\sinh(\gamma h) - 1\}. \tag{10}$$

It can be seen that the electric potential produced by the piezoelectric effect is a complex function of the coordinate z as a result of the thermoelastic stresses. An important factor here is that the total potential difference between the opposite surfaces of the sample ($z = h$) is zero, regardless of the propagation constant γ of the thermal wave. This result agrees with the data presented in Ref. 6, in which a similar result was obtained for the particular case $\varkappa = 0$, i.e., neglecting effects caused by diffusion of heat.

Formula (10) was used to calculate the amplitude and phase distribution of the oscillations of the potential difference $U(z) - U(0)$ over the thickness of the plate for various thermal diffusion lengths l_T and various light modulation frequencies for a particular sample. The results of these calculations are plotted in Fig. 2. It can be seen from Fig. 2a that the amplitude distribution of the potential oscillations over the thickness of the piezoelectric (Fig. 2a) has two maxima. The phase of the oscillations varies strongly with the coordinate, particularly near the minimum amplitude (Fig. 2b). At the maximum amplitudes of $U(z)$ the phases of the oscillations are shifted by approximately 180°, i.e., the voltage oscillations in these parts of the plate are in antiphase.

These results indicate that if $l_T \ll h$, then for a certain ratio of sample and transducer thicknesses the sample–transducer system may be barely sensitive to the photoacoustic signal. For a more specific estimate of the optimum geometric relations for this system, we need to know the layer-by-layer distribution of the potential difference for this particular case. Thus, we calculated the potential difference ΔU in a thin piezoelectric layer Δh as a function of the z coordinate of this layer:

$$\Delta U = -\int_{z-\Delta h/2}^{z+\Delta h/2} E_z dz = [(\alpha_T E d_{31} P_0)/(\varepsilon_{33}(1-\nu)\varkappa\gamma^2)] \times \{\Delta h/h + 12[(\cosh(\gamma h)-1)/(\gamma h \sinh(\gamma h)) - 1/2](z/h - 1/2)\Delta h/h - 2\cosh(\gamma(h-z)) \times \sinh(\gamma\Delta h/2)/\sinh(\gamma h)\}. \tag{11}$$

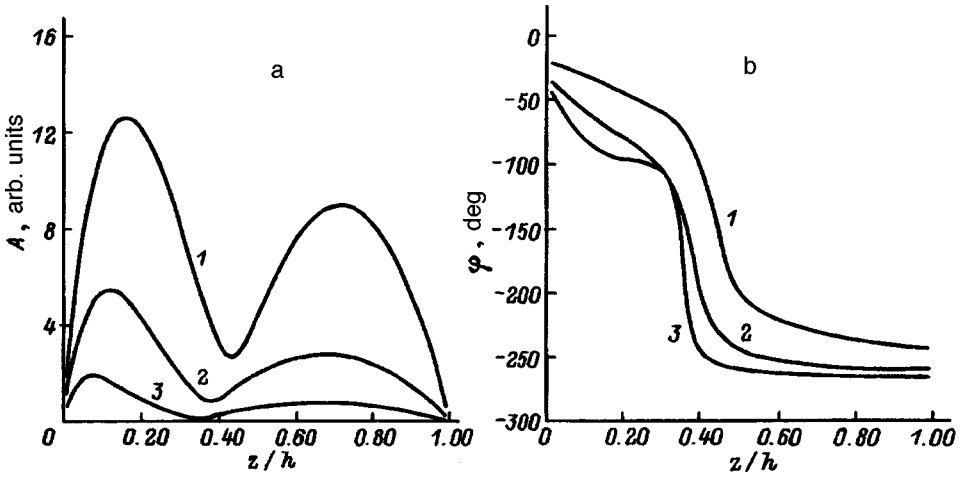

FIG. 2. Distribution of the variable component of the potential amplitude A (a) and phase φ of the oscillations (b) over the thickness of a piezoelectric plate: l_T/h: *1*—1/5, *2*—1/10, and *3*—1/20.

Formula (11) was used to calculate the amplitude and phase shift ΔU as a function of the coordinate z for $\Delta h/h=0.001$ and various values of l_T/h. The results of these calculations are plotted in Fig. 3. It can be seen from Fig. 3a that the amplitude dependence $\Delta U(z)$ has two minima, the first near the illuminated surface and the second closer to the opposite surface. As l_T increases, the minima are shifted to higher z. The phase shift is also a fairly complex function of the z coordinate of the layer (Fig. 3b). Near the illuminated surface, where $l_T/h \ll 1$, the phase changes appreciably (up to 250°) (curves *4–6*). The phase shift caused by the thermal wave is the decisive factor here. For $z > l_T$ the phase shift is mainly determined by the elastic deformations. A characteristic feature here is the phase reversal point, at which the phase shift of the variable component of the plate deformations changes sign. This point corresponds to the position about which the neutral plane oscillates, i.e., the plane in which the variable component of the deformations is zero at a given time. For $l_T/h \ll 1$ (curves *4–6*) the neutral plane is almost fixed and has the coordinate $z_0=2h/3$. In this case, the phase shifts fairly abruptly from $\pi/2$ to $-\pi/2$, passing through zero at $z=z_0$. As l_T increases, the phase reversal point shifts toward higher z and the dependence of the phase shift on the coordinate has a much gentler slope near the reversal point (curves *1* and *2*) as a result of an increase in the amplitude of the neutral plane oscillations.

Although these calculations were made for a homogeneous system, they may also be used for an approximate or qualitative analysis of sample–transducer system if the parameters of the elements are not too different.

For $l_T/h \ll 1$ (this is the most typical case for photoacoustic investigations using piezoelectric signal recording) the differences in the thermophysical parameters of the sample and the transducer do not play a significant role, since the thermal wave does not reach the transducer. Here attention is drawn to two possible cases. One is when the amplitude of the photoacoustic signal is large and depends strongly on the parameters of the system, particularly on l_T, and the other is when the phase of the photoacoustic signal is more sensitive to the parameters of the system. The first case will be found when the thickness of the piezoelectric transducer is equal to approximately 1/3 of the dimensions of the sample–transducer system ($\Delta h/h \simeq 0.33$, $z \simeq 0.83$). It can be seen from Fig. 3a (curves *2–6*) that the signal amplitude will have a maximum for this case and will vary strongly with l_T. However, it can be seen from Fig. 2b (curves *2* and *3*) that the change in the phase shift for a layer of thickness

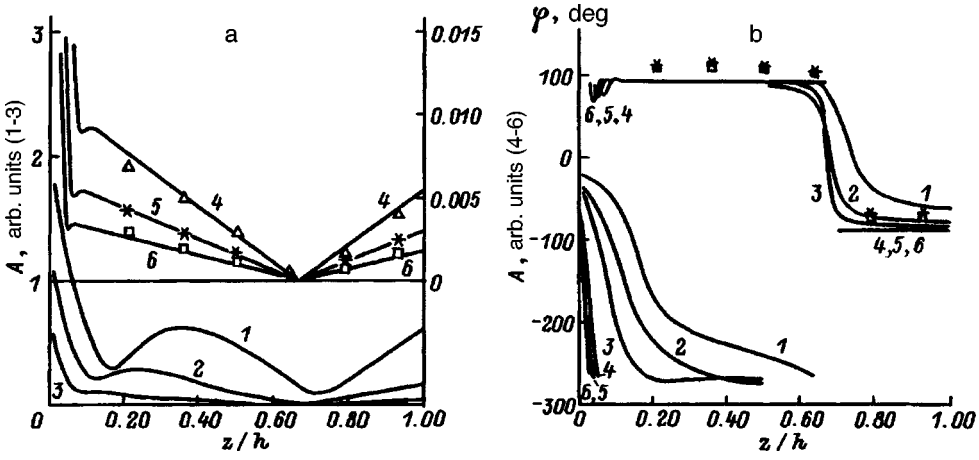

FIG. 3. Amplitude of the potential difference (a) and phase shift (b) for a thin layer Δh of a piezoelectric plate as a function of the coordinate (z/h) of this layer: l_T/h: *1*—1/5, *2*—1/10, *3*—1/20, *4*—1/75, *5*—1/100, and *6*—1/130.

$\Delta h/h \cong 0.3$ near the rear surface is relatively small. For instance, when the parameter l_T/h decreases from 0.1 to 0.05, the amplitude varies fourfold while the phase shift varies by approximately 8°. The second case is achieved when the thickness of the piezoelectric transducer is approximately 2/3 of the thickness of the entire system ($\Delta h/h \simeq 0.68$, $z \simeq 0.68$). It can be seen from Fig. 3 (curves *2–6*) that the amplitude has a minimum for this ratio but the phase shift depends strongly on l_T. For the same variations of the parameter l_T/h the phase of the oscillations varies by approximately 75°.

The experimental investigations were carried out using a layered disk formed by cementing together seven identical piezoelectric elements made of the PZT ceramic TsTS-19, each 0.6 mm thick and 20 mm in diameter (the piezoceramic is described by the same matrix constants as class C_{6v} crystals). One of the end surfaces of the disk was exposed to 60 mW modulated light from an LG-38 laser ($\lambda = 0.63\mu$m). Measurements were made of the amplitude and phase shift of the potential difference for each element of the sample (except for the first) for various light modulation frequencies. No measurements were made for the first layer since, in addition to the piezoelectric effect, the potential difference in this layer is strongly influenced by the potential generated as a result of the pyroeffect (TsTS-19 ceramic is a strong pyroelectric). At the minimum modulation frequency used in the experiments ($f = 42$ Hz), the thermal diffusion length in the ceramic, $l_T \simeq 50\mu$m, was substantially less than the thickness of the elements. Thus, the pyroelectric effect should only be observed in the first layer. Since the pyroelectric effect was not taken into account in the theory, the electric potential was not analyzed in this layer.

The results of these investigations for three modulation frequencies 42 ($l_T/h \simeq 1/75$), 75 ($l_T/h \simeq 1/100$), and 125 Hz ($l_T/h \simeq 1/130$) are shown by the asterisks in Fig. 3. The experimental data for the amplitude were normalized to the theoretical data using the curve corresponding to $f = 75$ Hz at the point which refers to the seventh layer. The data for the phase shift did not need to be normalized. It can be seen that the experimental data describing the potential distribution and its phase shift in a piezoelectric in the presence of the photothermoacoustic effect show reasonable agreement with the theory.

To conclude, a complex potential distribution over the thickness of a piezoelectric is established in the presence of the photothermoacoustic effect. The potential difference between the opposite surfaces of a piezoelectric plate when one is exposed to modulated light is zero, regardless of the propagation constant of the thermal wave and the relation between the thickness of the sample and the thermal wavelength. In measurements of the signal in a double-layer sample–transducer system for a sample whose thickness is substantially greater than the thermal diffusion length (the transducer is positioned at the back of the sample), attention is drawn to two possible cases. When the thickness of the sample is approximately twice that of the transducer, the maximum signal amplitude may be obtained. In this case, the phase of the signal oscillations is close to $-\pi/2$ and is almost independent of the thermal diffusion length. However, when the sample thickness is approximately half that of the transducer, the signal amplitude has a minimum although the phase of the signal oscillations varies substantially with changes in the properties of the sample.

[1] J. C. Rosencwaig, *Photoacoustics and Photoacoustic Spectroscopy* (Wiley, New York, 1980).
[2] V. P. Zharov, and V. S. Letokhov, *Laser Optoacoustic Spectroscopy* [in Russian] Nauka, Moscow (1984).
[3] W. Jackson and N. M. Amer, J. Appl. Phys. **51**, 3343 (1980).
[4] R. P. Vardapetyan, Akust. Zh. **35**, 1026 (1989) [Sov. Phys. Acoust. **35**, 597 (1989)].
[5] Yu. V. Gulyaev, A. I. Morozov, and V. Yu. Raevskiĭ, Akust. Zh. **31**, 469 (1985).
[6] T. Ikari, S. Shigetomi, and Y. Koga, Phys. Rev. B **37**, 886 (1988).
[7] J. Etxebarria, S. Uriate, J. Fernandez *et al.*, J. Phys. C **17**, 6601 (1984).
[8] V. F. Kosorotov, L. S. Kremenchugskiĭ, L. V. Levash *et al.*, Preprint No. 9 [in Russian], Institute of Physics, Academy of Sciences of the Ukrainian SSR, Kiev (1984).
[9] V. F. Kosorotov, L. S. Kremenchugskiĭ, L. V. Levsh *et al.*, Fiz. Tverd. Tela (Leningrad) **26**, 888 (1984) [Sov. Phys. Solid State **26**, 540 (1984); Erratum **26**, 1551 (1984)].
[10] A. D. Kovalenko, *Thermoelasticity* [in Russian], Vishcha Shkola, Kiev (1975).

Translated by R. M. Durham

Electromagnetic excitation of infrasound in a conducting medium

G. A. Lyakhov and N. V. Suyazov

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia
(Submitted July 16, 1996)
Zh. Tekh. Fiz. **68**, 80–83 (January 1998)

A comparative analysis is made of the mechanisms of interaction between the electromagnetic fields of a global resonator and hydrodynamic and acoustic disturbances in a conducting medium. A universal boundary condition at the interface between air and the conducting medium, which takes into account the motion of the electrolyte, is obtained in an explicit analytical form to calculate the long-wavelength electromagnetic fields. The intensity of the electromagnetic field excited by a vertical hydroacoustic wave is estimated together with the efficiency of excitation of infrasonic oscillations of a conducting medium in the field of a global resonator.
 [S1063-7842(98)01501-3]

INTRODUCTION

For electromagnetic oscillations in the ultralow-frequency (ULF) range ($f \leqslant 300$ Hz) the wavelength is comparable with the radius of the Earth $R_E = 6.4 \times 10^3$ km or with the altitude of the ionosphere $h_i \sim 70$ km, so that their propagation is strongly influenced by global resonances—resonant cavities formed between the Earth and the ionosphere.

For this range we typically find very low attenuation in air (around 0.2 dB/Mm at $f \approx 8$ Hz and around 1 dB/Mm at $f \approx 100$ Hz) and a large skin depth δ in electrolytes and other conducting media. The depth of penetration in seawater, for instance, is δ [m] $\approx 200 f^{-1/2}$ [Hz] which means that radio communication can be achieved with objects at depths of hundreds of meters.

Ultralow-frequency electromagnetic oscillations are effective in geophysics — to study the distribution of thunderstorms, electron density profiles in the ionosphere, geomagnetic disturbances, and solar activity.[1]

Ultralow-frequency hydrodynamic disturbances, on the other hand, have a substantial influence on biological objects in the ocean, which has not yet been fully identified. They can reveal characteristic changes at frequencies up to 10 Hz, which show up as fluctuations in the volume scattering of a probe signal caused by the displacement of individual scattering sources.[2,3]

It is therefore important to assess the efficiency of the interaction between ULF electromagnetic oscillations at frequencies close to the natural frequencies of a global resonator, with hydrodynamic, including infrasonic, disturbances in the ocean. This problem can naturally be generalized to the case of a conducting elastic medium of finite conductivity. An obvious practical application is the possibility of recording ULF electromagnetic fields using solid-state acoustic detectors.

MECHANISMS OF INTERACTION BETWEEN ULTRALOW-FREQUENCY ELECTROMAGNETIC OSCILLATIONS AND HYDRODYNAMIC DISTURBANCES OF A CONDUCTING MEDIUM

If, to be specific, we consider an ocean (the generalization is obvious), the first mechanism involves the Earth's static magnetic field $\mathbf{H}_0$, in which the electrical currents induced in the seawater by the alternating electromagnetic field of the global resonator are acted upon by an Ampère force, which then induces hydrodynamic and acoustic ULF disturbances of the water at the frequency of the electromagnetic oscillations. The opposite situation is also possible: the movement of highly conducting seawater induces an alternating electromagnetic field.

Studies of the inverse effect (see Refs. 4–6) have shown that the electromagnetic fields generated by the hydrodynamic motion can be observed: a vertical electric field of 30 μV/m and a magnetic field of the order of 3×10^{-6} Oe/m parallel to the flow and increasing with depth are generated in a slowly varying flow with a velocity of 1 m/s (Ref. 5). Note that these data only refer to the significantly lower-frequency disturbances ($f \lesssim 1$ Hz) or to the disturbances with a horizontal wavelength much shorter than h_i and R_E.

The second interaction mechanism is caused by a difference between the mobilities and masses of the cations and anions in the electrolyte. These ions are entrained differently by the moving solvent, which leads to charge separation, the formation of an electric current, and consequently leads to the generation of an electromagnetic field—the Debye effect[7]—(new characteristics were investigated in Refs. 8 and 9). Conversely, when the ions move under the action of the external field, the overall effect of their friction force with the solvent is uncompensated so that the solvent and the electrolyte move as a single entity.[10,11]

A system of hydrodynamic equations of motion for a multicomponent (solvent, cations, and anions) fluid is used as the initial system to describe these effects:

$$\rho_s d\mathbf{v}/dt = \left[\sum_j n_j V_j - 1\right]\nabla p + \rho_s \mathbf{g} + \sum_j \rho_j \nu_j (\mathbf{v}_j - \mathbf{v}),$$

$$\rho_j d\mathbf{v}_j/dt = -n_j V_j \nabla p + \rho_j \mathbf{g} - \rho_j \nu_j (\mathbf{v}_j - \mathbf{v}) + e_j n_j (\mathbf{E} + \mathbf{v}_j \times \mathbf{H}/c). \quad (1)$$

Here p is the pressure, $\mathbf{E}$ and $\mathbf{H}$ are the electric and magnetic field strengths in the electrolyte, c is the speed of light, $\mathbf{g}$ is the acceleration due to gravity, j is the number of the ion species, ρ_s, $\mathbf{v}$ and ρ_j, $\mathbf{v}_j$ are the average density and velocity

of the solvent and the ions, V_j is the ion volume (taking into account the solvation sheath), ν_j is the coefficient of friction, n_j and e_j are the ion concentration and charge, which satisfy the electroneutrality principle $\Sigma_j e_j n_j = 0$. In the low-frequency approximation $2\pi f \nu_j^{-1} \ll 1$ ($f \sim 10$–10^3 Hz, $\nu_j \sim 10^{13}$ s^{-1}; Ref. 8) for low ($|\mathbf{u}_j| \ll |\mathbf{v}|$) relative ion velocities $\mathbf{u}_j = \mathbf{v}_j - \mathbf{v}$, system (1) yields the expression

$$\mathbf{u}_j \approx \frac{e_j}{m_j \nu_j}\left[\mathbf{E} + \frac{\mathbf{v}\times\mathbf{H}}{c}\right] - \frac{m_j - V_j\rho}{m_j \nu_j}\left[\frac{d\mathbf{v}}{dt} - \mathbf{g}\right], \tag{2}$$

where $\rho = \rho_s + \Sigma_j \rho_j$ is the electrolyte density, and $m_j = \rho_j / n_j$ are the effective ion masses (including the sheath).

We write the equations for the electromagnetic field in the electrolyte, neglecting the displacement current ($\varepsilon' f \ll \sigma$, where ε' is the real part of the permittivity) compared with the conduction current $\mathbf{j} = \Sigma_j e_j n_j \mathbf{u}_j$, and expressing the latter using expression (2)

$$\nabla\times\mathbf{H} = (4\pi\sigma/c)(\mathbf{E} + \mathbf{v}\times\mathbf{H}/c) + (4\pi\gamma/c)(d\mathbf{v}/dt - \mathbf{g}),$$

$$\nabla\times\mathbf{E} = -(1/c)\partial\mathbf{H}/\partial t, \quad \nabla\cdot\mathbf{H} = 0, \tag{3}$$

where $\sigma = \Sigma_j e_j^2 n_j / \nu_j m_j$, $\gamma = -\Sigma_j e_j n_j (m_j - \rho V_j)/\nu_j m_j$.

For sea water with sodium and chlorine ions and 35% salinity the data given in Ref. 8 yield $\sigma \cong 6\times 10^{10}$ s^{-1}, $\gamma \cong 10^{-2}$ g$^{1/2}$ cm$^{-3/2}$.

Summing the system (1) and then substituting the relative velocities $\mathbf{u}_j$ from expression (2) and $\mathbf{E} + (\mathbf{v}\times\mathbf{H})/c$ from the first equation in system (3), we obtain the equation of motion for the electrolyte as a whole. Confining ourselves to the linear approximation with respect to the variable field and the electrolyte velocity, assuming that the disturbances are low-frequency $2\pi f(\beta - \gamma\eta/\sigma) \ll 1$, where $\eta = -\rho^{-1}\Sigma_j e_j n_j \nu_j^{-1}$ and $\beta = \rho^{-1}\Sigma m_j n_j \nu_j^{-1}$ (in particular, for seawater we have $\eta \cong 10^{-2}$ g$^{-1/2}$ cm$^{3/2}$ and $\beta \cong 2\times 10^{-14}$ s), and finally assuming that the magnetic field and velocity depend harmonically on time, we obtain

$$i\omega\mathbf{v} + \nabla p/\rho = -\mathbf{H}_0\times\nabla\times\mathbf{H}/4\pi\rho + (i\eta c\omega/4\pi\sigma)(\nabla\times\mathbf{H}),$$

$$\nabla^2\mathbf{H} - (4\pi i\sigma\omega/c^2)\mathbf{H} = (4\pi\sigma/c)(\nabla\times\mathbf{F}), \quad (\nabla\mathbf{H}) = 0. \tag{4}$$

Here we have $\omega = 2\pi f$, and the contribution of the electrolyte motion to the variation of the effective electromagnetic field is described by the vector

$$\mathbf{F} = \mathbf{H}_0\times\mathbf{v}/c - i\gamma\omega\mathbf{v}/\sigma. \tag{5}$$

The boundary conditions at the air-electrolyte interface in the linear approximation assume continuity of the magnetic field $\mathbf{H}$ and the tangential component of the electric field $\mathbf{E}_\tau = \mathbf{n}(\mathbf{E}\times\mathbf{n})$, where $\mathbf{n}$ is the normal to the interface. In accordance with expression (3), the electric field is related to the magnetic filed by

$$\mathbf{E} = (c/4\pi\sigma)(\nabla\times\mathbf{H}) + \mathbf{F}. \tag{6}$$

The equations for the electromagnetic field in the Earth–ionosphere cavity have the simple form

$$\nabla^2\mathbf{H} + (\omega^2/c^2)\mathbf{H} = 0, \quad \nabla\cdot\mathbf{H} = 0,$$

$$\mathbf{E} = -i(c/\omega)(\nabla\times\mathbf{H}). \tag{7}$$

The boundary with the ionosphere at altitude h_i is characterized by the effective impedance $Z_i = \varepsilon_i^{-1/2} \simeq (i\omega\nu_e)^{1/2}/\omega_e$, where ε_i is the permittivity of the ionosphere, ω_e is the plasma frequency which is determined by the electron density N_e ($\omega_e^2 = 4\pi e^2 N_e/m_e$), and ν_e is the collision frequency. For frequencies $f \simeq 10$ Hz the altitude is $h_i \simeq 70$ km, and $Z_i \simeq 10^{-2} i^{1/2}$ (Ref. 1), so that the ratio of the impedances of seawater and the ionosphere is $Z_0/Z_i \simeq 10^{-3}$ (Refs. 1 and 12).

We estimate the degree of mutual influence of the hydrodynamic and electromagnetic disturbances in seawater, assuming in Eq. (4)

$$|\nabla\times\mathbf{H}| \sim \frac{4\pi\sigma}{c}|\mathbf{F}|.$$

From this it follows that the self-consistent interaction of the subsystems as a result of the Earth's static magnetic field (Ampère mechanism) is characterized by the parameter

$$\varkappa_A = H_0^2\sigma/\rho c^2\omega. \tag{8}$$

For $H_0 \sim 0.3$ Oe at frequencies $f \sim 10$–10^3 Hz and typical parameters of seawater we have $\varkappa_A = 10^{-13}$–10^{-15}.

The coupling of the subsystems by the ion separation (Debye) mechanism is characterized by the parameter

$$\varkappa_A = \eta\gamma\omega\sigma^{-1}. \tag{9}$$

For these conditions we have $\varkappa_A = 10^{-13}$–10^{-11}.

The difference between the frequency dependences in expressions (8) and (9) leads to the existence of a boundary frequency f_b for a given electrolyte for a fixed external magnetic field:

$$f_b \approx H_0\sigma/2\pi|\gamma|c. \tag{10}$$

For frequencies $f < f_b$ the Ampère effect is the dominant mechanism of action of the electromagnetic field on the electrolyte, whereas for $f > f_b$ the Debye mechanism predominates. For seawater in the Earth's magnetic field we have $f_b \cong 10$ Hz.

BOUNDARY CONDITIONS AT THE AIR–ELECTROLYTE INTERFACE — EXCITATION OF OSCILLATIONS IN AN ELECTROMAGNETIC RESONATOR BY THE MOTION OF A CONDUCTING MEDIUM

In the assumed harmonic approximation we have $\mathbf{v}(\mathbf{r},t) = \mathbf{v}\exp(i\omega t - i\mathbf{kr})$, where $\mathbf{k} = \{\mathbf{k}_\tau, k_z\}$, and the z axis is directed vertically upward ($z = 0$ at the interface). Only fairly large-scale motion of the electrolyte may be an effective source of ultralong-wavelength electromagnetic waves. For instance, in the oceanic problem with a global cavity formed by the Earth's surface and the ionosphere, the horizontal scale of this motion, k_τ^{-1}, is considerably greater than the skin depth δ (see Introduction). Let us assume that the depth of the ocean, h_0, also exceeds δ

$$k_\tau^{-1}, h_0[M] \gg 200 f^{-1/2}[\mathrm{Hz}]. \tag{11}$$

If conditions (11) are satisfied, it is natural to use the approximation of horizontal homogeneity to solve the equations (4). In addition, we have $|Z_0/Z_i| \cong 10^{-3} \ll 1$, so that the field in air above the interface at $z=0$ satisfies the condition $|Z_0| \ll |\partial \ln H/\partial z|c/\omega \ll |Z_0|^{-1}$. This allows us to obtain an explicit relation between the horizontal component $\mathbf{E}_\tau$ of the electric field directly above the interface and the amplitude of the velocity $\mathbf{v}$:

$$\mathbf{E}_\tau = -\{(\nabla \times \mathbf{H}_0)_\tau/c + i\gamma\omega\mathbf{v}_\tau/\sigma\}/\{1-(i/2)^{1/2}\delta k_z\}. \tag{12}$$

Relation (12) exhaustively describes, in particular, the influence of the movement of seawater on the electromagnetic field above the water and in the Earth–ionosphere cavity. It specifies a boundary condition at the surface of the water, which refines the condition $\mathbf{E}_\tau = 0$ conventionally used to calculate the fields of a global resonator.

Disturbances at frequencies around 10 Hz can only be caused by acoustic waves with the vertical (to ensure horizontal homogeneity of the amplitude of the disturbance) wave vector $k_\tau \to 0$ (the surface gravitational waves make a small contribution because of their small-scale nature). From this it follows that the mechanism of ion separation is an ineffective source of electromagnetic field generation in a global resonator: the corresponding contribution in Eq. (12) is proportional to the horizontal component of the velocity $\mathbf{v}_\tau$, which tends to zero for $k_\tau \to 0$.

We shall now consider the Ampère mechanism which is associated with the static magnetic field H_0 and vertical acoustic waves. For acoustic disturbances a sound wave reflected from the surface $k_z \to -k_z$ must be taken into account in Eq. (20). With allowance for the phase of the reflection the boundary condition for the electromagnetic field has the form

$$\mathbf{E}_\tau(z=+0) \approx -(v/c)(\mathbf{n} \times \mathbf{H}_{0\tau})/(1+if/f_0), \tag{13}$$

where $\mathbf{H}_{0\tau}$ is the horizontal component of the Earth's static magnetic field, v is the amplitude of the velocity perturbation in the sound wave at the interface, and $f_0 = 2\sigma c_s^2/c^2$ is the frequency for which the skin depth is a quarter of the sound wavelength (c_s is the velocity of sound). For seawater we have $f_0 \cong 2.8$ Hz.

The electromagnetic modes propagating in the Earth–ionosphere cavity have different frequencies and vertical structures:[1,12] $(\omega_l^2 c^{-2} - k_\tau^2)^{1/2} h_i = \pi l$, where the vertical number l is the number of half waves between the surface of the Earth and the ionosphere. Modes with $l \gg 1$ only exist at relatively high frequencies $f \gtrsim 2$ kHz and are strongly attenuated at distances of the order of R_E. At frequencies $f \gtrsim 10$ kHz their power attenuation is 2 dB/Mm, and as the cutoff frequency is approached the attenuation increases to 10–20 dB/Mm at $f = 1$–10 kHz (Ref. 12). Thus for modes with $l \gg 1$ there is no interference between waves repeatedly encircling the globe.[1] The lower frequencies $f \simeq 1$–10^2 Hz correspond to the waveguide TM mode (horizontal magnetic field) with $l=0$, which depends weakly on the vertical coordinate. Because this mode is weakly attenuated, the interference of waves repeatedly encircling the globe produces Shumann resonances[13]

$$\omega_n/c \approx [n(n-1)]^{1/2}/R_E - iZ_i/2h_i.$$

The lowest ($n=1$–3) Shumann frequencies f are 7.8, 13.8, and 17.9 Hz for Q factors of 4.63, 5.76, and 6.56, respectively.[1,12] Here attention is drawn to the fact that for seawater the boundary frequency f_b from Eq. (10) and the frequency $4f_0$ (for which the skin depth is half the infrasound wavelength) are very close to the frequencies of the lowest Shumann resonances.

Expression (13) gives an estimate of the electromagnetic field excited by a hydroacoustic wave. For the Shumann ($n=1$) and vertical ($l=1$) resonances $f=8$ Hz and 2 kHz we take the amplitude of the vertical acoustic wave to be of the order of that typical for sea noise in a band equal to the width of the resonance maximum $\Delta f = f/Q$ ($\Delta f \approx 1.7$ and 400 Hz (Refs. 1 and 12): $v \approx 0.3$ and 0.06 cm/s (this corresponds to spectral noise densities of 3×10^3 and 40 Pa $\cdot$ Hz$^{-1/2}$ (Ref. 3)). An analysis of the equations (7) with the boundary conditions (13) shows that the gain for a vertical resonator caused by the resonant buildup which takes place if the size of the region of acoustic excitation in the ocean is around 10^2 km, is $(1+|R_i|)/(1-|R_i|) \approx 4$ (Ref. 12), where R_i is the coefficient of reflection of an electromagnetic wave by the ionosphere. Expression (13) then gives

$$|E_\tau| \simeq 10^{-4}\ \ \mu\text{V/m}, \quad |H_\tau| \simeq 4\times10^{-15}\ \text{Oe}. \tag{14}$$

For a Shumann resonance, assuming that the excitation region has global dimensions, the resonant gain may be considerably greater: the radial (vertical) component of the field E_R exceeds E_τ given by expression (13) by a factor of $cQ/\omega h_i$. However, the bounded nature of the excitation region reduces E_R by a factor of $(n+1/2)S/2\pi R_E^2$, where S is the area of the excitation region. As a result, for the first Shumann resonance $f \cong 8$ Hz (for an excitation region of 3×10^3 km) we obtain

$$|E_R| \simeq 0.3\ \ \mu\text{V/m}, \quad |H_\tau| \simeq 10^{-11}\ \ \text{Oe}. \tag{15}$$

The amplitudes of the electromagnetic fields estimated in formulas (14) and (15) can be measured although this obviously requires precision methods of processing the received signal at a level below the noise level of the global resonator. For the dominant storm sources, for example, the average amplitudes of the noise fields are $(\langle H_\tau^2 \rangle)^{1/2} \sim 10^{-8}$ Oe in the frequency band of the first Shumann resonance.

EXCITATION OF AN ACOUSTIC FIELD IN A CONDUCTING MEDIUM BY THE ELECTROMAGNETIC FIELDS OF A GLOBAL RESONATOR

In order to estimate the efficiency of the inverse process to that considered, we now assume that the electromagnetic field in the global resonator is given. This corresponds to the situation where the right-hand side of the second equation in (4) vanishes. We consider the frequency range $f < f_b$ from Eq. (10), in which the Ampère mechanism predominates.

The solution of Eq. (4) with a harmonic acoustic field for a conducting layer of depth h determines the required acoustic field in the electrolyte

$$v_z = \frac{i^{3/2}\xi(f_0/f)^{1/2}(H_\tau/H_0)c_s}{1+i(f_0/f)} Z(z). \tag{16}$$

Here $\xi = H_{0\tau}^2/4\pi\rho c_s^2$, R_s is the coefficient of reflection of sound by the underlying surface, and the depth dependence in Eq. (16) describes the function

$$Z(z) = \exp[(1+i)z/\delta] + \frac{\exp(-i\omega z/c_s) + R_s \exp[i\omega(z-2h)/c_s]}{1 - R_s \exp(-2i\omega h/c_s)}.$$

In the oceanic problem the excitation of an infrasonic wave has a low efficiency. It is more promising to use the described excitation of infrasonic waves in a conducting medium for the concentrated reception of regular electromagnetic signals in the range 10–100 Hz. The detection system should incorporate a conducting volume made of a highly elastic, highly conducting material positioned in the field of a permanent magnet with a fairly strong field H_0. For a copper plate in a moderate field $H_0 \simeq 2\times 10^3$ Oe, for example, we find $\xi \simeq 10^{-7}$ so that it follows from Eq. (16) that the acoustic rate of excitation for $H_\tau = 10^{-6}$ Oe (an antenna with a current of 100–200 A at a voltage of 7 kV, having a radiation efficiency of 10^{-3}) is $v_z \simeq 4\times 10^{-11} Q$ [cm/s]. The Q factor of the copper plate may reach $Q \simeq 2\times 10^3$. Under these conditions we find $v_z \simeq 8\times 10^{-8}$ cm/s so that the amplitude of the acoustic displacement at 8 Hz is $a = v_z/2\pi\omega \cong 10^{-8}$ cm. These displacements can be detected by optical interferometric systems.

CONCLUSIONS

Estimates have shown that this effect involving the excitation of infrasonic oscillations of a conducting medium by the electromagnetic fields of a global resonator and the inverse effect involving the generation of ULF electromagnetic waves by the acoustic motion of an electrolyte may be observed with not too stringent constants on the excitation and detection parameters. The next logical step is to optimize the detection parameters and possibly select a more efficient detection system. Attention is therefore drawn to the possible use of sensitive bimodal systems,[14] for some of which the infrasonic frequency range contains known resonances of interest in their own right, where the amplitude of the generated electromagnetic fields may be greater than the mean-square background values, even in a strongly disturbed atmosphere.

This work was partially supported by the Russian Fund for Fundamental Research (Project No. 96-02-17642).

[1] P. V. Bliokh, A. P. Nikolaenko, and Yu. F. Filippov, *Global Electromagnetic Resonances in the Earth–Ionosphere Cavity* [in Russian], Naukova Dumka, Kiev (1977).
[2] *Oceanology, Physics of the Ocean, Vol. 2, Hydrodynamics of the Ocean* [in Russian], edited by V. M. Kamenkovich and A. S. Monin, Nauka, Moscow (1978).
[3] V. A. Mozgovoĭ, Akust. Zh. **33**, 1101 (1987) [Sov. Phys. Acoust. **33**, 640 (1987)].
[4] V. V. Sochel'nikov, *Principles of the Theory of a Natural Electromagnetic Field in the Sea* [in Russian], Gidrometizdat, Leningrad (1979).
[5] *Electromagnetic and Acoustic Processes* [in Russian], Far East University Press, Vladivostok (1987).
[6] J. C. Larsen, J. Mar. Res. **29**, 28 (1971).
[7] P. Debye, J. Chem. Phys. **1**, 13 (1933).
[8] A. L. Virovlyanskiĭ and A. N. Malakhov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **24**, 851 (1981).
[9] G. A. Lyakhov and Yu. P. Svirko, Pis'ma Zh. Tekh. Fiz. **18**(1), 51 (1992) [Sov. Tech. Phys. Lett. **18**, 20 (1992)].
[10] *Physical Acoustics: Principles and Methods*, edited by W. P. Mason [Academic Press, New York, 1968; Mir, Moscow, 1973].
[11] A. L. Virovlyanskiĭ and A. N. Malakhov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **27**, 1349 (1984).
[12] W. Harth, *Theory of Low Frequency Wave Propagation*, Vol. 2 of *CRC Handbook of Atmospherics* (CRC Press, Boca Raton, Fla., 1983).
[13] W. O. Shumann, Angew. Phys. **9**, 373 (1957).
[14] S. M. Osovets, D. A. Ginzburg, V. S. Gurfinkel' *et al.*, Usp. Fiz. Nauk **141**, 103 (1983) [Sov. Phys. Usp. **26**, 801 (1983)].

Translated by R. M. Durham

Two-dimensional scattering of electromagnetic waves from a permeable inclusion in an anisotropic medium

N. P. Zhuk, S. N. Shul'ga, and A. G. Yarovoĭ

Kharkov State University, 310077 Kharkov, Ukraine
(Submitted May 22, 1996)
Zh. Tekh. Fiz. **68**, 84–88 (January 1998)

Integrodifferential equations for the cross section of the scatterer and a collocation method are used to obtain a numerical solution of the problem of scattering of an H-polarized wave by an anisotropic layer of a three-layer dielectric structure. © *1998 American Institute of Physics.* [S1063-7842(98)01601-8]

1. Theoretical modeling of the interaction between harmonic electromagnetic waves and anisotropic objects is of interest for many physical applications, such as waveguide optics, nondestructive testing, and remote sensing. Scattering of waves by anisotropic objects in an unbounded homogeneous isotropic medium was considered, in particular, in Refs. 1–4, where the scatterers were simulated by a homogeneous cylinder of circular[1] and arbitrary[2] cross section, by an inhomogeneous cylinder of arbitrary cross section,[3] and by a three-dimensional inhomogeneity of arbitrary shape.[4] A more complex case where an anisotropic medium surrounds an ideally conducting or inhomogeneous permeable cylinder was investigated in Refs. 5 and 6. A wide range of topics relating to the propagation and scattering of waves in anisotropic waveguiding structures was addressed in Ref. 7. Here we generalize Refs. 5 and 6 to the case where an inhomogeneous permeable cylinder is embedded in a homogeneous layer of a three-layer structure, where the materials forming the layer and the inclusion may be anisotropic. The present study may also be considered to be a generalization of Refs. 8 and 9, where the authors analyzed two-dimensional scattering by an isotropic inhomogeneous scatterer in an isotropic plane-layered medium.

2. The geometry of the problem is shown in Fig. 1. A homogeneous anisotropic layer of thickness h is characterized by the permittivity tensor

$$\hat{\varepsilon}_s=\begin{bmatrix} \varepsilon_{xx}^{(s)} & 0 & 0 \\ 0 & \varepsilon_{yy}^{(s)} & \varepsilon_{yz}^{(s)} \\ 0 & \varepsilon_{zy}^{(s)} & \varepsilon_{zz}^{(s)} \end{bmatrix} \tag{1}$$

with the constant elements $\varepsilon_{xx}^{(s)},\ldots,\varepsilon_{zz}^{(s)}$. The layer is deposited on a homogeneous isotropic substrate having the permittivity ε_c and its upper boundary is with free space. The region S is the cross section of an inhomogeneous anisotropic cylindrical inclusion which is situated entirely within the layer and is oriented along the x axis. The permittivity distribution within the inclusion is characterized by a tensor $\hat{\varepsilon}_p$ similar to (1), whose elements $\varepsilon_{xx}^{(p)},\ldots,\varepsilon_{zz}^{(p)}$ may depend on $\mathbf{r}=(0,y,z)$. These permittivity tensors describe uniaxial crystals with an inclined optic axis lying in the yz plane or a cold electron plasma in a static external magnetic field directed along the x axis. The magnetic permeability is assumed to be 1 everywhere.

To simplify the problem, we assume that the primary electromagnetic field $\mathbf{E}^{\text{in}}$, $\mathbf{H}^{\text{in}}$, generated by harmonic ($\sim e^{-i\omega t}$) external sources in the absence of an inclusion, does not depend on x. Then this field, like the field $\mathbf{E}$, $\mathbf{H}$ generated in the presence of an inclusion, separates into E- and H-polarized components which propagate independently. It follows from the Maxwell equations that the propagation and scattering of the E-polarized component takes place as in an isotropic structure with the permittivity given by the xx elements of the tensors $\hat{\varepsilon}_s$ and $\hat{\varepsilon}_p$. This problem has already been considered in Refs. 8 and 9. Thus we focus our attention on the more complex case of the H-polarization, where anisotropy effects are pronounced both in the layer and in the inclusion.

3. It follows from Maxwell's equations that in this particular case of H polarization, the nonzero components of the electric field E_y and E_z may be expressed in terms of a single nonzero magnetic field component H_x. When the observation point $\mathbf{r}$ is in free space, we have

$$E_y=-\frac{1}{ik_0}\frac{\partial H_x}{\partial z},\quad E_z=\frac{1}{ik_0}\frac{\partial H_x}{\partial y}, \tag{2}$$

but for the region within the anisotropic half space but outside the inclusion, the corresponding formulas are:

$$\begin{aligned} E_y&=-\frac{1}{ik_0a}\left(\varepsilon_{yz}\frac{\partial H_x}{\partial y}+\varepsilon_{zz}\frac{\partial H_x}{\partial z}\right),\\ E_z&=\frac{1}{ik_0a}\left(\varepsilon_{yy}\frac{\partial H_x}{\partial y}+\varepsilon_{zy}\frac{\partial H_x}{\partial z}\right). \end{aligned} \tag{3}$$

Similar expressions for the substrate are obtained formally from Eq. (2) after dividing the right-hand side by ε_c and, for points inside the inclusion, by replacing the subscript s with p on the right-hand side of Eq. (3). Here $k_0=\omega/c$ is the wave number and c is the speed of light in vacuum,

$$a_s=\varepsilon_{yy}^{(s)}\varepsilon_{zz}^{(s)}-\varepsilon_{yz}^{(s)}\varepsilon_{zy}^{(s)},\quad a_p=\varepsilon_{yy}^{(p)}\varepsilon_{zz}^{(p)}-\varepsilon_{yz}^{(p)}\varepsilon_{zy}^{(p)}. \tag{4}$$

Note that in Eqs. (2) and (3) terms containing external field sources have been omitted to simplify the notation. Their absence is strictly justified when the field sources are positioned at infinity, which, for specificity, we shall henceforth assume.

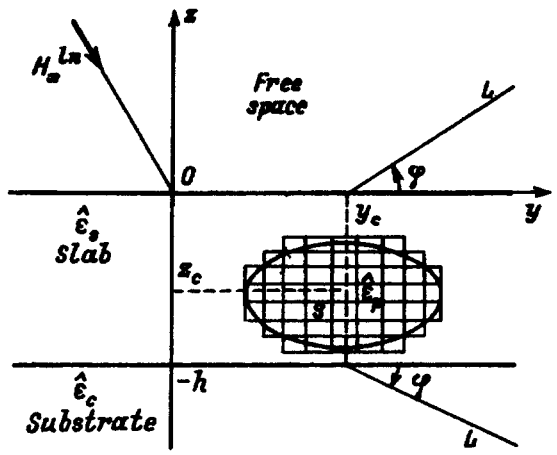

FIG. 1.

After elimination of E_y and E_z, Maxwell's equations yield equations for the main unknown function H_x. In free space and in the substrate, these have the following form, respectively:

$$(\nabla_\perp^2 + k_0^2)H_x = 0, \tag{5}$$

$$(\nabla_\perp^2 + k_c^2)H_x = 0, \tag{6}$$

where $\nabla_\perp$ is the del operator in the yz plane, and $k_c = k_0\sqrt{\varepsilon_c}$ is the wave number in the substrate material.

We write the equation for H_x in the region $-h < z < 0$, assuming that the field within this region in the presence of an inclusion is identical to the field in the absence of an inclusion which would excite electric currents with the volume density

$$\mathbf{J}_p - (ik_0 c/4\pi)[\hat{\varepsilon}_p - \hat{\varepsilon}_s]\mathbf{E}, \tag{7}$$

distributed within S (Ref. 10). Here $\mathbf{E}$ is the (unknown) electric field inside the inclusion. The resulting equations for H_x for $-h < z < 0$ have the standard form

$$\left[\varepsilon_{yy}^{(s)}\frac{\partial^2}{\partial y^2} + \varepsilon_{zz}^{(s)}\frac{\partial^2}{\partial z^2} + (\varepsilon_{yz}^{(s)} + \varepsilon_{zy}^{(s)})\frac{\partial^2}{\partial y\,\partial z} + k_0^2 a_s\right]H_x = a_s q \tag{8}$$

for points positioned both outside and inside the inclusion. The quantity

$$q = \frac{4\pi}{c a_s}\left[\left(\varepsilon_{zy}^{(s)}\frac{\partial}{\partial y} + \varepsilon_{zz}^{(s)}\frac{\partial}{\partial z}\right)J_{py} - \left(\varepsilon_{yy}^{(s)}\frac{\partial}{\partial y} + \varepsilon_{yz}^{(s)}\frac{\partial}{\partial z}\right)J_{pz}\right] \tag{9}$$

has the meaning of secondary sources ''induced'' by the primary field in the inclusion, and J_{py} and J_{pz} have nonzero values outside the region S and are given by

$$J_{py} = -\frac{ik_0 c}{4\pi}(\eta_{yy}E_y + \eta_{yz}E_z),$$
$$J_{pz} = -\frac{ik_0 c}{4\pi}(\eta_{zy}E_y + \eta_{zz}E_z) \tag{10}$$

within S. Here the functions $\eta_{jk} \equiv \varepsilon_{jk}^{(p)} - \varepsilon_{jk}^{(s)}$ determine the electrical contrast of the inclusion relative to the surrounding medium $(j,k = y,z)$. In addition to the formulated equations, the function H_x should obey the condition of continuity at the boundaries of the layer $z = 0, -h$ and should thereby ensure continuity of E_y and E_z.

4. We introduce the solution $G(\mathbf{r},\mathbf{r}')$ of the boundary-value problem for $H_x(\mathbf{r})$ from the previous section, which corresponds to the point source $q(\mathbf{r}) = \delta(\mathbf{r}-\mathbf{r}')$ positioned inside the layer $(-h < z' < 0)$ and satisfies the radiation conditions at infinity $(|\mathbf{r}-\mathbf{r}'| \to +\infty)$. This solution—a Green function—may be obtained by applying a Fourier transformation with respect to $y - y'$ and solving the resultant boundary-value problem with an independent variable z using a standard procedure.[11] The final expressions for $G(\mathbf{r},\mathbf{r}')$ are given in the Appendix.

Having the Green function, we can invert the boundary-value problem for H_x and replace it by the integral relation

$$H_x(\mathbf{r}) = H_x^{\text{in}}(\mathbf{r}) + \int G(\mathbf{r},\mathbf{r}')q(\mathbf{r}')\,d\mathbf{r}', \tag{11}$$

where integration is performed over the entire plane yz.

Taking into account the definition (9) of q and integrating by parts in Eq. (11), we obtain

$$H_x(\mathbf{r}) = H_x^{\text{in}}(\mathbf{r}) + ik_0 F[E_y, E_z](\mathbf{r}). \tag{12}$$

In this relation $\mathbf{r}$ is an arbitrary point in the yz plane,

$$F[E_y,E_z](\mathbf{r}) = \int_S d\mathbf{r}'[L_y(\mathbf{r},\mathbf{r}')E_y(\mathbf{r}') + L_z(\mathbf{r},\mathbf{r}')E_z(\mathbf{r}')], \tag{13}$$

$$a_s L_y(\mathbf{r},\mathbf{r}') = [\varepsilon_{zy}^{(s)}\varepsilon_{yy}^{(p)}(\mathbf{r}') - \varepsilon_{yy}^{(s)}\varepsilon_{zy}^{(p)}(\mathbf{r}')]\partial G(\mathbf{r},\mathbf{r}')/\partial y' + [\varepsilon_{zz}^{(s)}\eta_{yy}(\mathbf{r}') - \varepsilon_{yz}^{(s)}\eta_{zy}(\mathbf{r}')]\partial G(\mathbf{r},\mathbf{r}')/\partial z', \tag{14}$$

$$a_s L_z(\mathbf{r},\mathbf{r}') = [\varepsilon_{zz}^{(s)}\varepsilon_{yz}^{(p)}(\mathbf{r}') - \varepsilon_{yz}^{(s)}\varepsilon_{zz}^{(p)}(\mathbf{r}')]\partial G(\mathbf{r},\mathbf{r}')/\partial z' + [\varepsilon_{zy}^{(s)}\eta_{yz}(\mathbf{r}') - \varepsilon_{yy}^{(s)}\eta_{zz}(\mathbf{r}')]\partial G(\mathbf{r},\mathbf{r}')/\partial y'. \tag{15}$$

We substitute expression (12) for H_x into the direct formulas for E_y and E_z, assuming that the point $\mathbf{r}$ belongs to the region S. As a result, we obtain a system of two coupled integrodifferential equations

$$E_y(\mathbf{r}) = -\frac{1}{a_p(\mathbf{r})}\left[\varepsilon_{yz}^{(p)}(\mathbf{r})\frac{\partial}{\partial y} + \varepsilon_{zz}^{(p)}(\mathbf{r})\frac{\partial}{\partial z}\right] \times [H_x^{\text{in}}(\mathbf{r})/ik_0 + F(\mathbf{r})], \tag{16}$$

$$E_z(\mathbf{r}) = \frac{1}{a_p(\mathbf{r})}\left[\varepsilon_{yy}^{(p)}(\mathbf{r})\frac{\partial}{\partial y} + \varepsilon_{zy}^{(p)}(\mathbf{r})\frac{\partial}{\partial z}\right] \times [H_x^{\text{in}}(\mathbf{r})/ik_0 + F(\mathbf{r})], \tag{17}$$

$(\mathbf{r} \in S)$ for the electric field inside the inclusion. If the solution of these equations is known, formulas (12) and (13) can be used to find the function H_x at any point in the yz plane and the corresponding direct formulas for E_y and E_z can be used to calculate these values everywhere outside the inclusion. The outlined sequence of procedures forms the basis of the algorithm for this scattering problem.

5. We search for a numerical solution of Eqs. (16) and (17) by first using the rectangular mesh shown in Fig. 1 to approximate the cross section of the inclusion by a stepped figure consisting of identical cells having the dimensions $\Delta y \times \Delta z$. We then assumed as an approximation that, within each cell, the unknowns E_y and E_z are constant and the permittivity distribution is homogeneous. After taking the integrodifferential equations (16) and (17) at the center of each cell, we finally obtain a system of linear algebraic equations

$$E_{ym} = \sum_{n=1}^{M} (K_{mn}^{yy} E_{yn} + K_{mn}^{yz} E_{zn}) + Q_m^y, \tag{18}$$

$$E_{zm} = \sum_{n=1}^{M} (K_{mn}^{zy} E_{yn} + K_{mn}^{zz} E_{zn}) + Q_m^z \tag{19}$$

$(m = 1,2,\ldots,M)$. Here E_{ym} and E_{zm} are the unknown values of E_y and E_z in the mth cell and M is the total number of cells approximating S. The explicit form of the coefficients $K_{mn}^{yy}, \ldots, K_{mn}^{zz}$ and the right-hand sides Q_m^y and Q_m^z is easily determined from relations (16) and (17). For instance, we have

$$K_{mn}^{yy} = -\frac{1}{a_p(\mathbf{r}_m)} \left[\varepsilon_{yz}^{(p)}(\mathbf{r}) \frac{\partial}{\partial y} + \varepsilon_{zz}^{(p)}(\mathbf{r}) \frac{\partial}{\partial z} \right] \Bigg|_{\mathbf{r}=\mathbf{r}_m} \times \int_{S_n} d\mathbf{r}' L_y(\mathbf{r},\mathbf{r}'), \tag{20}$$

$$Q_m^y = -\frac{1}{ik_0 a_p(\mathbf{r}_m)} \left[\varepsilon_{yz}^{(p)}(\mathbf{r}) \frac{\partial}{\partial y} + \varepsilon_{zz}^{(p)}(\mathbf{r}) \frac{\partial}{\partial z} \right] \times H_x^{\text{in}}(\mathbf{r})|_{\mathbf{r}=\mathbf{r}_m}, \tag{21}$$

where S_n is the interior of the nth cell and $\mathbf{r}_m$ is the central point in the mth cell.

The coefficients K_{mn}^{jk} $(j,k = y,z)$ are calculated as follows. Using a suitable expression for the Green function given in the Appendix, the values of L_y and L_z from formulas (14) and (15), and then the coefficients K_{mn}^{jk} may be expressed as a sum of two components (for instance, $K_{mn}^{jk} = K_{mn}^{jk(0)} + K_{mn}^{jk(r)}$), the first describing the scattering from an inclusion in an unbounded medium with the parameters of an anisotropic layer and the second taking into account the finite thickness of the layer. The numerical calculations of $K_{mn}^{jk(0)}$ which reduce to calculation of the linear integrals over the contour of the cell, are described in Refs. 9 and 12. The integrals with respect to $\mathbf{r}'$ within the nth cell encountered in calculations of the coefficients $K_{mn}^{jk(r)}$ can simply be replaced by the product of the integrand taken at the center $\mathbf{r}' = \mathbf{r}_n$ of the cell and the area $\Delta y \Delta z$ of this cell. Finally the Sommerfeld integrals, which appear in the expressions for $K_{mn}^{jk(r)}$ because of the Fourier integral on the right-hand side of formula (A2) for the correction caused by reflections $G^{(r)}$, are calculated by shifting the original integration contour from the real axis into the complex plane $\varkappa$. Details of the numerical implementation of this procedure are given in Ref. 9.

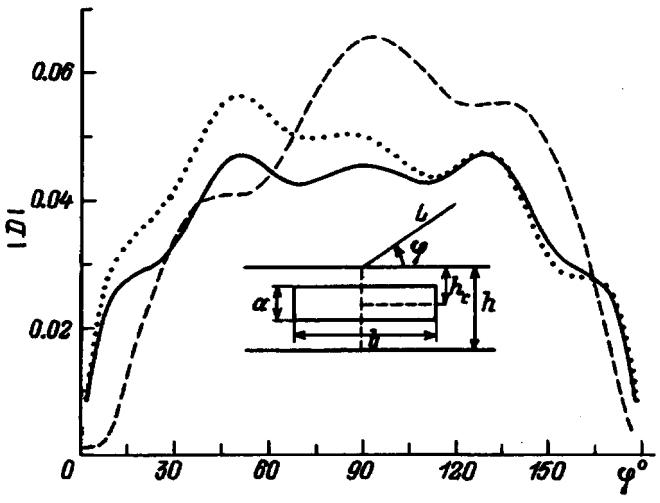

FIG. 2.

The system of linear algebraic equations (18) and (19) for the following examples was solved numerically by the Gauss method.

6. Figures 2–4 illustrate a numerical solution of the problem for a scattered plane H-polarized wave of unit amplitude $A^{\text{in}} = 1$ incident from the upper half space normal to the layer. The layer of electrical thickness $k_0 h = 2.0$ is situated in a vacuum (i.e., $\varepsilon_c = 1.0$). The rectangular inclusion (shown in the inset to Fig. 2) has the dimensions $k_0 a = 0.4$, $k_0 b = 0.8$, and its center of symmetry is located at the distance $k_0 h_c = 2.0$ from the upper boundary of the layer. For the numerical calculations, the inclusion was divided into 72 cells (twelve horizontal divisions and six vertical divisions). In these figures the solid curve refers to the base model of an isotropic inclusion with the permittivity $\varepsilon_c = 2.0$, $\varepsilon_p = 4.0$, respectively. The dashed curve corresponds to the case which differs from the previous one in that the layer material is a uniaxial dielectric. The principal values of the permittivity of this dielectric (i.e., the values along the optic axis and in the transverse direction) are $\varepsilon_{\parallel} = 3.0$, $\varepsilon_{\perp} = 2.0$ and the optic axis lying in the yz plane is deflected downward at the

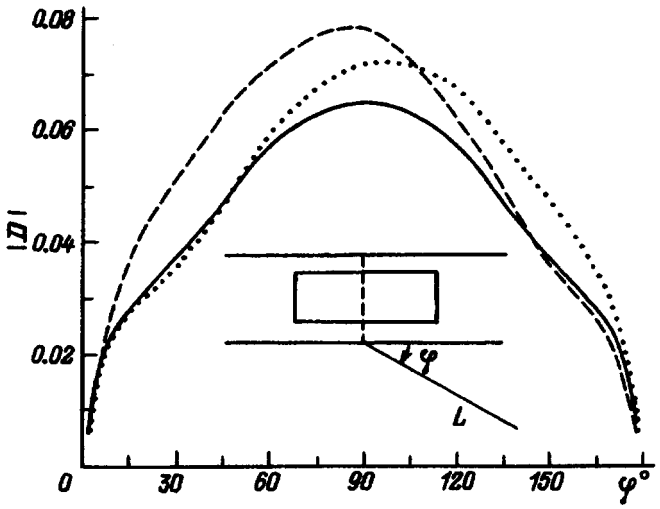

FIG. 3.

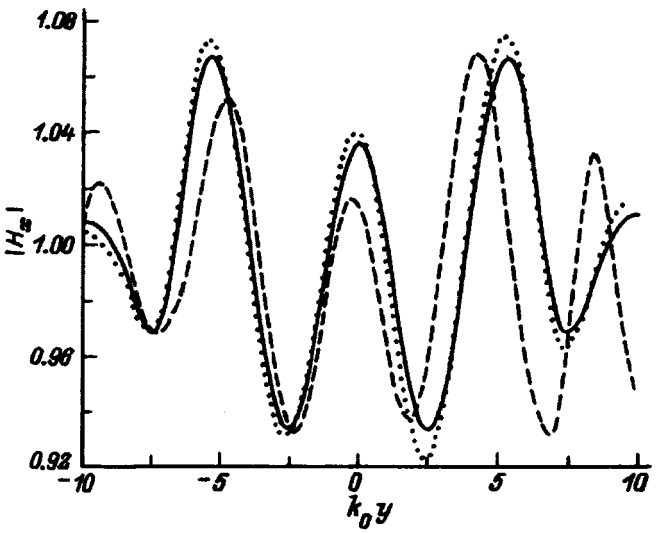

FIG. 4.

angle $\varphi_0=30°$ relative to the y axis. The elements of the layer permittivity tensor $\hat{\varepsilon}_s$ in which we are interested have the following values

$$\varepsilon_{yy}^{(s)}=\varepsilon_{\parallel}^{(s)}\cos^2\varphi_0+\varepsilon_{\perp}^{(s)}\sin^2\varphi_0=2.25,$$
$$\varepsilon_{zz}^{(s)}=\varepsilon_{\parallel}^{(s)}\sin^2\varphi_0+\varepsilon_{\perp}^{(s)}\cos^2\varphi_0=2.75, \tag{22}$$
$$\varepsilon_{yz}^{(s)}=\varepsilon_{zy}^{(s)}=(\varepsilon_{\perp}^{(s)}-\varepsilon_{\parallel}^{(s)})\sin\varphi_0\cos\varphi_0=-0.43.$$

The dotted curve describes the additional case where the layer material is an isotropic dielectric with the permittivity $\varepsilon_s=2.0$ and the inclusion is a uniaxial dielectric for which the principal values of the permittivity are $\varepsilon_{\parallel}^{(p)}=5.0$, $\varepsilon_{\perp}^{(p)}=4.0$, and the optic axis lies in the yz plane and is inclined downward at the angle $\varphi_0=30°$ relative to the y axis. The elements of the permittivity tensor $\hat{\varepsilon}_p$ are calculated using formulas similar to (22) and are given by

$$\varepsilon_{yy}^{(p)}=4.25,\quad \varepsilon_{zz}^{(p)}=4.75,\quad \varepsilon_{yz}^{(p)}=\varepsilon_{zy}^{(p)}=-0.43. \tag{23}$$

Figures 2 and 3 show the angular distribution of the field calculated in the upper half space (Fig. 2). This distribution is characterized by the value of $D(\varphi)$ which is determined by the asymptotic formula for the x component of the scattered magnetic field in the far-field zone outside the layer

$$H_x^{sc}(\mathbf{r})=A^{\text{in}}D(\varphi)\frac{e^{ik_0L}}{\sqrt{L}} \tag{24}$$

($k_0L\gg1$). The polar coordinates L, φ for the observation point in the upper or lower half space are defined in Fig. 1. In this particular case of a rectangular inclusion, the point y_c, z_c coincides with the center of symmetry of this inclusion. Figures 2 and 3 clearly demonstrate how anisotropy in the layer or inclusion influences the formation of the scattered field.

Figure 4 shows the distribution of the total field $|H_x|$ at the upper boundary of the layer. As was to be expected, the figure is symmetric relative to the point $y=0$ for an isotropic inclusion and layer since a plane wave is incident normal to the layer. The oscillatory behavior of these curves reflects the significant role of interference effects in the formation of the scattered and total fields. With increasing distance from the inclusion along the surface of the layer, these oscillations become smoother and the field $|H_x|$ reaches the constant value $|H_x^{\text{in}}|$.

APPENDIX

If the observation point is located in the layer $(-h<z<0)$, then $G(\mathbf{r},\mathbf{r}')$ is expressed as a sum $G=G^{(0)}+G^{(r)}$ of the Green function $G^{(0)}(\mathbf{r},\mathbf{r}')$ for an unbounded medium with the permittivity of the layer and the correction $G^{(r)}(\mathbf{r},\mathbf{r}')$, which takes into account the reflection from the boundaries $z=0,-h$. We then have

$$G^{(0)}(\mathbf{r},\mathbf{r}')=\frac{a_s}{4i\sqrt{\nu}}H_0^{(1)}(k_0s), \tag{A1}$$

$$G^{(r)}(\mathbf{r},\mathbf{r}')=\frac{a_s}{2\pi i\varepsilon_{zz}^{(s)}}\int_{-\infty}^{+\infty}\frac{d\varkappa}{(\gamma_+-\gamma_-)P}\,e^{i\varkappa(y-y')}$$
$$\times[\exp(i\gamma_+z)AR+\exp(i\gamma_-z)BQ], \tag{A2}$$

where $H_0^{(1)}$ is a Hankel function of the first kind,

$$\nu=\varepsilon_{yy}^{(s)}\varepsilon_{zz}^{(s)}-(1/4)(\varepsilon_{yz}^{(s)}+\varepsilon_{zy}^{(s)})^2, \tag{A3}$$
$$s^2=(a_s/\nu)[\varepsilon_{yy}^{(s)}(z-z')^2+\varepsilon_{zz}^{(s)}(y-y')^2-(\varepsilon_{yz}^{(s)}+\varepsilon_{zy}^{(s)})$$
$$\times(y-y')(z-z')], \tag{A4}$$
$$\gamma_{\pm}=-\varkappa\frac{\varepsilon_{yz}^{(s)}+\varepsilon_{zy}^{(s)}}{2\varepsilon_{zz}^{(s)}}\pm\left[\frac{k_0^2a_s-\varkappa^2\varepsilon_{yy}^{(s)}}{\varepsilon_{zz}^{(s)}}\right.$$
$$\left.+\varkappa^2\left(\frac{\varepsilon_{yz}^{(s)}+\varepsilon_{zy}^{(s)}}{2\varepsilon_{zz}^{(s)}}\right)^2\right]^{1/2}. \tag{A5}$$

$$A=e^{i(\gamma_+-\gamma_-)h}[e^{-i\gamma-z'}+Qe^{-i\gamma_+z'}], \tag{A6}$$
$$B=e^{-i\gamma_+z'}+\text{Re}^{i(\gamma_+-\gamma_-)h-i\gamma_-z'}, \tag{A7}$$
$$P=1-QR\exp[i(\gamma_+-\gamma_-)h], \tag{A8}$$
$$Q=-\frac{\varepsilon_{zz}^{(s)}\gamma_++\varepsilon_{yz}^{(s)}\varkappa-\gamma_0a_s}{\varepsilon_{zz}^{(s)}\gamma_-+\varepsilon_{yz}^{(s)}\varkappa-\gamma_0a_s}, \tag{A9}$$
$$R=-\frac{\varepsilon_{zz}^{(s)}\gamma_-+\varepsilon_{yz}^{(s)}\varkappa+\gamma_ca_s/\varepsilon_c}{\varepsilon_{zz}^{(s)}\gamma_++\varepsilon_{yz}^{(s)}\varkappa+\gamma_ca_s/\varepsilon_c}, \tag{A10}$$
$$\gamma_0=(k_0^2-\varkappa^2)^{1/2},\quad \gamma_c=(k_0^2\varepsilon_c-\varkappa^2)^{1/2}, \tag{A11}$$

($0\leqslant\arg\gamma_0,\gamma_c<\pi$). If the observation point is outside the layer, then $G(\mathbf{r},\mathbf{r}')$ is expressed as the Fourier integral

$$G(\mathbf{r},\mathbf{r}')=\frac{a_s}{2\pi i\varepsilon_{zz}^{(s)}}$$
$$\times\int_{-\infty}^{+\infty}\frac{d\varkappa}{(\gamma_+-\gamma_-)P}e^{i\varkappa(y-y')}F(\varkappa,z,z'), \tag{A12}$$

where for $z>0$

$$F(\varkappa,z,z')=(1+Q)e^{i\gamma_0 z}$$
$$\times[e^{-i\gamma_+(z'+h)}+\mathrm{Re}^{-i\gamma_-(z'+h)}], \qquad \text{(A13)}$$

and for $z<-h$

$$F(\varkappa,z,z')=(1+R)e^{-i\gamma_c(z+h)}[e^{-i\gamma_- z'}+Qe^{-i\gamma_+ z'}]. \qquad \text{(A14)}$$

Note that in these formulas it is assumed that $-h<z'<0$.

[1] J. C. Monzon, IEEE Trans. Antennas Propag. **AP-25**, 670 (1987).
[2] B. Beker, K. R. Umashankar, and A. Taflove, IEEE Trans. Antennas Propag. **AP-37**, 1573 (1989).
[3] R. D. Graglia and P. L. E. Uslenghi, IEEE Trans. Antennas Propag. **AP-32**, 867 (1984); **AP-35**, 225 (1987).
[4] R. D. Graglia, P. L. E. Uslenghi, and R. S. Zich, Proc. IEEE **37**, 750 (1989).
[5] J. C. Monzon, IEEE Trans. Antennas Propag. **AP-36**, 1401 (1988).
[6] O. V. Bagatskaya, N. P. Zhuk, and S. N. Shul'ga, Radiotekh. Elektron. **40**, 869 (1995).
[7] E. P. Kurushin and E. I. Nefedov, *Electrodynamics of Anisotropic Structures* [in Russian], Nauka, Moscow (1980).
[8] N. P. Zhuk and A. G. Yarovoĭ, Zh. Tekh. Fiz. **62**(7), 1 (1992) [Sov. Phys. Tech. Phys. **37**, 741 (1992)].
[9] N. P. Zhuck and A. G. Yarovoy, IEEE Trans. Antennas Propag. **AP-42**, 16 (1994).
[10] N. A. Khizhnyak, *Integral Equations of Macroscopic Electrodynamics* [in Russian], Naukova Dumka, Kiev (1986).
[11] A. N. Tikhonov and A. A. Samarskii, *Equations of Mathematical Physics* (Pergamon Press, Oxford, 1964) [Russian original, 4th ed., Nauka, Moscow, 1972].
[12] C. C. Su, IEEE Trans. Antennas Propag. **AP-35**, 1418 (1987).

Translated by R. M. Durham

Transition radiation of a charge in media with a nonuniform potential

V. L. Fal'ko, S. I. Khankina, and V. M. Yakovenko

Institute of Radiophysics and Electronics, Ukrainian National Academy of Sciences, 310085 Kharkov, Ukraine

I. V. Yakovenko

Molniya Scientific-Research and Design Institute, Kharkov Physicotechnical Institute, 310013 Kharkov, Ukraine

(Submitted May 28, 1996)
Zh. Tekh. Fiz. **68**, 11–14 (January 1998)

The influence of a potential barrier on the transition radiation in the form of volume and surface electromagnetic waves emitted by a charged particle crossing an interface between media is investigated. It is shown that the volume-wave radiation field arises not only as a result of the jump in the dielectric constant at the boundary but also as a result of the velocity jump and the reflection of an electron induced by the presence of a nonuniform potential barrier. The angular distribution of the transition radiation intensity is obtained. © *1998 American Institute of Physics.* [S1063-7842(98)00201-3]

1. The radiation emitted by charged particles on crossing an interface between media with different electromagnetic properties has been the subject of a large number of publications (see the references cited in Ref. 1 and, for example, Refs. 2–7). This phenomenon is of interest because transition radiation is encountered quite often in very diverse problems related to astrophysics, accelerator physics, plasma physics, and solid-state physics.

The presence of a potential barrier at the interface between two media is usually not taken into account in studies of transition radiation. Nevertheless, its role is very significant. This was shown, for example, in Refs. 3–5 in studies of the interaction of charged particles with surface plasmons.

In the present work the features of the transition radiation of electromagnetic waves by a particle are investigated with consideration of the influence of the potential barrier U on the interface between two media.

Let a charged particle move uniformly and linearly in medium 1 (for example, in a vacuum, $z<0$) with a velocity v_1 along a normal (the z axis) toward the interface between the media. It is assumed that $U(z)$ has the form

$U(z)=0 \quad \text{for } -\infty<z<0,$

$U(z)=U_0 \quad \text{for } z\geqslant 0$

and that the height of the wall U_0 is smaller than the kinetic energy of the particle in the vacuum $E=(m_0v_1^2)/2$. Then the velocity of the particle in medium 2 ($z\geqslant 0$) equals

$$v_2=\sqrt{\frac{2(E-U_0)}{m}}; \quad \mathbf{v}_2 \| z. \tag{2}$$

The reflection coefficient F of the particle from the barrier is determined from the Schrödinger equation and the boundary conditions in Ref. 8:

$$F=\left(\frac{v_1-v_2}{v_1+v_2}\right)^2. \tag{3}$$

In medium 1 a current is induced by the particle as it moves toward the wall or after it has been reflected from it,

$$\mathbf{j}_1=e\mathbf{v}_1\delta(\boldsymbol{\rho})[\delta(z-v_1t)-F\delta(z+v_1t)], \tag{4}$$

and the current in medium 2 is induced by the particle after it has passed over the barrier

$$\mathbf{j}_2=De\mathbf{v}_1\delta(\boldsymbol{\rho})\delta(z-v_2t). \tag{5}$$

Here $D=1-F$ is the transmission coefficient of the particle over the barrier, and $\boldsymbol{\rho}$ is a vector in the plane of the interface. The electromagnetic field in each of the media is determined from Maxwell's equations, in which the current of charged particles is assigned by (4) or (5). The boundary conditions are the conditions of continuity of the tangential components of the electric (**E**) and magnetic (**H**) fields in the $z=0$ plane at the interface and the condition of radiation at $z=\pm\infty$. Because of the axial symmetry of Maxwell's equations in an isotropic medium with a current along the z axis, it is convenient to introduce a cylindrical system of the coordinates ρ, φ, and z, in which the TM (H_φ, E_ρ, E_z) and TE (E_φ, H_ρ, H_z) modes propagate independently. Only TM waves are excited by a charged particle moving along the z axis. We represent the dependences of the field components of this wave on the time in the form of an expansion in Fourier integrals and their dependences on ρ in terms of Fourier-Bessel integrals:

$$E_z(\rho,z,\omega)=\int_0^\infty \varkappa E_z(z,\varkappa)J_0(\varkappa\rho)d\varkappa, \tag{6}$$

$$E_\rho(\rho,z,\omega)=\int_0^\infty E_\rho(z,\varkappa)J_1(\varkappa\rho)d\varkappa, \tag{7}$$

where $J_n(\varkappa\rho)$ is a Bessel function of order n.

The relation between the magnetic (H_φ) and electric (E_ρ) fields is defined by the equation

$$\frac{\partial H_\varphi}{\partial z}=\frac{i\omega c}{\varepsilon_i}E_\rho, \tag{8}$$

where $i=1,2$ labels the medium, and ε_i is the dielectric constant of the ith medium.

The delta function $\delta(\boldsymbol{\rho})$ can be written in terms of a Bessel function:

$$\delta(\boldsymbol{\rho})=\frac{1}{2\pi}\int_0^\infty \varkappa J_0(\varkappa\rho)d\varkappa. \tag{9}$$

As a result we find that the components of the electric and magnetic fields have the form

$$E_z^{(i)}=\int_0^\infty d\varkappa\varkappa J_0(\varkappa\rho)\Bigg\{A^{(i)}\exp(-i\lambda_i z)+B^{(i)}\exp(i\lambda_i z)+\frac{ie}{\pi\omega}(1-\beta_i^2\varepsilon_i)f_i\Bigg[C^{(i)}\exp\left(i\frac{\omega}{v_i}z\right)+D^{(i)}\times\exp\left(-i\frac{\omega}{v_i}z\right)\Bigg]\Bigg\}, \tag{10}$$

$$H_\varphi^{(i)}=-i\frac{\omega}{c}\varepsilon_i\int_0^\infty d\varkappa J_1(\varkappa\rho)\Bigg\{A^{(i)}\exp(-i\lambda_i z)+B^{(i)}\exp(i\lambda_i z)-i\frac{e\varkappa^2 v_i^2}{\pi\omega^3}f_i\Bigg[C^{(i)}\exp\left(i\frac{\omega}{v_i}z\right)+D^{(i)}\exp\left(-i\frac{\omega}{v_i}z\right)\Bigg]\Bigg\}. \tag{11}$$

The expression for the field components $E_\rho^{(i)}$ is easily obtained from Eqs. (8) and (11). In (10) and (11) we introduced the following notation: the coefficients $A^{(i)}$, $B^{(i)}$, $C^{(i)}$, and $D^{(i)}$ in media 1 and 2 equal, respectively,

$$A^{(1)}=A(\varkappa),\quad B^{(1)}=0,\quad C^{(1)}=1,\quad D^{(1)}=-F, \tag{12}$$

$$A^{(2)}=0,\quad B^{(2)}=B(\varkappa),\quad C^{(2)}=\frac{v_1}{v_2}(1-F),\quad D^{(2)}=0,$$

$$\lambda_i=\sqrt{\frac{\omega^2}{c^2}\varepsilon_i-\varkappa^2}\ (\mathrm{Re}\lambda_i>0); \tag{13}$$

$$f_i=\frac{1}{\varepsilon_i(\beta_i^2\varepsilon_i-(\varkappa^2 v_i^2/\omega^2)-1)};\quad \beta_i=\frac{v_i}{c}. \tag{14}$$

The coefficients $A(\varkappa)$ and $B(\varkappa)$ in the expressions are found from the boundary conditions on the interface between the media at $z=0$. They specify the transition radiation field. The coefficient $A(\varkappa)$ corresponds to a wave propagating along $z<0$, and $B(\varkappa)$ corresponds to a wave propagating along $z>0$. We are interested in the radiation field in medium 1, which is described by the first terms in formulas (10) and (11):

$$A(\varkappa)=\frac{ie\varkappa^2 v_1}{\pi\omega^2\Delta(\omega,\varkappa)}\Bigg\{(f_1-f_2)\varepsilon_2-\frac{1}{\omega}\sqrt{\frac{\omega^2}{c^2}\varepsilon_2-\varkappa^2}\times(f_1\varepsilon_1 v_1+f_2\varepsilon_2 v_2)+F\Bigg[(f_1+f_2)\varepsilon_2+\frac{1}{\omega}\sqrt{\frac{\omega^2}{c^2}\varepsilon_2-\varkappa^2}(f_1\varepsilon_1 v_1-f_2\varepsilon_2 v_2)\Bigg]\Bigg\}. \tag{15}$$

Here

$$\Delta(\omega,\varkappa)=\varepsilon_2\sqrt{\frac{\omega^2}{c^2}\varepsilon_1-\varkappa^2}+\varepsilon_1\sqrt{\frac{\omega^2}{c^2}\varepsilon_2-\varkappa^2}. \tag{16}$$

We note that the term proportional to F in (15) appears as a result of reflection of the particle from the potential barrier (1).

Let us consider media with different values of the dielectric constant $\varepsilon_{1,2}(\omega)>0$. The radiation field in medium 1 is obtained using the stationary-phase method. This radiation has the form of a spherical wave, in which the field components equal

$E_\rho(\omega)=E(\omega)\cos\Theta,\quad E_z=E(\omega)\sin\Theta,$

$$E(\omega)=\frac{e\beta_1\cos\Theta\,\sin\Theta}{\pi c[\varepsilon_2\cos\Theta+\sqrt{\varepsilon_1(\varepsilon_2-\varepsilon_1\sin^2\Theta)}]}\times\frac{\exp(i(\omega/c)\sqrt{\varepsilon_1}R)}{R}\times\Bigg\{\frac{(\varepsilon_2-\varepsilon_1)(1+\beta_1\sqrt{\varepsilon_2-\varepsilon_1\sin^2\Theta}-\varepsilon_1\beta_1^2)}{(1-\varepsilon_1\beta_1^2\cos^2\Theta)(1+\beta_1\sqrt{\varepsilon_2-\varepsilon_1\sin^2\Theta})}+\frac{\varepsilon_1(\beta_2-\beta_1)\sqrt{\varepsilon_2-\varepsilon_1\sin^2\Theta}}{(1+\beta_1\sqrt{\varepsilon_2-\varepsilon_1\sin^2\Theta})(1+\beta_2\sqrt{\varepsilon_2-\varepsilon_1\sin^2\Theta})}+F\Bigg(\frac{\varepsilon_2+\varepsilon_1\beta_1\sqrt{\varepsilon_2-\varepsilon_1\sin^2\Theta}}{1-\varepsilon_1\beta_1^2\cos^2\Theta}+\frac{\varepsilon_1}{1+\beta_2\sqrt{\varepsilon_2-\varepsilon_1\sin^2\Theta}}\Bigg)\Bigg\}. \tag{17}$$

Here we have introduced the angle Θ and the distance R from the point of contact of the particle with the interface between the media at $z=0$ to the point of observation of the radiation in medium 1 such that $\mathbf{R}=\boldsymbol{\rho}\sin\Theta-\mathbf{i}z\cos\Theta$ ($\mathbf{i}$ is a unit vector parallel to the z axis); it is assumed that the condition $(\omega/c)R\gg 1$ holds. The radiated energy flux (17) in the solid-angle element $d\Omega=\sin\Theta d\Theta d\varphi$ is easily calculated from the formula

$$\frac{d^2W}{d\Omega d\omega}=cR^2|E(\omega)|^2. \tag{18}$$

It is seen from expression (17) that the radiation field consists of three parts. The first is the radiation which is caused by the jump in the dielectric constants on the boundary and exists in the absence of a potential barrier ($U_0=0$). The second part describes the radiation caused by the velocity jump on the boundary ($U_0\neq 0$) without consideration of electron reflection from the potential barrier. The third term specifies the fraction of the radiation associated with propagation of the de Broglie wave ''reflected'' from the boundary.

In the case of an infinitely high barrier ($U_0\to\infty$, $F=1$) we obtain

$$E(\omega)=\frac{2e\varepsilon_2\beta_1\cos\Theta\,\sin\Theta}{\pi c[\varepsilon_2\cos\Theta+\sqrt{\varepsilon_1(\varepsilon_2-\varepsilon_1\sin^2\Theta)}]}$$

$$\times\frac{1}{(1-\varepsilon_1\beta_1^2\cos^2\Theta)}\,\frac{\exp(i(\omega/c)\sqrt{\varepsilon_1}R)}{R}. \quad (19)$$

The expression for the radiation field in the absence of a potential barrier ($U_0=0$) is known.[1] We note that in this case ($U_0=0$) the amplitude of the field and the radiated energy are smaller than in the presence of an infinitely high potential barrier ($U_0\to\infty$). For example, if medium 2 is an ideal conductor($\varepsilon_2\to\infty$), the field $E(\omega)$ (Ref. 1) is two times smaller than the radiation field of a particle in the presence of an infinitely high potential, and the energy flux differs by a factor of 4.

Let us assume that a particle moves in a semiconductor with a $p-n$ junction, at which the bottom of the conduction band can be described using a potential barrier of form (1) (U_0 is a finite quantity). Since the dielectric constant is determined only by the properties of the crystal lattice, we must set $\varepsilon_1=\varepsilon_2=\varepsilon$ in Eqs. (17) and (18). The radiation field in such a semiconductor has the form

$$E(\omega)=\frac{e\sin\Theta}{2\pi c(1+\beta_2\sqrt{\varepsilon}\cos\Theta)}\,\frac{\exp(i(\omega/c)\sqrt{\varepsilon}R)}{R}\times\left[\frac{(\beta_2-\beta_1)\sqrt{\varepsilon}\cos\Theta}{1+\beta_1\sqrt{\varepsilon}\cos\Theta}+F\frac{2+(\beta_2-\beta_1)\sqrt{\varepsilon}\cos\Theta}{1-\beta_1\sqrt{\varepsilon}\cos\Theta}\right]. \quad (20)$$

It can be seen from (20) that the angular distribution of the intensity varies and that, in comparison to the classical case ($U_0=0$, and $\varepsilon_1\neq\varepsilon_2$), the radiation pattern is ''pressed'' against the $z=0$ plane. It should be noted that in the general case [see (17)] the angular distribution of the field $E(\omega)$ is characterized by the presence of a sharp maximum, which appears in the vicinity of values of the angle Θ for which the condition for the Vavilov–Cherenkov effect is satisfied in medium 1:

$$\cos^2\Theta=\frac{1}{\varepsilon_1\beta_1^2}$$

(here we are dealing with a maximum, rather than a singularity, since damping of the wave in the medium must be taken into account under real conditions).

This feature is present both in the first term (it is associated with reflection of the Cherenkov radiation produced by the particle as it moves in the positive direction of the z axis away from the boundary at $z=0$) and in the third term (the Cherenkov radiation in the same medium caused by the particle after reflection from the barrier). In a semiconductor with a $p-n$ junction the maximum in the field distribution $E(\Theta)$ (20) is determined only by the Vavilov–Cherenkov radiation of the particle after reflection from the boundary.

As we know, surface waves can propagate on an interface between media, if the dielectric constant of one of the media has a negative value. Let us assume that $\varepsilon_2<0$ and $|\varepsilon_2|>\varepsilon_1$. In this case the function $\Delta(\omega,\varkappa)$ (16) vanishes when

$$\varkappa=\varkappa_p=\frac{\omega}{c}\sqrt{\frac{\varepsilon_1|\varepsilon_2|}{|\varepsilon_2|-\varepsilon_1}}. \quad (21)$$

Equation (21) is the dispersion relation of a surface polariton. The contribution from pole (21) describes the transition radiation field of a cylindrical surface wave

$$E_z(\omega)=E\exp\left(-\left|\varkappa\right|_p\sqrt{\frac{\varepsilon_1}{|\varepsilon_2|}}z\right)\sqrt{\frac{|\varkappa_p|}{\rho}}\exp(i\varkappa_p\rho),$$

$$E_\rho(\omega)=i\sqrt{\frac{\varepsilon_1}{|\varepsilon_2|}}E_z(\omega),$$

$$H_\varphi(\omega)=-\sqrt{\frac{\varepsilon_1(|\varepsilon_2|-\varepsilon_1)}{|\varepsilon_2|}}E_z(\omega),$$

$$E=-\sqrt{\frac{2}{\pi}}\frac{e\beta_1}{c}\times\frac{|\varepsilon_2|^{5/2}\varepsilon_1^{3/2}}{(|\varepsilon_2|-\varepsilon_1+\beta_1^2\varepsilon_1^2)(|\varepsilon_2|-\varepsilon_1+\beta_2^2\varepsilon_2^2)(\varepsilon_2^2-\varepsilon_1^2)}\times\exp\left(i\frac{\pi}{4}\right)\left\{(1+F)\left[\frac{|\varepsilon_2|-\varepsilon_1+\beta_2^2\varepsilon_2^2}{\varepsilon_1}+i\beta_2\frac{|\varepsilon_2|-\varepsilon_1+\beta_1^2\varepsilon_1^2}{\sqrt{|\varepsilon_2|-\varepsilon_1}}\right]+(-1F)\left[\frac{|\varepsilon_2|-\varepsilon_1+\beta_1^2\varepsilon_1^2}{|\varepsilon_2|}+i\beta_1\frac{|\varepsilon_2|-\varepsilon_1+\beta_2^2\varepsilon_2^2}{\sqrt{|\varepsilon_2|-\varepsilon_1}}\right]\right\}. \quad (22)$$

This result refers to the case where the pole $\varkappa_p$ and the stationary-phase point

$$\varkappa_s=\frac{\omega}{c}\sqrt{(\varepsilon_1)}\sin\Theta$$

are far enough from one another that their contributions to integrals (10) and (11) can be treated independently.

The energy flux of wave (22) through a circular area $(\rho,\ \rho+d\rho)$ at $z=0$ equals

$$\frac{\partial^2 W}{\partial\rho\partial\omega}=2\pi\omega\frac{\varepsilon_1^{3/2}}{|\varepsilon_2|^{1/2}}|E|^2. \quad (23)$$

The energy of a cylindrical wave in the absence of a potential barrier ($U_0=0$, $F=0$) (Ref. 1) is $4\varepsilon_2^2/(|\varepsilon_2|+\varepsilon_1)^2$ times smaller than the energy in the case of a perfectly reflecting boundary ($U_0\to\infty$, $F=1$).

Thus, the radiation field in medium 1 is formed by a spherical volume wave and a cylindrical surface wave. The spherical wave forms at large distances from the point of contact of the particle with the interface between the media [$R\gg c/(\omega\sqrt{\varepsilon_1})$], as follows from the condition for applicability of the stationary-phase method), and its intensity is distributed in the range of angles $0<\Theta<\pi/2$. The cylindrical wave propagates along the interface between the media ($\Theta=\pi/2$) and damps at a depth

$$L \sim \frac{c}{\omega\sqrt{\varepsilon_1}}\sqrt{\frac{|\varepsilon_2|-\varepsilon_1}{\varepsilon_1}}.$$

In the region of values of the frequency ω and the angle Θ (near $\pi/2$) which satisfy the conditions

$$|\pi/2-\Theta|<2\sqrt{\frac{2c}{R\omega\sqrt{\varepsilon_1}}}\ll 1, \qquad |\varepsilon_2|\ll\varepsilon_1, \tag{24}$$

the distance between the pole and the stationary-phase point becomes smaller than the width of the lines of the features of the integrands in (10) and (11). Then the van der Waerden method[9] should be used to calculate the integrals in (10) and (11). We shall not present the expressions for the radiation field because of their cumbersome nature. We note that a cylindrical surface wave and a spherical volume wave exist in the range of angles (24), but their amplitudes are small because of this inequality.

Let us next investigate the radiation of a moving charged particle in a homogeneous medium with a potential barrier in the form of a rectangle or a δ function. Such a potential can appear, for example, in a semiconductor medium due to the presence of an impurity or a defect. The reflection coefficient of the particle in this case can be represented in the following manner:

$$F=\frac{\psi(E,U_0)}{1+\psi(E,U_0)}, \tag{25}$$

where the form of the function $\psi(E,U_0)$ is determined by the form of the potential $U(z)$.

In the case of a rectangular barrier of width a

$$U(z)=\begin{cases} 0 & (z<0), \\ U_0 & (0<z<a), \\ 0 & (a<z), \end{cases} \tag{26}$$

the function $\psi(E,U_0)$ equals[8]

$$\psi=\frac{U_0^2}{4E(U_0-E)}\sinh^2\frac{a}{\hbar}\sqrt{2m(U_0-E)}\,(U_0>E), \tag{27}$$

$$\psi=\frac{U_0^2}{4E(E-U_0)}\sin^2\frac{a}{\hbar}\sqrt{2m(E-U_0)}\,(E>U_0). \tag{28}$$

If the potential has the form of a δ function, i.e.,

$$U(z)=V_0\delta(z), \tag{29}$$

then

$$\psi=\frac{mV_0^2}{2E\hbar^2}. \tag{30}$$

The expression for ψ can be obtained from formula (27) if $U_0\gg E$ and $(a\sqrt{2mU_0})/\hbar\ll 1$, where $V_0=U_0a$.

The radiation field of the particle is a spherical wave and can be described in the region $z<0$ by (17) and (18), if we set $\varepsilon_1=\varepsilon_2=\varepsilon$ and $\beta_1=\beta_2=\beta$ in them, and the reflection coefficient $F=F(E,U_0)$ can be found from formulas (25)–(30):

$$E(\omega)=\frac{e\beta\,\sin\,\Theta}{\pi c}F\frac{\exp(i\,(\omega/c)\sqrt{\varepsilon}R)}{R}. \tag{31}$$

If $E>U_0$, then

$$\frac{d^2W}{d\Theta\,d\omega}=\frac{e^2\beta^2U_0^4\,\sin^3\,\Theta}{8E^2(E-U_0)^2}\sin^4\frac{a}{\hbar}\sqrt{2m(E-U_0)}. \tag{32}$$

The radiated intensity oscillates, vanishing under the condition $(a/\hbar)p=\pi n(p=\sqrt{2m(E-U_0)})$, i.e., when a whole number of half de Broglie wavelengths $\lambda_D=(2\pi\hbar)/p$ fit into the barrier width. In this case there is some analogy with the transition radiation of a particle passing through a thin isotropic insulating plate, which exhibits oscillations as a result of variation of the ratio of the thickness of the plate to the wavelength of the charge $\lambda_e=(2\pi v)/\omega$.

We note that the phenomenon considered here can be utilized in the spectroscopy of solids.

[1] V. L. Ginzburg and V. N. Tsytovich, *Transition Radiation and Transition Scattering*, Hilger, New York–Bristol (1990).
[2] I. A. Anisimov, I. Yu. Kotlyarov, and S. M. Levitskiĭ, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **32**, 1034 (1989).
[3] A. M. Belenov, P. N. Luskinovich, and A. G. Sobolev *et al.*, Zh. Tekh. Fiz. **56**, 1902 (1986) [Sov. Phys. Tech. Phys. **31**, 1138 (1986)].
[4] M. V. Burtyka, V. M. Yakovenko, and I. V. Yakovenko, Fiz. Nizk. Temp. **21**, 628 (1995) [Low Temp. Phys. **21**, 489 (1995)].
[5] V. M. Yakovenko and I. V. Yakovenko, Phys. Lett. A **196**, 290 (1994).
[6] V. V. Krechetov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **38**, 639 (1995).
[7] I. I. Kalikinskiĭ, Zh. Tekh. Fiz. **61** (9), 20 (1991) [Sov. Phys. Tech. **36**, 978 (1991)]; **65** (10), 131 (1995) [Phys. Tech. Phys. **40**, 1047 (1995)].
[8] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, 3rd ed., Pergamon Press, Oxford (1977).
[9] É. A. Kaner and V. M. Yakovenko, Zh. Éksp. Teor. Fiz. **42**, 471 (1962) [Sov. Phys. JETP **15**, 330 (1962)].

Translated by P. Shelnitz

Influence of local dispersion on transient processes accompanying the generation of rf radiation by an electromagnetic shock wave

A. M. Belyantsev and A. B. Kozyrev

Institute of Physics of Microstructures, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia
(Submitted July 15, 1996)
Zh. Tekh. Fiz. **68**, 89–95 (January 1998)

Transient processes accompanying the conversion of a video pulse into a radio pulse in a nonlinear transmission line having hysteretic properties are studied. It is established that the transition process leading to the establishment of ''steady-state'' (close in amplitude) oscillations has a minimum when the electromagnetic shock wave front is phase-matched with the wave excited by it at a frequency near the minimum local dispersion of the group velocity. [S1063-7842(98)01701-2]

INTRODUCTION

Electromagnetic shock waves and processes involving the propagation of electromagnetic radiation in nonlinear media (transmission lines) have been studied for some time. The main fundamental principles and various possible applications associated with the propagation of electromagnetic signals in dispersive nonlinear media can be found in the first reviews and books.[1–4]

Recently published studies have examined the generation of rf oscillations by direct conversion of a video pulse into a radio pulse during its propagation in dispersive nonlinear media (transmission lines).[5–8]

In Refs. 5 and 6 the authors discuss the generation of a train of solitons in LC lines with a nonlinear capacitance varying as some function $C(V)$, and in dispersive periodic structures (media) similar to an LC line. However, this method of generating rf oscillations has several significant disadvantages.[1] The amplitude of the generated oscillations decays rapidly in the train, the spectrum of the generated train is fairly wide because of the nonmonochromatic nature of the generation process, and the number of generated oscillations rapidly saturates as the video pulse propagates in a nonlinear transmission line.[5,6]

The authors of Refs. 7 and 8 proposed a significantly more effective mechanism for direct conversion of a video pulse into a radio pulse with a quasimonochromatic carrier (filling) during its propagation in a dispersive nonlinear medium (transmission line). In this mechanism, the front of an intense electromagnetic shock wave excites a synchronously traveling rf wave. It was shown in these studies that for monochromatic generation, the nonlinear medium behind the front should saturate and remain saturated for a long time (compared with the period of the excited wave). Under these conditions, rf perturbations of finite amplitude will propagate behind the electromagnetic shock wave front, as in a linear medium. It is observed that the generation may be efficient and the carrier frequency may be stable because of the phase matching with the shock front and the wave excited by it. An undoubted advantage of this approach compared with soliton generation is that first, the generated oscillations are almost monochromatic, and second, the decay of the amplitude of the steady-state oscillations in the radio pulse is only caused by rf losses in the linear transmission line. It is also possible to control the frequency of the generated oscillations by varying the initial conditions. It was noted in these studies that in addition to the hysteretic dependence of the nonlinearity, the dispersion of the medium (transmission line) plays an important role in the generation of monochromatic oscillations.

Here we investigate the transient processes accompanying the conversion of a video pulse into a radio pulse in a nonlinear transmission line of finite length which exhibits hysteretic properties, and we determine the constraints on the dispersion from the point of view of minimizing the transition process leading to the establishment of ''steady-state'' (close in amplitude) monochromatic oscillations in the case of weak rf losses.

QUALITATIVE DESCRIPTIONS OF THE TRANSITION PROCESS AND THE ROLE OF LOCAL DISPERSION IN THE CONVERSION OF A VIDEO PULSE TO A RADIO PULSE IN A NONLINEAR MEDIUM

It was shown numerically in Ref. 8 that the transition process accompanying the generation of rf oscillations by an electromagnetic shock wave consists of the formation of the shock wave front, generation of quasimonochromatic oscillations (with increasing amplitude), and deformation of the radio pulse envelope as a result of the dispersion of a transmission line with saturated nonlinearity. As the shock wave propagates, the number of oscillations behind the shock wave front increases in proportion to the length traveled by it, and the spectral composition of the generated radio pulse ''is enriched'' at the frequency of a wave phase-matched with the steady-state front of the electromagnetic shock wave [$v_p(\omega_0)=v_s$, where v_p is the phase velocity and v_s is the electromagnetic shock wave velocity]. Since in this case, the wave packet is ''rigidly attached'' to the shock wave front (traveling radiation source) and moves along behind it, the deformation of the wave packet in a linear medium behind the shock wave front will entail both a lengthening of the packet as a result of the generation of new oscillations and also its dispersional spreading. In general, this process can only be investigated by numerical methods.

However, some constraints imposed on the dispersion of

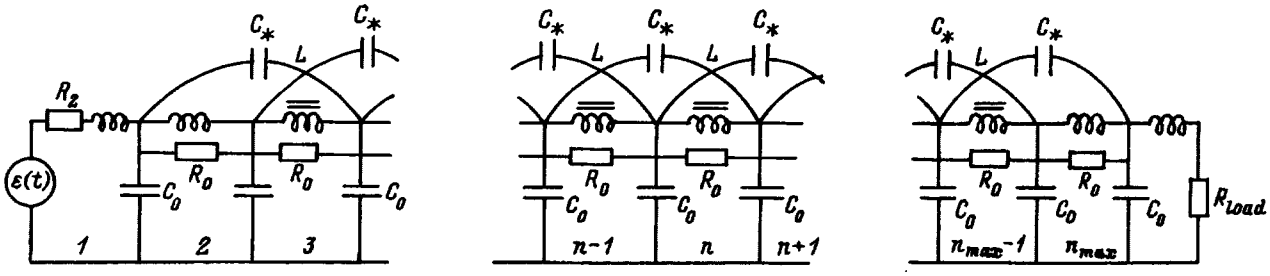

FIG. 1. Equivalent circuit of a transmission line with spatial dispersion.

a transmission line with saturated nonlinearity with a view to minimizing the transition process, especially on the minimum time taken for establishment of quasisteady-state decay of the generated radio pulse may be predicted *a priori* by analogy with the propagation of wave packets in linear dispersive media. It is known (see Ref. 9) that the rate of deformation of the envelope of a wave packet with a narrow frequency spectrum ($\Delta\omega/\omega_0 \ll 1$, where $\Delta\omega$ is the width of the wave packet spectrum and ω_0 is the average frequency) is determined by the local dispersion of the group velocity near the average frequency. For the generation of a radio pulse by an electromagnetic shock wave, the case of a narrow frequency spectrum is achieved when the shock wave front traverses a specific length of line, i.e., when a fairly extended radio pulse is generated. The local behavior of the dispersion near the average frequency $\omega=\omega_0$ may be expressed in the form

$$k(\omega)=k(\omega_0)+\left(\frac{dk}{d\omega}\right)_{\omega_0}(\omega-\omega_0)+\frac{1}{2}\left(\frac{d^2k}{d\omega^2}\right)_{\omega_0}(\omega-\omega_0)^2 + \frac{1}{6}\left(\frac{d^3k}{d\omega^3}\right)_{\omega_0}(\omega-\omega_0)^3+\ldots, \tag{1}$$

where k is the wave number.

In the first approximation with respect to $\Delta\omega/\omega_0 \ll 1$ (where $\Delta\omega=\omega-\omega_0$), the wave packet propagates without distortions with the group velocity

$$v_g=\left(\frac{dk}{d\omega}\right)_{\omega_0}^{-1}.$$

It follows from the second approximation with respect to $\Delta\omega/\omega_0$ (Ref. 9), that the spreading of the wave packet decreases, the stronger the inequality

$$\left(\frac{d^2k}{d\omega^2}\right)_{\omega_0}\Delta\omega \ll \left(\frac{dk}{d\omega}\right)_{\omega_0}$$

or

$$\frac{1}{v_g}\frac{dv_g}{d\omega}\bigg|_{\omega_0}\Delta\omega=v_g^{-2}\frac{dv_g}{dk}\bigg|_{k=k_0}\Delta k \ll 1. \tag{2}$$

It can be seen from Eq. (2) that the wave packet spreads more slowly as its frequency spectrum becomes narrower and the dispersion of the group velocity becomes smaller, i.e., $(dv_g/d\omega)_{\omega_0}$ or $(dv_g/dk)_{k=k_0}$. It is known that the distortion of the envelope of an extended wave packet ($\Delta\omega/\omega_0 \ll 1$) and its phase structure as it propagates over a path l will be negligible if

$$l \leqslant \lambda_0^2\left(\frac{\omega_0}{\Delta\omega}\right)^2 \Big/ \left(\frac{dv_g}{d\omega}\right)_{\omega=\omega_0}, \tag{3}$$

where $\lambda_0=2\pi v_p/\omega_0$.

Obviously, for $(dv_g/d\omega)_{\omega=\omega_0}=0$ the value of l will be determined by higher derivatives of v_g and will be of a higher order in $(\omega_0/\Delta\omega)$. For the generation of rf oscillations by an electromagnetic shock wave (direct conversion of an extended video pulse into a radio pulse), the number of generated oscillations in the radio pulse increases as the shock wave front propagates in a dispersive nonlinear transmission line. Thus, its spectral composition also ''narrows'' ($\Delta\omega/\omega_0$ decreases). In accordance with condition (3), as the shock wave front propagates in the transmission line, the length l for which the dispersion distortion of the wave packet remains negligible, will increase. Thus, for a semiinfinite (or fairly extended) video pulse it may be expected that the decay profile of the generated radio pulse will stabilize with time, even with low or no rf losses. In this case, only part (in the region between the leading and trailing edges) of the electromagnetic shock wave structure will come close to steady-state.

However, this reasoning is qualitatively valid from the instant that the spectrum of the radio pulse generated by the electromagnetic shock wave becomes sufficiently narrow ($\Delta\omega/\omega_0 \ll 1$). At the same time, the transition process begins with the generation of a single oscillation, i.e., at the initial stage of generation $\Delta\omega \sim \omega_0$.

Thus, in general the influence of local dispersion on the rate of formation of a quasisteady-state radio pulse can only be investigated by numerical methods.

DESCRIPTION OF MODEL. METHODS OF INVESTIGATION

As in Refs. 7 and 8, we shall consider the transmission line to be an LC line with capacitive cross couplings via one element (Fig. 1). An advantage of this system is that the dispersion is easily controlled by a single parameter — the cross coupling coefficient $\gamma_*=C_*/C_0$ (C_* is the cross coupling capacitance and C_0 is the main capacitance of the LC line) and that for specific values of γ_* the dependence of v_g on k or the phase shift per element has a minimum. To simplify the analysis, we shall assume that one inductance is nonlinear, in which magnetization reversal of the ferrite

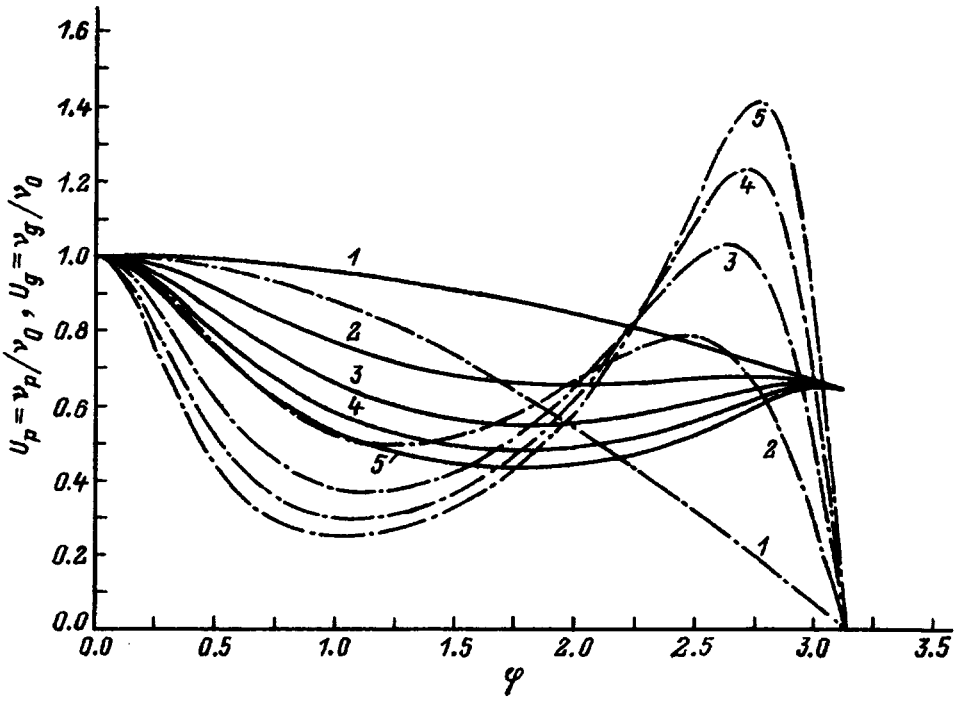

FIG. 2. Phase (solid curves) and group (dot-dash curves) velocities as a function of the phase shift per element. γ_* : *1* — 0, *2* — 0.2, *3* — 0.4, *4* — 0.6, and *5* — 0.8.

takes place incoherently in strong fields. Typical dispersion characteristics ($U_p = v_p/v_0$ is the relative phase velocity, $v_p = d\cdot\omega/\varphi$, $U_g = v_g/v_0$ is the relative group velocity, $v_g = d\cdot d\omega/d\varphi$, d is the period of the system, φ is the phase shift per element, and $v_0 = d\cdot\tau_0^{-1}$) are plotted in Fig. 2. These were calculated for several values of γ_* using the dispersion relation

$$\sin^2\frac{\varphi}{2} - 4\gamma_*\frac{\omega^2}{\omega_c^2}\sin^2\varphi = \frac{\omega^2}{\omega_c^2}, \tag{4}$$

where $\omega_c = 2/\sqrt{L_0C_0}$.

The nonlinear processes in an LC line with cross couplings and ferrite nonlinearities are described by the system of differential equations

$$i_n - i_{n+1} + \frac{1}{r_0}(u_{n-1} - 2u_n + u_{n+1}) + \gamma_*\frac{d}{d\tau}(u_{n-2} - 2u_n + u_{n+2}) = \frac{du_n}{d\tau},$$

$$\frac{di_n}{d\tau} = u_{n-1} - u_n - 4\pi\eta q_0(1 - m_n^2)i_n,$$

$$\frac{dm_n}{d\tau} = q_0(1 - m_n^2)i_n.$$

The first two equations are essentially the Kirchhoff equations for the nth element in the line. The last equation describes fast incoherent magnetization reversal of the ferrite. In these equations i_n and u_n are the dimensionless current and voltage in the nth element of the transmission line, $r_0 = R_0/Z_0$, $Z_0 = \sqrt{L_0/C_0}$ is the wave impedance of the transmission line without cross coupling, $m_n = M_n/M$ is the dimensionless magnetization of the ferrite (M is the saturation magnetization: $4\pi M = B_s$), $\tau = t/\tau_0$ is the dimensionless time ($\tau_0 = \sqrt{L_0C_0}$), η is the filling factor of the inductance by the ferrite, q_0 is the dimensionless switching coefficient [$q_0 = (\alpha\gamma_0 M\tau_0)/(1+\alpha^2)$, α is the dissipation coefficient, and $\gamma_0 = 1.76\times10^7$ Oe^{-1} s^{-1} is the absolute value of the gyromagnetic ratio for an electron.] This system of equations is valid for $n \geqslant 3$. For the first two elements we have

$$\frac{du_1}{d\tau} = i_1 - i_2 + \gamma_*\frac{d}{d\tau}(u_3 - u_1),$$

$$\frac{di_1}{d\tau} = \varepsilon(\tau) - i_1 r_g - u_1,$$

$$\frac{du_2}{d\tau} = i_2 - i_3 + \gamma_*\frac{d}{d\tau}(u_4 - u_2),$$

$$\frac{di_2}{d\tau} = u_1 - u_2.$$

For the last two elements we have

$$\frac{du_{n_{\max}-1}}{d\tau} = i_{n_{\max}-1} - i_{n_{\max}} + \gamma_*\frac{d}{d\tau}(u_{n_{\max}-3} - u_{n_{\max}-1}),$$

$$\frac{di_{n_{\max}-1}}{d\tau} = u_{n_{\max}-2} - u_{n_{\max}-1},$$

$$\frac{du_{n_{\max}}}{d\tau} = i_{n_{\max}} - i_{n_{\max}+1} + \gamma_*\frac{d}{d\tau}(u_{n_{\max}-2} - u_{n_{\max}}),$$

where $n_{\max}$ is the number of elements in the transmission line, r_g is the internal resistance of the input video pulse generator, normalized to Z_0, and r_{load} is the load resistance normalized to Z_0.

The initial conditions are: $i_n(\tau=0) = 0$, $u_n(\tau=0) = 0$, $m_n(\tau=0) = m_0$ ($-1 \leqslant m_0 \leqslant 1$), where m_0 is the relative magnetization of the ferrite. By varying the initial magneti-

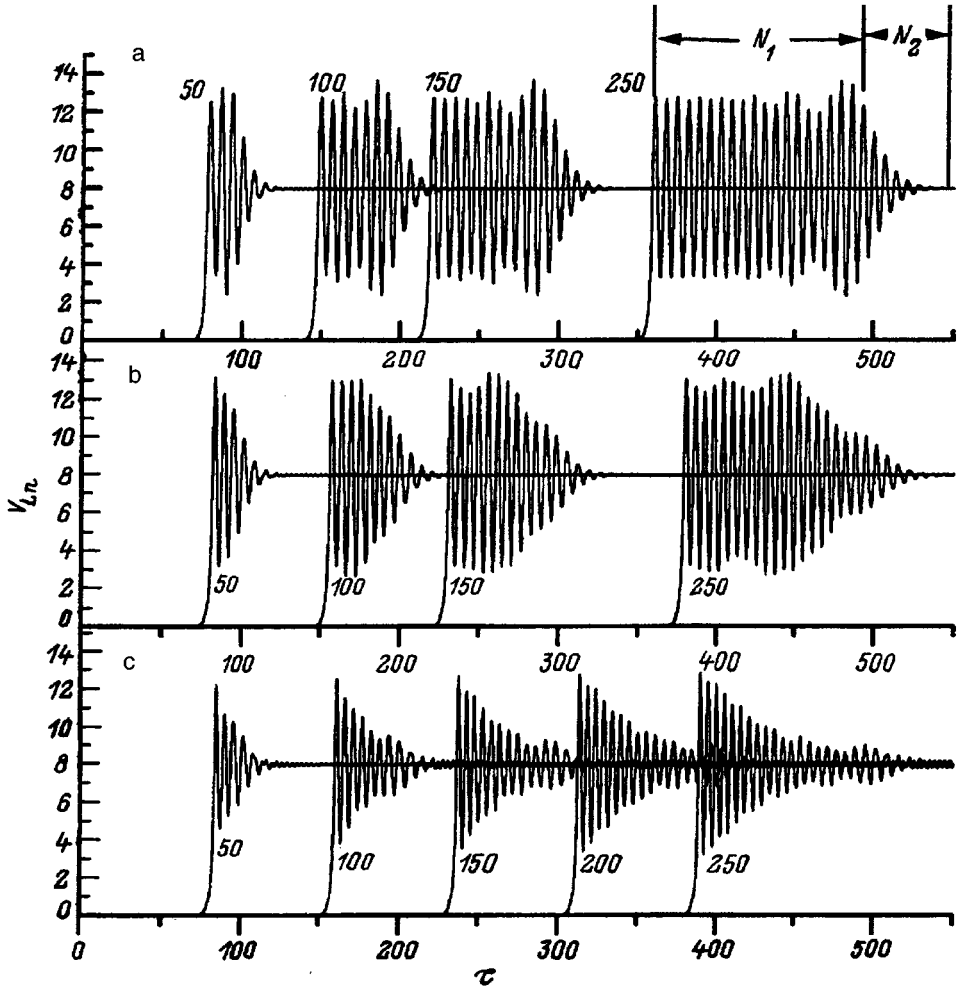

FIG. 3. Oscilloscope traces of the voltage in 50, 100, 150, and 250 elements of the transmission line. φ: a—1.26, b—1.55, c—1.85; v_p/v_0: a—0.7115, b—0.6728, c—0.6547; m_0: a—0.068, b—−0.049, c—−0.273; ε_0: a—14.43, b—14.69, c—15.17.

zation, we can control the generation process by altering the shock wave velocity and thus the generation frequency (the position of the operating point on the dispersion curve can be varied). The boundary condition at the entrance (the output voltage of the video pulse generator $\varepsilon(\tau)$ supplied to the transmission line) has the form

$$\varepsilon(\tau)=\varepsilon_0\begin{cases}\sin^2\dfrac{\omega_1\tau}{2}, & 0\leqslant\omega_1\tau\leqslant\pi,\\ 1, & \pi\leqslant\omega_1\tau.\end{cases}$$

The values of q_0 and η were taken as 0.4 and 0.5, respectively. We assumed $r_0=5000$ so that the rf attenuation was negligible.

DISCUSSION OF THE NUMERICAL RESULTS

Figure 3 shows typical calculated oscilloscope traces of the voltages $V_{Ln}=u_n-(1/2)(di_n/d\tau)$ at the midpoint of the inductance in a particular set of elements of the transmission line (Fig. 1) with $\gamma_*=0.2$ under phase matching conditions corresponding to different points on the dispersion characteristic (Fig. 2). The electromagnetic shock wave velocity and therefore the phase matching $v_s=v_p(\omega)$ at different frequencies for different local dispersion of the group velocity was regulated by varying the initial magnetization of the ferrite and the output voltage of the input video pulse generator, with all the other parameters kept constant. The traces for $\varphi_1=1.26$ correspond to phase matching at the minimum of the group velocity $(dv_g/d\varphi)_{\varphi=\varphi_1}=0$, and for φ_2 and φ_3, respectively, $0<(dv_g/d\varphi)_{\varphi_1}<(dv_g/d\varphi)_{\varphi_2}$. It can be seen from Fig. 3 that as the local dispersion of the group velocity varies, the transition process and specifically duration of the trailing edge of the radio pulse vary substantially.

With some degree of arbitrariness, the train of oscillations formed behind the front may be divided into two characteristic sections: the first $N_1(n)$ (where n is the number of the element in the line) comprises oscillations of almost identical amplitude, and the second $N_2(n)$ comprises oscillations making up the trailing edge of the radio pulse (Fig. 3).

Figure 4 gives N_1 and N_2 in different elements of a transmission line with $\gamma_*=0.2$ for various electromagnetic shock wave velocities corresponding to phase matching $v_s=v_p(\omega)$ at frequencies with very different values of the first derivative of the group velocity. It can be seen from Fig. 4 that for phase matching at the minimum of the group velocity ($\varphi_1=1.26$) the trailing edge of the radio pulse is of a quasisteady nature as the shock wave propagates through 150 elements of the transmission line (see also Fig. 3), after which N_2 remains constant in the subsequent elements. In this case, the number of oscillations N_1 increases almost lin-

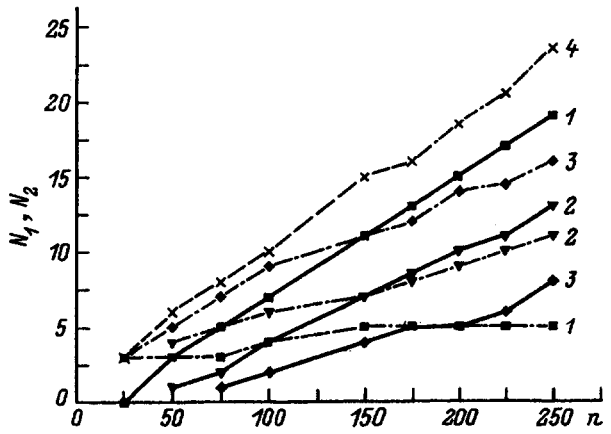

FIG. 4. Number of oscillations N_1 (solid curves) and N_2 (dot-dash curves) as a function of the element number in the line (for phase matching corresponding to different points on the dispersion curve). φ: *1* — 1.26, *2* — 1.55, *3* — 1.7, and *4* — 1.85; v_p/v_0: *1* — 0.7115, *2* — 0.6728, *3* — 0.6613, and *4* — 0.6547; m_0: *1* — 0.068, *2* — −0.049, *3* — −0.229, and *4* — 0.273; ε_0: *1* — 14.43, *2* — 14.69, *3* — 15.07, and *4* — 15.17.

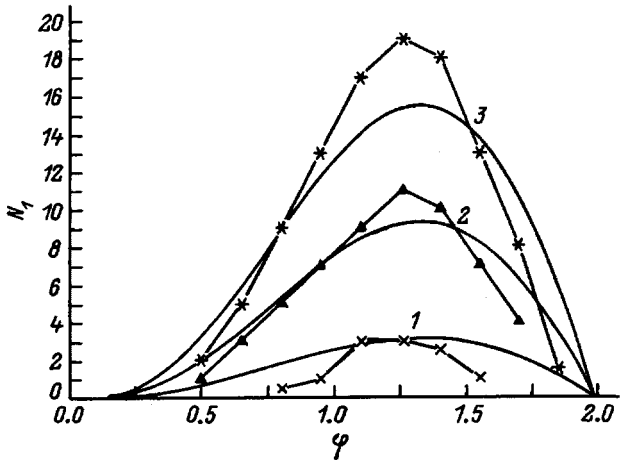

FIG. 5. Number of oscillations N_1 as a function of the phase shift per element obtained by computer calculations (curves with asterisks) and estimated using formula (5). n: *1* — 50, *2* — 150, and *3* — 250.

early with the element number in the transmission line. The traces plotted in Fig. 3a show that from the 150th element both the number of oscillations on the trailing edge of the radio pulse and the envelope in this region remain constant.

For phase matching outside the minimum of the group velocity v_g the increment of N_1 decreases and N_2 increases with increasing $dv_g/d\varphi$. An increase in the local dispersion of the group velocity lengthens the process of formation of the quasisteady-state field structure because of the dispersional spreading of the wave packet. Note that for the length of transmission line being studied, the trailing edge of the radio pulse generated by the electromagnetic shock wave does not have time to form a steady-state structure in cases of appreciable local dispersion even in the first order in $\Delta\omega/\omega$ $[(dv_g/d\omega)(\omega/v_g)$ is 0.37, 0.49, and 0.56, respectively for $\varphi=1.55$, 1.7, and 1.85]. The narrowing of the radio pulse spectrum $[\Delta\omega/\omega\sim 1/(N_1+N_2)]$ is clearly insufficient [see Eq. (3)] to stabilize the trailing edge of the generated radio pulse with this dispersion after propagation through 250 elements in the line. Note that the number of oscillations N_1 in a radio pulse of the same amplitude will be highest for phase matching at the minimum of v_g (Fig. 5). Figure 5 gives the results of computer calculations and data obtained using an approximate formula for N_1 for the establishment of a quasisteady-state regime. This formula is derived from the condition for backflow of energy from the electromagnetic shock wave front at the frequency of the excited oscillations and has the form (see also Ref. 8)

$$N_1(n)=\frac{(v_s-v_g)\varphi}{2\pi v_s}n. \tag{5}$$

Some difference between the estimate and the numerical results may be attributed to the arbitrary division of the nonsteady-state generated train into a main section and a trailing edge, made when analyzing the results.

The influence of the local dispersion in orders higher than the first in $\Delta\omega/\omega$ on the transition processes of generation may be investigated by analyzing the phase matching at minima of v_g having different slopes. For the transmission line shown in Fig. 1, $d^2v_g/d\varphi^2$ may be varied at the minimum of v_g by varying the cross coupling coefficient. Figure 6 gives N_1 and N_2 as a function of the element number for the propagation of an electromagnetic shock wave in transmission lines with $\gamma_*=0.2$ and 0.8. It can be seen from Fig. 6 that even for a small variation in $v''_{g\varphi\varphi}$ $[(v''_{g\varphi\varphi})_{\gamma_*=0.8}/(v''_{g\varphi\varphi})_{\gamma_*=0.2}\approx 1.07]$, the number of oscillations in the quasisteady-state trailing edge varies by a factor of 1.5–2. However, this quasisteady-state trailing edge forms after propagation through approximately the same number of elements in transmission lines with $\gamma_*=0.2$ and 0.8. This is evidently because, in the line with $\gamma_*=0.8$, the spectrum of the generated radio pulse narrows faster than in the line with $\gamma_*=0.2$ $[N(n)|_{\gamma_*=0.8}/N(n)|_{\gamma_*=0.2}>1$, $N=N_1+N_2]$. Figure 7 shows the spectral density of the radio pulses formed in 50, 150, and 250 elements of transmission lines with $\gamma_*=0.2$ and 0.8. It can be seen that in both cases, the spectral density of the radio pulses at the wave frequency phase-

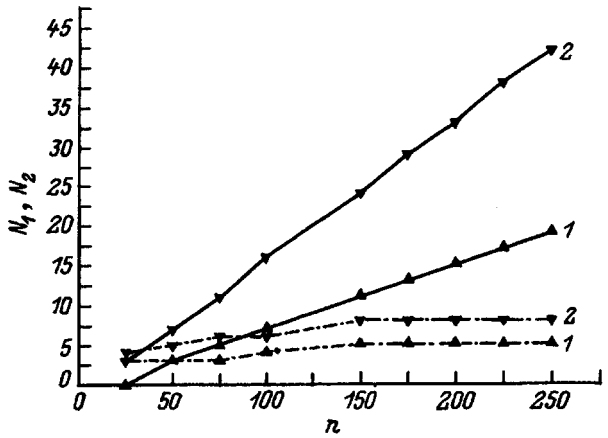

FIG. 6. Dependences of N_1 (solid curves) and N_2 (dot-dash curves) on the element number. γ_*: *1* — 0.2, *2* — 0.8; φ: *1* — 1.26, *2* — 1.04; v_p/v_0: *1* — 0.7115, *2* — 0.5198; m_0: *1* — 0.068, *2* — −0.720; ε_0: *1* — 14.43, *2* — 13.69.

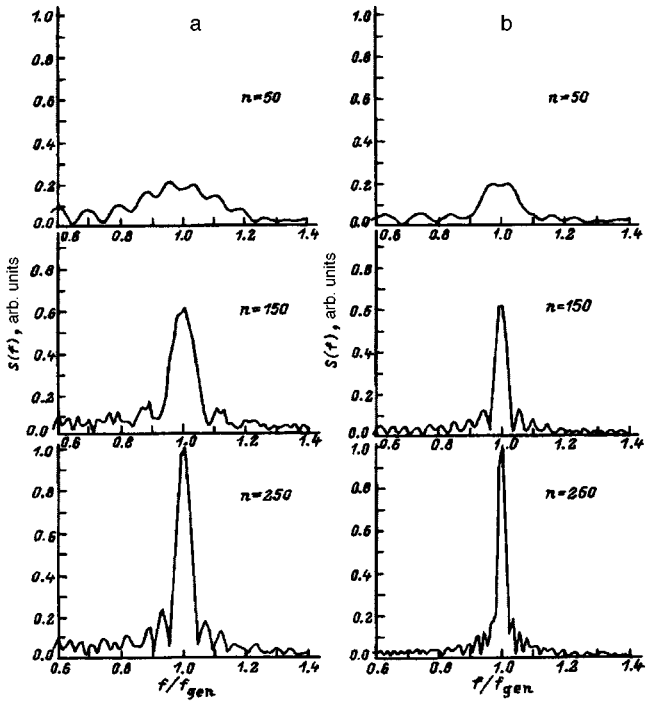

FIG. 7. Evolution of the spectral density of the radio pulses. γ_*: a — 0.2, b — 0.8; v_p/v_0: a — 0.7115, b — 0.5198; m_0: a — 0.068, b — −0.720; ε_0: a — 14.43, b — 13.69.

matched with the electromagnetic shock wave increases with increasing n. In these cases, the condition $\Delta\omega/\omega \ll 1$ ($\Delta\omega$ is the half width of the spectrum at half maximum) is satisfied for propagation of the shock wave through 150 elements. For $\gamma_* = 0.8$ the spectrum narrows more rapidly.

Knowing the spectral composition and the local dispersion, we can easily understand the nature of the amplitude modulation in the radio pulse generated by the electromagnetic shock wave (Fig. 3). In particular, under phase matching $v_s = v_p(\omega_0)$ at the minimum of $v_g(\omega_0)$ all the spectral components $\omega_0 \pm \Delta\omega$ for $\Delta\omega/\omega \ll 1$ have the group velocity $v_g(\omega) > v_g(\omega_0)$. The energy of these spectral components ''lags'' more slowly behind the shock wave front than that at the frequency of the phase-matched wave, and this shows up as negligible amplitude modulation in the radio pulse (Fig. 3a). For phase matching $v_s = v_p(\omega_0)$ outside the minimum of $v_g(\omega_0)$, modulation is also observed at the trailing edge of the radio pulse (Fig. 3c) because components with $v_g(\omega) < v_g(\omega_0)$ are present in a narrow band of the radio pulse spectrum. Strictly speaking, the amplitude-modulated part of the generated radio pulse, like its trailing edge, should be assigned to the transition process. It is easy to see from Fig. 3a that for phase matching at the minimum of v_g the transition process is stabilized after propagation through 150 elements of the transmission line, since the number of oscillations and the profile of the transition region of the radio pulse are almost indistinguishable on the oscilloscope traces for 150 and 250 elements. After the transition process has stabilized, the length of the radio pulse increases as a result of an increase in the number of oscillations with an amplitude that decreases with increasing distance from the front, as in a steady-state electromagnetic shock wave, because of rf losses in a transmission line with saturated nonlinearity[7] whereas at the trailing edge of the radio pulse, this is caused by dispersion.

CONCLUSIONS

Thus, from the point of view of optimizing the transition process accompanying the direct conversion of a video pulse into a radio pulse as this pulse propagates in nonlinear transmission lines possessing hysteretic properties, the optimum solution involves phase matching between the front of the electromagnetic shock wave and the wave excited by it at a frequency near the minimum local dispersion of the group velocity. It has been shown for the example of an LC transmission line with capacitive cross coupling that under phase matching $v_s = v_p(\omega_0)$ at the minimum of the group velocity $v_s(\omega_0)$, the trailing edge of the radio pulse relatively rapidly acquires a steady-state profile and its duration may be 4–5 periods of the generated oscillations. The possible direct conversion of a video pulse into a radio pulse with steep leading and trailing edges is highly promising from the practical point of view.

This work was supported financially by the Russian Fund for Fundamental Research (Project No. 94-02-05443).

[1] A. V. Gapanov, L. A. Ostrovskiĭ, and G. I. Freĭdman, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **10**, 1376 (1967).
[2] A. V. Gapanov, L. A. Ostrovskiĭ, and M. I. Rabinovich, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **13**, 164 (1970).
[3] A. Scott, *Active and Nonlinear Wave Propagation in Electronics* [Wiley-Interscience, New York, (1970); Sov. Radio, Moscow, (1977)].
[4] G. B. Whitham, *Linear and Nonlinear Waves* [Wiley Interscience, New York, (1974); Mir, Moscow, (1977)].
[5] H. Ikezi, S. S. Wojtowicz, R. E. Waltz *et al.*, J. Appl. Phys. **64**, 3277 (1988).
[6] T. Kuusela and J. Hietarinta, Rev. Sci. Instrum. **62**, 2266 (1991).
[7] A. M. Belyantsev and S. L. Klimin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **36**, 1011 (1993).
[8] A. M. Belyantsev, A. I. Dubnev, S. I. Klimin *et al.*, Zh. Éksp. Teor. Fiz. **65**(8), 132 (1995) [Tech. Phys. **40**, 820 (1995)].
[9] M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukhov, *Wave Theory* [in Russian], Nauka, Moscow (1990).

Translated by R. M. Durham

Some properties of the envelope equations used in the design and adjustment of electron-optical beams

Yu. V. Zuev

D. V. Efremov Scientific-Research Institute for Electrophysical Apparatus, 189631 St. Petersburg, Russia
(Submitted June 17, 1996)
Zh. Tekh. Fiz. **68**, 96–102 (January 1998)

An analytical transformation of the usual envelope equations is used to derive scale transformation equations for a first-order electron-optic lens. Analysis of these equations, which take into account both the particle temperature and the particle space charge, leads to the identification of certain general principles for constructing and scaling beam systems of various types. The properties of the transformation are illustrated by the example of the equations for an axisymmetric electrostatic lens.
[S1063-7842(98)01801-7]

INTRODUCTION

The design of beam optics always includes, to some extent or other, the solution to an inverse problem of dynamics, i.e., that of finding structures and force-field intensities that provide the required beam characteristics. In turn, any force field used to shape a beam is in the final analysis always determined by certain axial functions and their derivatives [i.e., controls $u(z)$]. Using the paraxial equation, we can write the linear $r(z)$ and angular $r'(z)$ envelopes of a beam in terms of the following integrals:

$$r(z)=r_0+\int_0^z r'(z)dz, \quad r'(z)=r_0'+\int_0^z \Phi[u(z),p]dz.$$

Analogous expressions can be written down for aberration corrections that characterize the intensity of beam heating. From this starting point, the beam design should include the successive solution of two problems: 1) finding an on-axis field that ensures the desired optical properties of the system; and 2) extrapolating this field into the surrounding space to determine the necessary field-shaping elements. The first problem reduces to finding that control $u(z)$ which, for the boundary conditions

$$u(0)=u_0, \quad u(L)=u_L, \tag{1}$$

would assign to functionals of the form

$$f(u)=\int_0^L \Phi[u(z),p]dz,$$

$$h(u)=\int_0^L \int_0^z \Phi[u(\zeta),p]d\zeta dz$$

two completely determined values

$$|h(u)-A|\leqslant \delta_A, \quad |f(u)-B|\leqslant \delta_B,$$

while ensuring as far as possible that

$$g(u)=\int_0^L \Omega[u(z),p]dz\leqslant C \text{ or } g(u)\to\min. \tag{4}$$

Sometimes these requirements are combined:

$$\omega_1[h(u)-A]^2+\omega_2[f(u)-B]^2+\omega_3[g(u)-C]^2\leqslant\delta. \tag{5}$$

Here and in what follows, $h(u)$ corresponds to the size and $f(u)$ to the slope angle of the envelope at the target, while $g(u)$ estimates the finite temperature in the beam. The current, charge, mass, and velocity of the particles, the transverse dimensions of the beam cross section at the entrance, the slope of the envelopes, the emittance, etc. play the role of parameters, here labeled by the set p.

In this formulation, such problems belong to the isoperimetric class of variational problems and should be solved using the corresponding methods. However, the use of these methods is hindered in practice by the large number of constraints imposed on the desired control, which makes it impossible to determine the functionals for all values of their arguments. Usually these constraints, which ensure physical realizability of the control, are expressed in terms of the beam aperture

$$R_{\min}\leqslant\int_0^z\int_0^\zeta \Phi[u(\zeta),p]d\zeta dz\leqslant R_{\max}, \quad 0\leqslant z\leqslant L, \tag{6}$$

the length of the structure

$$L_{\min}\leqslant L\leqslant L_{\max}, \tag{7}$$

the achievable field intensity

$$u_{\min}\leqslant u(z)\leqslant u_{\max}, \tag{8}$$

the rate of change of the control

$$|u'(z)|\leqslant u'_{\max} \tag{9}$$

etc. The literature to date contains only isolated cases where these variational methods are used to solve problems of this kind.

More often, an attempt is made to describe the control by a finite number of parameters and express the functionals in terms of them. Once this is done, the problem is reduced to finding an extremum of some effectiveness function of n variables. The quantity usually used as an effectiveness function is the so-called merit (quality, Q) function of the device, which also includes deviations of the beam characteristics from their normal values. The search for the extremum is

carried out by the methods of optimal control or multicriterion optimization. Unfortunately, this transformation is made feasible largely by using a matrix description of the system, which is unsuitable for beams in which an important role is played by the self-charge. Hence, in the overwhelming majority of cases the design procedure consists of finding multiple solutions to the direct problem, i.e., selecting parameters that describe the force field and then calculating beam characteristics until the necessary transformation is obtained. The methods most widely used in this case are nonlinear programming techniques.

Computer programs based on principles such as these, while guaranteeing physical realizability of the system designs, can only find a design solution if the initial approximation is good. This category includes such widely known programs as TRANSPORT and TRACE.[1] The most important cause of this difficulty is the fact that the inverse problem is ill-posed. Problems of this kind, in which a cause and effect relation is inverted, can have either a multiplicity of solutions or no solution at all, depending on the initial constraints. Moreover, the method of solution also plays a nontrivial role. In the problem under discussion here, our task is not to optimize a function or functional, but rather to attach definite values to them. These values do not necessarily lie in the ''bottom'' that is closest to the starting point.

The goal of this paper is to develop a nonoptimizational method for solving this problem within the formulation used currently in the design of beam optics, i.e., matching the parameters of the beam to the transfer characteristics of the accelerating structures or transport channels [conditions (2), (3) but without condition (4)]. The few direct methods known to the author for solving such problems apply to systems with a certain field structure (like, e.g., those in 1–4) and negligibly small perveance.[1–3,5,6]

TRANSFORMATION OF THE ENVELOPE EQUATIONS BY A COUPLING FUNCTION. GENERALIZED EQUATIONS FOR SCALE TRANSFORMATIONS

Ordinarily, the envelope equations most often used to design first-order lenses in optical systems have the following forms. The envelope equation in a solenoid is

$$r''=\Phi_S(r,B), \tag{10}$$

where $\Phi_S(r,B)=Q_S|r+\varepsilon^2|r^3-\text{const}\ rB^2$; B is the magnetic induction on the axis; Q_S is a space charge parameter which depends on the value of the beam current and on the type and velocity of the particles; and ε is the emittance.

In a quadrupole channel the envelope equations are

$$x''=Gx+F_x, \quad y''=-Gy+F_y. \tag{11}$$

Here the reduced gradient G of the lens plays the role of a control; $F_x(x,y)=Q/(x+y)+\varepsilon_x^2/x^3$, and $F_y(x,y)=Q/(x+y)+\varepsilon_y^2/y^3$. The system of Eqs. (11) can be reduced to a single equation which relates only the transverse dimensions x and y of the beam:

$$yx''+xy''=\Phi_q(x,y), \tag{12}$$

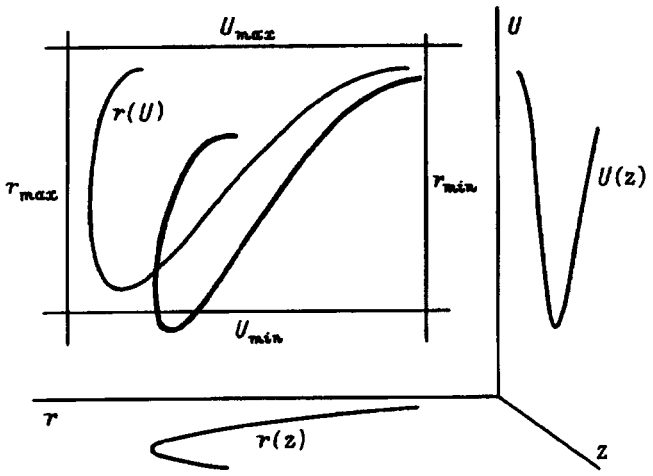

FIG. 1. A description of a lens–beam system in the space (U,r,z).

where $\Phi_q(x,y)=yF_x+xF_y$. Then $G=(x''-F_x)/x=(y''-F_y)/(-y)$ (in this type of description the specified dimensions of the beam act as the control).

In a deflecting magnet the envelope equations are

$$x''=-\frac{(1-n)}{R^2}x+F_x, \quad y''=-\frac{n}{R^2}y+F_y$$

or

$$yx''+xy''=\Phi_M(x,y). \tag{13}$$

Here R is the radius of deflection of the magnet, n is the decay exponent of the field, and $\Phi_M(x,y)=yF_x+xF_y-(xy)/R^2$, $nR^{-2}=R^{-2}+(x''-F_x)/=(y''-F_y)/(-y)$.

It is most convenient to witness the advantages of this approach for the example of the envelope equations in an axisymmetric electrostatic lens. The field of the lens is determined only by the axial distribution of the potential $U(z)$, and the behavior of the envelope is determined by the initial conditions $r(0)=r_0$, $r'(0)=r_0'$ and the equation

$$r''U+0.5r'U'+0.25rU''=\Phi_U(r,U), \tag{14}$$

where $\Phi_U(r,U)=Q_U/(\sqrt{U}r)+\varepsilon_H^2/r^3$, and ε_H is the normalized emittance.

Consequently, the lens–beam system can be described by a curve in space (the solid curve in Fig. 1) in the coordinates Urz, which are specified, e.g., by the projections $U(z)$ and $r(z)$. Equation (14) is normally used in looking for projections of this curve on the rz plane for a known projection on the Uz plane (or conversely). The idea of this method is to start from the projection on the Ur plane. Then, by specifying some function that initially relates the control to the beam, i.e., a coupling function, we determine beforehand all the values that both the beam radius and the potential in the lens can take. Particularly important among these values are their minimum, maximum, and boundary values, i.e., those that satisfy conditions (1), (2), (6), and (8); see Fig. 1.

There are also a number of advantages associated with the equation that establishes a relation between the coupling function and the optical axis. Its form

TABLE I.

$$\ddot{z}=\frac{\dot{z}}{h_1(\zeta)}[h_2(\zeta)-h_3(\zeta)\dot{z}^2]$$

Type of optical system	Coupling function	ζ	z	h_1	h_2	h_3
Solenoid	$r(B)$	B	$\frac{dz}{dB}$	$\frac{dr}{dB}$	$\frac{d^2r}{dB^2}$	Φ_S
	$B(r)$	r	$\frac{dz}{dr}$	1	0	Φ_S
Axisymmetric electrostatic system	$r(U)$	U	$\frac{dz}{dU}$	$U\frac{dr}{dU}+0.25r$	$0.5\frac{dr}{du}+U\frac{d^2r}{dU^2}$	Φ_U
	$U(r)$	r	$\frac{dz}{dr}$	$U+0.25r\frac{dU}{dr}$	$0.5\frac{dU}{dr}+0.25\frac{d^2U}{dr^2}$	Φ_U
Quadrupole channel	$y(x)$	x	$\frac{dz}{dx}$	$y+x\frac{dy}{dx}$	$x\frac{d^2y}{dx^2}$	Φ_q
	$x(y)$	y	$\frac{dz}{dy}$	$x+y\frac{dx}{dy}$	$y\frac{d^2x}{dy^2}$	Φ_q
Deflecting magnet	$y(x)$	x	$\frac{dz}{dx}$	$y+x\frac{dy}{dx}$	$x\frac{d^2y}{dx^2}$	Φ_M
	$x(y)$	y	$\frac{dz}{dy}$	$x+y\frac{dx}{dy}$	$y\frac{d^2x}{dy^2}$	Φ_M

$$\ddot{z}=\frac{\dot{z}}{h_1}(h_2-h_3\dot{z}^2) \tag{15}$$

is invariant with respect to selection of a new independent variable and is correct for all the equations we have discussed above (see Table I). This latter fact makes it possible to use unified principles to design systems of different types, based on the properties of this equation.

Equation (15) can have nodal and saddle-point singularities. The character and position of these singularities determine the uniqueness or multiplicity of realizations of the coupling function, or whether it is fundamentally nonrealizable (by a realization we here mean a solution to the ordinary envelope equation which gives a corresponding projection on the Ur plane). Multiplicity allows one to choose that realization which has at the boundaries not only the required values of the control and beam radius but also all the necessary derivatives, i.e., the realization that completely solves the problem under consideration. Identification and analysis of the singularities is a simple matter. Their types and positions do not depend on the form in which the coupling function is represented.

The fundamental equation preserves its form when we convert from an explicit representation to a parametric one:

$$\dot{M}=\frac{M}{f_1(\lambda)}[f_2(\lambda)-f_3(\lambda)M^2] \tag{16}$$

(the expressions for f_1, f_2, f_3 are given in Table II; the dots indicate differentiation with respect to the parameter λ).

However, the variable M now plays the same role in all the equations — it is the differential coefficient of a scale transformation, since by definition $M\equiv dz/d\lambda$. Furthermore, Eq. (16) reveals yet another series of properties that are useful not only in designing the optics but also in adjusting them.

Because the expressions for $f_2(\lambda)$ and $f_3(\lambda)$ correspond to the left and right sides of Eqs. (10) and (12)–(14) when λ

TABLE II.

$$M=\frac{M}{f_1(\lambda)}[f_2(\lambda)-f_3(\lambda)M^2]$$

Type of optical system	Coupling function	$f_1(\lambda)$	$f_2(\lambda)$	$f_3(\lambda)$	$\frac{df_1(\lambda)}{d\lambda}$
Solenoid	$\{B(\lambda),r(\lambda)\}$	$\dot{r}$	$\ddot{r}$	Φ_S	$f_2(\lambda)$
Electrostatic system	$\{U(\lambda),r(\lambda)\}$	$0.25\dot{U}+U\dot{r}$	$\ddot{r}U+0.5r\dot{U}+0.25r\ddot{U}$	Φ_U	$f_2(\lambda)+0.75\dot{r}\dot{U}$
Quadrupole channel	$\{X(\lambda),Y(\lambda)\}$	$X\dot{Y}+Y\dot{X}$	$X\ddot{Y}+Y\ddot{X}$	Φ_q	$f_2(\lambda)+2\dot{X}\dot{Y}$
Deflecting magnet	$\{X(\lambda),Y(\lambda)\}$	$X\dot{Y}+Y\dot{X}$	$X\ddot{Y}+Y\ddot{X}$	Φ_M	$f_2(\lambda)+2\dot{X}\dot{Y}$

is replaced by z, it is now completely obvious that $M(\lambda)\equiv 1$, when the parametrically specified coupling function coincides with the solution to the ordinary equations. The latter suggests the idea of using as the coupling function the beam dimensions and the control of the system which is to be adjusted or which is being designed if it is necessary to change its characteristics. From this point of view, Eq. (16) is a generalized equation for a scale transformation of first-order optical systems with a coupling function used as an invariant. If the parametrically specified coupling function corresponds to a solution of the ordinary envelope equation and admits multiple realizations, integration of Eq. (16) with $M_0>1$ gives a system with a greater length $L=\int M(\lambda)d\lambda$ but smaller first derivatives along the axis $d/dz=M^{-1}d/d\lambda$. Integration with $M_0<1$ has the opposite effect, since the integral curves of Eq. (16) do not intersect except for singular points, and the quantity $(M-1)$ preserves its sign. Readjustments (redesign) of systems for a new value of the current, emittance, type of particle, etc., can be reduced to analogous integrations.

In the general case, on the other hand, parametric specification of the coupling function, assuming that the parameter used is the coordinate along the optical axis, may be regarded as an approximate (qualitative) description of the desired behavior of the beam and control in the system. Elementary analysis of the singular points determines whether the desired configuration is valid (realizable). The fundamental equation of the transformation rescales the optical axis so that the beam and control specified by the coupling function are matched with respect to their derivatives in exact correspondence with the ordinary envelope equation.

ANALYSIS AND PROPERTIES OF SINGULAR POINTS OF THE FUNDAMENTAL EQUATION OF THE TRANSFORMATION

It is not difficult to show that Eq. (16) with given initial conditions has a unique solution over the entire range of variation of the independent variable λ except for singular points at which both the numerator and denominator of the equation equal zero:

$$f_1(\tilde{\lambda})=0, \quad \tilde{M}[f_2(\tilde{\lambda})-f_3(\tilde{\lambda})\tilde{M}^2]=0.$$

At zeroes of the characteristic function $f_1(\lambda)$, the coupling function generates either one singularity (always zero) or three different singularities (one zero and two symmetric) corresponding to the real roots of Eq. (18). The form of the singularities will be the same as for a singular point of the linearized equation $\dot{M}=\tilde{M}+a(M-\tilde{M})/(\lambda-\tilde{\lambda})$ with coefficients $a_1=f_2(\tilde{\lambda})/\dot{f}_1(\tilde{\lambda})$ for $\tilde{M}_1=0$, $a_{2,3}=-2a_1$ for $\tilde{M}_{2,3}=\pm\sqrt{f_2(\tilde{\lambda})/f_3(\tilde{\lambda})}$. If $a\geqslant 0$ $[\dot{f}_1(\tilde{\lambda})\neq 0]$, then the singular point is a node. In the opposite case, a saddle-point singularity occurs. When $\dot{f}_1(\tilde{\lambda})=0$, the nature of the singular point can only be determined by investigating terms of higher order (additional information is given in Ref. 7). The coupling function is unrealizable in principle only when the corresponding integral curve of the transformation equation passes through zero or suffers a discontinuity. Such solutions to Eq. (16) can only be singular. The absence of singular points guarantees a multiplicity of realizations of the coupling function, specified by the initial conditions for integration of Eq. (16). To each M_0 there corresponds a different scaling of the optical axis for which the derivatives of the control and the beam derivatives change but the functions themselves still satisfy both the ordinary envelope equation and the coupling function chosen. In this case the ratio of their first derivatives remains unchanged: $r'/U'=\dot{f}/\dot{U}$.

The possibility of realization of a coupling function that forms zeroes of the characteristic function f_1 depends on the character of the singularities that arise in this case. The presence of nonzero singular points $(\tilde{M}\neq 0)$ always guarantees that a suitable solution to Eq. (16) will exist near $\tilde{\lambda}$. This is either a unique solution passing through the center of a nonzero saddle point (which uniquely determines the admissible initial value M_0 and the only possible value of the boundary derivatives), or an infinite set of solutions with a common point at a nonzero node. Provided that $f_2(\tilde{\lambda})\neq -0.5\dot{f}_1(\lambda)$, the integral curves passing through this nonzero node will also have a common tangent at the singular point. This allows us to join any two solutions obtained by integrating Eq. (16) from the left and the right of the singularity at the node in an ideal way that will ensure continuity of all the derivatives. This latter property also makes it possible to solve the problem (2), (3) formally by specifying the required boundary values as initial conditions.

In general, there can be several zeroes of the characteristic function. Necessary conditions for realizability of a coupling function in this situation naturally is realizability in the vicinity of each zero, i.e., $f_2(\tilde{\lambda}_i)f_3(\lambda_i)>0$, $i=1-N$. This condition is also sufficient if the adjacent nonzero singularities that form are of different type or are nodes. For adjacent saddle points a sufficient condition of general form has not yet been found.

There are two more properties of the transformation we have introduced that can be useful in adjusting and redesigning optical systems.

Property A. Let $\{U(\lambda),r(\lambda)\}$ be a realizable coupling function in parametric form, C_1 and C_2 are arbitrary positive constants; then $\{C_1U(\lambda),C_2r(\lambda)\}$ is also a realizable coupling function provided that $f_3(U,r)f_3(C_1U,C_2r)_{\lambda=\tilde{\lambda}_i}>0$, since the positions of the zeroes $\tilde{\lambda}_i$ and the character of the singular points of Eq. (16) are unchanged by such a transformation of the coupling function.

Property B. Let $p_1,\ldots,p_K$ be the parameters entering into f_3 of a realizable system with a coupling function $\{U(\lambda),r(\lambda)\}$, and let $\chi_1\ldots$, χ_k be some functions of λ. Then the coupling function $\{U(\lambda), r(\lambda)\}$ will be realizable with parameters $\chi_1p_1,\ldots,\chi_Kp_K$ if $f_3(U,r,p_1,\ldots,p_K)$ $f_3(U,r,\chi_1p_1,\ldots,\chi_Kp_K)_{\lambda=\tilde{\lambda}_l}>0$, since the character of the singular points for such changes in the parameters remains as before.

The only exception (but only from the point of view of realizability) that can occur here is when adjacent nonzero saddle points are present. A unique integral curve that connects the centers of the saddle points exists only under certain conditions, which can be violated as a result of scaling

of the coupling function or as a consequence of changes in the parameters.

GENERAL PRINCIPLES FOR CONSTRUCTING AND SCALING BEAM SYSTEMS

From what was said above, it follows that the use of transformation equations reduces the procedure for design (adjustment) of systems with respect to first-order optical characteristics to a rather arbitrary choice (correction) of a certain arbitrary function — the coupling function. For specified boundary values, this function should (or should not) generate some given sort of singularities of the fundamental equation. Analysis of the transformation equation reveals certain principles that are common for lenses of different types.

Any coupling function ensures the required transformation of the beam with respect to size if it is realizable and has the appropriate boundary values.

The same coupling function is realizable, as a rule, over a wide region of variation of current, emittance, particle type, etc. (property B). The boundaries of this region can be used to determine the limiting values of these parameters.

The coupling function can be subjected to linear scaling with preservation of the character of the singular points (property A). This scaling of the envelope or control can significantly accelerate a redesign of optical systems that involves changes in the restrictions on maximum aperture and force-field intensity, making a readjustment to nearby beam parameters easier.

When the coupling function has a nonzero node adjacent to the boundary, it is possible to obtain a given beam size for any slope angle of the same sign.

When a nonzero saddle-point singularity is adjacent to the boundary, the boundary value of the slope for a beam that realizes the chosen coupling function is uniquely specified. Certain combinations of boundary values of the characteristic function $f_1(\lambda)$, for example $f_1(0)<0$, $f_1(L)>0$ when $f_3(\lambda)>0$, entail the formation of such points automatically. In order to obtain a required slope angle in this case, as in the traditional design process, from a formal standpoint, it remains only to exhaust all variants of the coupling with the specified boundary values. However, problems with such ''inconvenient'' boundary conditions can be reduced to an equivalent problem by including in the system a suitable segment with optical properties that are known beforehand, which converts the boundary values of the beam and the control into more convenient values. A typical example of such boundary conditions is a converging input beam at zero force-field intensity. The simplest equivalent system in this case need only contain a drift region of a length such that beams that are converging at the input begin to diverge after passing through the crossover.

By varying the positions of singularities and (or) coupling functions between them, we can decrease the aberration (4) and optimize the length and other properties of the system without changing the size and slope angle of the beam at the boundary. In contrast to the classical formulation, optimization problems approached in this way are made easier by the fact that Eq. (16), as a special case of the Bernoulli equation, can be converted into a linear inhomogeneous equation with a solution in the form of quadratures.[7]

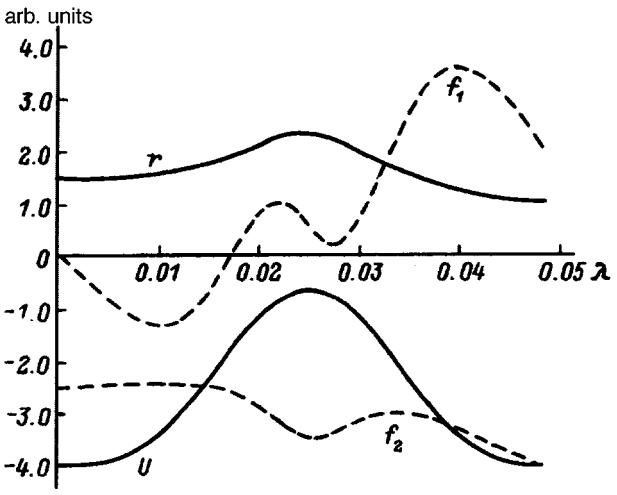

FIG. 2. Characteristics of a unipotential lens and a beam at a current of 30 mA.

The use of Eq. (16) allows us in principle to formalize the procedure for designing a system of the desired length. Any length that is not less than a certain length can *a priori* be guaranteed by that choice of coupling function for which two adjacent nonzero singularities of the node type are found on the interval of variation of λ. Then it remains only to choose properly a suitable integral curve that joins these two nodes, i.e., to pick M_0. Unfortunately, such a situation is not always realizable. A corresponding change in the sign of the function $f_3(\lambda)$ is necessary. And the equation for a solenoidal lens, for example, can never have nonzero nodal points under any conditions, since $f_2 \equiv df_1/d\lambda$.

As an illustration of several possible uses of the method proposed here, in what follows we give a number of transformations of an axially symmetric unipotential lens with a proton beam that are typical for injection systems in accelerating structures with spatially uniform quadrupole focusing (energy 60 keV, current 30 mA, normalized emittance 0.25 mm mrad). The axial distribution of the lens potential, which gives the basic coupling function in parametric form, was approximated by a sixth order polynomial with zeroes of the first and second derivatives at the boundaries. The envelope function corresponding to it is found by numerical integration of Eq. (14) with initial values $r(0)=1.5$ mm, $r'(0)=0$. The shapes of the electrodes that realize the required axial distribution of the potential were determined by the first-approximation equation

$$\Psi = U(z) - 0.25U''(z)R_A^2(z) + Q_U\{1+2\ln[R_A(z)/r(z)]\}/\sqrt{U(z)},$$

where Ψ is the electrode potential and R_A is the aperture radius.

The potential of the middle electrode was chosen so as to form a beam at the input of the system with a radius of about 1 mm and a convergence angle of 10 mrad. Figure 2 shows the characteristics of this system. Since the coupling function was obtained by integration of the ordinary envelope equa-

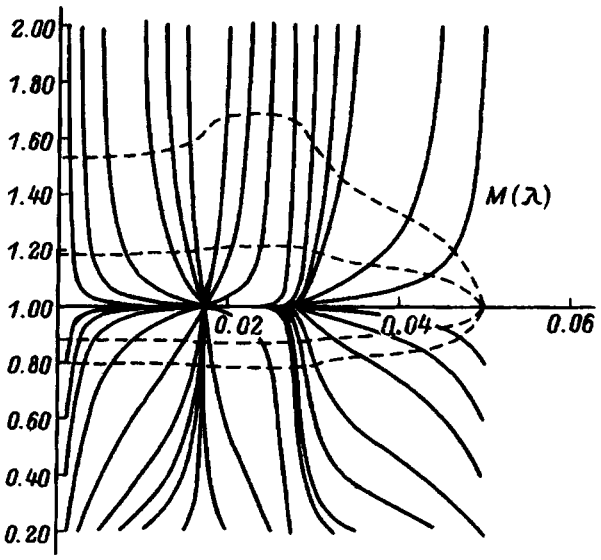

FIG. 3. Integral curves of Eq. (16). The solid curves are for a current of 30 mA; the dashed curves are (from top to bottom) currents of 10, 20, 40 and 50 mA.

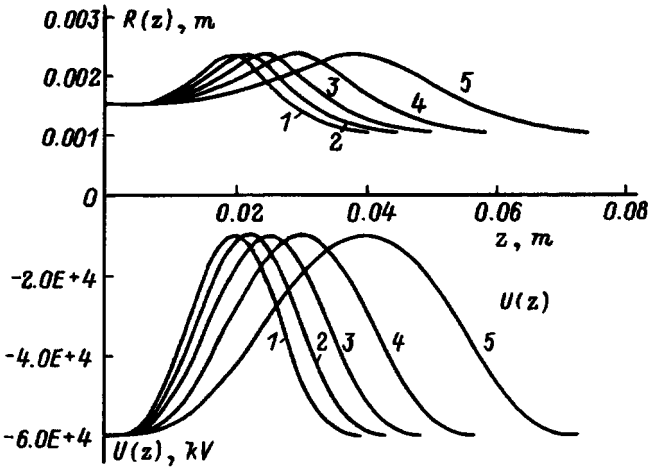

FIG. 5. Changes in the axial potential distribution and envelope as the beam current changes. *1–5* — 50, 40, 30, 20, 10 mA, respectively.

tion, we have $f_3 \equiv f_2$. The solution to Eq. (16) for a current of 30 mA is given by the solid curves in Fig. 3. The integral curve $M=1$ corresponds to the basic realization of the coupling function. The remaining realizations differed in the value of the slope angle of the envelope at the input. On the interval from the nonzero node to the saddle point at the input, all of the realizations coincide. Figure 4 shows the profile of the electrodes with potentials of the basic system that realize a coupling function with an angular beam envelope at the output of 5, 10, and 20 mrad. The dashed curves in Fig. 3 show the solution to Eq. (16) for the same system at different currents; the beam angle at the output is kept equal to 10 mrad. The axial potential distributions obtained from these solutions and the envelopes for the beam corresponding to them are shown in Fig. 5. Figure 6 shows the electrode profiles that allow the same transformation of beam radius and slope angle to be implemented at currents of 20, 30, 40 mA without changing the voltage on the electrodes.

Of course, it is by no means true that every mathematical realization of a coupling function is realizable physically. The approach given here only allows us to formally satisfy a number of constraints imposed on the control. The physical realizability of a control can be improved by correcting the coupling function, most importantly by changing the beam envelope, on whose behavior remarkably few restrictions have been imposed. A still more universal approach is the following procedure, which combines the advantages of a scale transformation with the possibility of explicit specification of the control: 1) determine the basic coupling function by solving the ordinary envelope equation with a physically realizable control; 2) if the control chosen does not ensure the necessary beam transformation, correct the coupling function such that it satisfies the required boundary values; 3) with the help of the generalized equation for scale transformation, find a control that mathematically realizes the corrected coupling function; 4) if the control obtained is not realizable physically, replace it by the closest form that admits this realization and return to point 1.

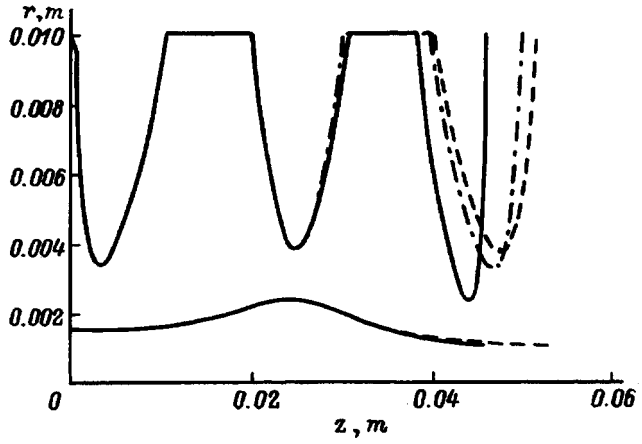

FIG. 4. Profiles of electrodes for shaping a beam with the radius of 1 mm and angular envelope 20 mrad (—), 10 mrad (– · –), 5 mrad (- - -).

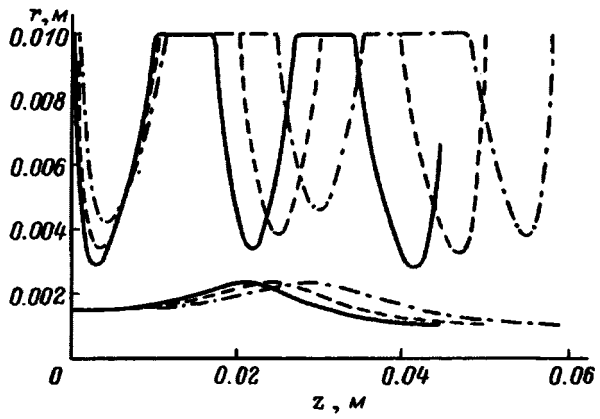

FIG. 6. Geometry of a lens that implements the same beam transformation at currents of 20 (– · –), 30 (- - -), 40 mA (—).

CONCLUSION

The transformation described in this paper leads to a unified approach to solving the problem of matching beams to structures of various types. It is based on introducing a coupling function, which initially defines a relation between the dependent variables of the ordinary envelope equation and is invariant under the transformation. The properties of singularities of the basic transformation equation can be used to formalize the process of design (adjustment) of beam systems based on their first-order optical characteristics, while taking into account particle temperature and space charge.

In many cases, these equations can be used to scale a previously designed system; to redesign it for new values of the current, emittance, beam energy, or new types of particles; to analyze the realizability of constraints on the force-field intensity, channel aperture, length, etc. For this reason, the problem of creating a catalog of standardized (basic) coupling functions that correspond to the type of beam transformations that occur in optical elements is of topical interest. The approach described here can also be used to solve other boundary value problems in dynamics.

[1] G. Swain, P. Busch, and M. Burns, *Proceedings of the 1989 IEEE Particle Accelerator Conference*, Chicago, Ill., 1989, p. 598.
[2] I. P. Yudin, in *Proceedings of the 1993 IEEE Particle Accelerator Conference*, Washington, D.C., 1993, p. 191.
[3] G. E. Lee-Whiting and N. Bezic, Nucl. Instrum. Methods **71**, 61 (1969).
[4] K. Takayama, Part. Accel. **21**, 259 (1987).
[5] J. Picht, in *Einführung in die Theorie der Elektronenoptik*, 2nd ed. (Leipzig, 1957), p. 137.
[6] N. I. Tarantin, *Magnetic Static Charged-Particle Analyzers: Fields and Linear Optics* [in Russian], Énergoatomizdat, Moscow (1986), pp. 31–36.
[7] Yu. V. Zuev, Preprint NIIEFA No. P-0940 [in Russian], TsNII Atominform, Moscow, 1995.

Translated by Frank J. Crowne

On envelope equations for electron beams in magnetic fields

N. D. Naumov

Central Physicotechnical Institute, 141300 Sergiev Posad, Russia
(Submitted November 14, 1996)
Zh. Tekh. Fiz. **68**, 103–107 (January 1998)

A self-similarity approach is used to obtain envelope equations for an annular beam propagating along a magnetic field, and also an electron beam injected at an angle to a magnetic field. An exact solution is constructed for the self-consistent problem of transverse oscillations of a cold annular beam in a magnetic field, and a comparison is made with approximate results from the method of envelope equations. © *1998 American Institute of Physics.* [S1063-7842(98)01901-1]

INTRODUCTION

The use of envelope equations, which make it possible to construct a number of analytical self-consistent models for beam propagation, has led to substantial progress in the description of the dynamics of charged-particle beams.[1–4] The existence of self-similar solutions to the equations of motion for a gas of charged particles makes it possible to convert partial differential equations to ordinary differential equations.[5] For curvilinear beams approximate solutions of self-similar type can be constructed when the ratio of transverse dimensions of the beam to its radius is small.[6] However, in this case the adequacy of such approximate solutions to the nonlinear problem remains an open question.

In this paper transverse oscillations of a cold annular beam of electrons in a magnetic field are discussed. It will be shown that an approximate solution of self-similar type can be constructed for a thin annular beam by using the method of envelope equations, while the Green's function method can be used to derive an exact solution to the nonstationary self-consistent problem. A comparison of these two solutions is of undoubted methodological interest.

In this paper envelope equations for a thin helical beam in a magnetic field are also derived. These equations find practical application in connection with the use of electron beams to study the ionosphere. The injection of an electron beam into an ionospheric plasma at an angle to the geomagnetic field was discussed in Ref. 7, but the results obtained there, as these authors themselves showed, cannot be applied to the case of injection at pitch angles close to 90°. The model derived here fills in this gap for time periods in which the ratio of transverse beam dimensions to its radius of curvature remains a small quantity.

ANNULAR BEAMS

The behavior of an axisymmetry cold beam of nonrelativistic electrons propagating along a magnetic field is described by the self-consistent system of equations of motion for a gas of charged particles

$$\frac{d}{dt}V_r-\frac{V_\theta^2}{r}=-\omega_c V_\theta+\frac{4\pi e^2}{mr}\int_0^r n(x)x\,dx, \tag{1}$$

$$\frac{d}{dt}V_\theta+\frac{V_r V_\theta}{r}=\omega_c V_r, \quad \frac{\partial n}{\partial t}+\frac{\partial nV_r}{\partial r}=0, \tag{2}$$

where $\omega_c=e_0B_0/mc$ is the cyclotron frequency, $e=-e_0$ is the electron charge, and $d/dt=\partial/\partial t+V_r\partial/\partial r$.

It is not difficult to see that there is a stationary solution to the equation for the azimuthal velocity of the form $V_\theta=\omega_L(r+C^2/r)$, where $\omega_L=\omega_c/2$ is the Larmor frequency. The constant C is determined from the condition that particles located a distance $r=C$ from the beam axis rotate with at the cyclotron frequency. This result is a consequence of the fact that the azimuthal component of the generalized momentum of a particle $P_\theta=r(p_\theta+mr\omega_L)=m\omega_L C^2$ is conserved for an axisymmetric beam.

An approximate nonstationary solution can be constructed for the radial equation. As was shown previously,[6] such a solution can be obtained when the ratio of the transverse dimensions of a beam to its radius is small, in which case the terms entering into Eq. (1) that involve the azimuthal velocity can be expanded and only the first-order terms kept with respect to the quantity d/r_1, where r_1 is the inner radius of the beam and d is the beam thickness. The almost linear variation of the self-electric field of the beam can be exploited by choosing a corresponding profile for the beam density:

$$n(\mathbf{x},t)=\frac{I}{2\pi e_0 urd}H(\xi-\xi^2), \tag{3}$$

where I and u are, respectively, the current and longitudinal velocity of the beam, $H(x)$ is the Heaviside step function, and $\xi=(r-r_1)/d$ is the self-similarity variable.

For this class of nonstationary motions, the velocity of the electron gas depends linearly on the self-similar variable; therefore, the radial velocity of the beam has the following form: $V_r=\dot{r}_1+\xi\dot{d}$. In this case the particle density (3) satisfies the equation of continuity (2). Substituting this expression into the linearized equations for the radial velocity (1) leads to an equation for the internal radius of the beam and its thickness:

$$\ddot{r}_1+\omega_L^2\left(r_1-\frac{C^4}{r_1^3}\right)=0, \quad \ddot{d}+\omega_L^2\left(1+3\frac{C^4}{r_1^4}\right)d=d_0\omega_p^2, \tag{4}$$

where $\omega_p=\sqrt{4\pi n_0 e^2/m}$ is the beam plasma frequency and $n_0=I/2\pi u e_0 r_{10} d_0$.

 1063-7842/98/010093-04$15.00

It should be noted that the collective field at the inner surface of the beam equals zero. Therefore, the equation for oscillation of the inner radius of the beam, in contrast to the equation for the beam thickness, is exact.

It is not difficult to see that for a cold beam the inner radius does not depend on time if at the initial time the particles at the inner surface of the beam rotate with the cyclotron frequency, i.e., when $C=r_{10}$. In this case, as follows from Eq. (3), the thickness and outer radius of a thin beam vary periodically with time:

$$r_2=r_{10}+d, \quad d=d_0[g+(1-g)\cos\,\omega_c t], \tag{5}$$

where $g=\omega_p^2/\omega_c^2$.

It is obvious that for $d=1$ a stationary state of the beam is realized.

THE METHOD OF GREEN'S FUNCTIONS

An exact solution can be obtained for the self-consistent problem of transverse motion of an axisymmetric cold beam in a magnetic field. Although the problem as posed is essentially hydrodynamic, it is convenient in solving it to start from the distribution function of the electron gas,

$$F(\chi,P_\theta,p_z,t)=f(\chi,t)\,\delta(P_\theta-m\omega_L C^2)\,\delta(p_z-mu),$$

where the collection of variables r, p_r is denoted by χ for brevity.

The solution to the Vlasov equation for the radial function $f(\chi,t)$

$$Lf(\chi,t)=0,$$

$$L=\frac{\partial}{\partial t}+\frac{p_r}{m}\frac{\partial}{\partial r}+\left[eE+m\omega_L^2\left(\frac{C^4}{r^3}-r\right)\right]\frac{\partial}{\partial p_r}$$

can be written by using the Green's function of the operator L:

$$H(t)f(\chi,t)=\int G(\chi,\chi_0;t)f(\chi_0,0)d\chi_0,$$

$$LG(\chi,\chi_0;t)=\delta(t)\,\delta(\chi-\chi_0).$$

Here $f(\chi,0)$ is the initial function for the beam distribution, which in the present case has the form $f(\chi,0)=n_0 r\,\nu(r)\,\delta(p_r)$, where $n_0\nu(r)=n(r,0)$ is the initial density distribution of particles.

The Green's function is determined by the radial motion of a single particle $r(t;\chi_0)$, $p_r(t;\chi_0)$ in the combined external and collective fields for $P_\theta=m\omega_L C^2$:

$$G(\chi,\chi_0;t)=H(t)\,\delta[r-r(t;\chi_0)]\,\delta[p_r-p_r(t;\chi_0)].$$

The motion of an electron satisfies the following conditions: $r(0;\chi_0)=r_0$, $p_r(0;\chi_0)=p_{r0}$.

The basic difficulty in implementing this method of solving the self-consistent problem is connected with including the influence of the self-field of the beam on the motion of the particles, which is unknown beforehand and which changes as the beam propagates. This problem is simplified if the layers of particles translate in the radial direction in ''single file,'' without overtaking each other. Then the value of the collective field acting on a particle does not depend on time and is determined both by the initial value of its radial position r_0 and by the preset initial density distribution of the particles:

$$E_r[r(t;\chi_0),t]=4\pi e n_0\frac{Q(r_0)}{r}, \quad Q(r_0)=\int_{r_{10}}^{r_0}\nu(x)x\,dx. \tag{6}$$

In summary, we find for the function $f(\chi,t)$

$$f(\chi,t)=n_0\int dr_0 r_0\nu(r_0)\,\delta[r-s(t,r_0)]$$

$$\times\,\delta[p_r-m\dot{s}(t,r_0)],$$

where $s(t,r_0)$ is the solution to the equation

$$\ddot{s}=\omega_L^2\left(\frac{C^4}{s^3}-s\right)+Q(r_0)\frac{\omega_p^2}{s} \tag{7}$$

with the initial condition $s(0,r_0)=r_0$, $\dot{s}(0,r_0)=0$. For the hydrodynamic characteristics of the flow we obtain

$$n(r,t)=\int F(\chi,P_\theta,p_z,t)d^3p=n_0\frac{\rho(r,t)\,\nu[\rho(r,t)]}{r|S(r,t)|}, \tag{8}$$

$$V_r(r,t)=\frac{1}{mn}\int p_r F(\chi,P_\theta,p_z,t)d^3p=U(r,t). \tag{9}$$

Here $\rho(r,t)$ is the solution to the transcendental equation $s(t,r_0)=r$, i.e., $s[t,\rho(r,t)]\equiv r$, and the following notations have been introduced: $S(r,t)=R[t,\rho(r,t)]$, $R(t,r_0)=\partial s/\partial r_0$, $U(r,t)=\dot{s}[t,\rho(r,t)]$. A simpler method can be used to calculate the hydrodynamic characteristics of the flow. It follows from expressions (8), (9), that at time t the particle density and radial velocity of the beam at a point $r=s(t,r_0)$ equal

$$n(s,t)=n_0\frac{r_0\nu(r_0)}{s(t,r_0)|R|}, \quad V_r(s,t)=\dot{s}(t,r_0). \tag{10}$$

Let us find the equation for the function R by differentiating Eq. (7) with respect to r_0:

$$\ddot{R}=\nu(r_0)\omega_p^2\frac{r_0}{s}-\left[\omega_L^2\left(1+3\frac{C^4}{s^4}\right)+Q(r_0)\frac{\omega_p^2}{s^2}\right]R.$$

It is obvious that the initial conditions for R have the form $R_0=1$, $\dot{R}_0=0$.

Thus, calculation of the hydrodynamic characteristics of an axisymmetric cold flow of charged particles in a magnetic field reduces to solving two ordinary differential equations. It is obvious that for a solid beam the upper limit of integration in Eq. (6) should be set equal to zero. For an annular beam, Eq. (10) can be used to obtain the characteristics of the beam at a given time t by stepping r_0 gradually with a sufficiently small step size from r_{10} to r_{20}. Note that when $s_0=r_{10}$, Eq. (7), which in this case becomes the first of Eqs. (4), determines the oscillations of the inner radius of the beam. When $s_0=r_{20}$, Eq. (7) describes the time dependence of the exterior radius of the beam; choosing the initial particle density in the form Eq. (3), i.e., $\nu=r_{10}/r$, we obtain

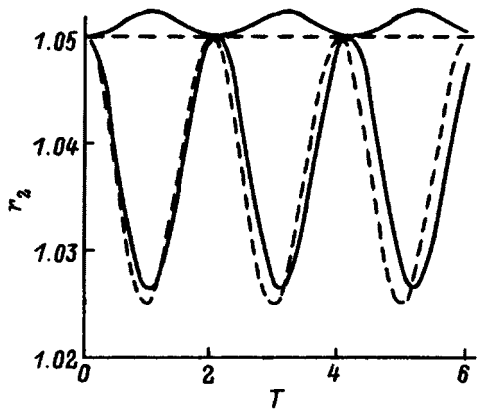

FIG. 1. Oscillations of the external beam radius.

$$\ddot{r}_2=\omega_L^2\left(\frac{C^4}{r_2^3}-r_2\right)+d_0\omega_p^2\frac{r_{10}}{r_2}. \tag{11}$$

Figure 1 shows the results of calculating the external radius of the beam for initial conditions $C=r_{10}$, $r_{20}=1.05r_{10}$ for $g=1$ and 0.75. The solid curves correspond to the solution to Eq. (11), the dashed curves to analogous data from the self-similar approximation Eq. (5); the variable $T=\omega_c t$. We can assert that the equation for envelopes gives a suitable description of the dynamics of the transverse dimensions of the beam at least at the initial stage of motion. Figure 2 shows a comparison of the exact and approximate solutions for particle densities at $g=0.75$ at time $t=5\pi/\omega_c$ (curves *1*, *2*), $t=5.5\pi/\omega_c$ (curves *3*, *4*), and $t=6\pi/\omega_c$ (curves *5*, *6*). For the particle density the exact and approximate results differ from each other to a larger extent.

HELICAL BEAM

Assume that the beam axis coincides with the trajectory of a single electron in the magnetic field, i.e., it is in the form of a vortex line

$$\mathbf{Y}(\sigma)=\frac{1}{k}\sin\alpha\cos k\sigma\mathbf{e}_x+\frac{1}{k}\sin\alpha\sin k\sigma\mathbf{e}_y+\sigma\cos\alpha\mathbf{e}_z.$$

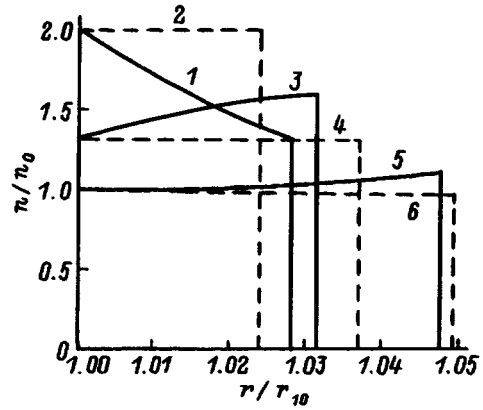

FIG. 2. Variation in particle density.

Here $k=\omega_c/u$, u is the particle velocity, σ is the path length, and α is the angle between the magnetic field and initial electron velocity vectors.

The transverse beam dynamics are conveniently treated in a system of curvilinear coordinates ρ,σ,ζ

$$\mathbf{x}=\mathbf{Y}(\sigma)+\rho\mathbf{e}_r+\zeta\mathbf{b},$$

where $\mathbf{b}=\mathbf{t}\times\mathbf{n}$, and **t, n, b** are the vectors of a Frenet triad connected with the curve $\mathbf{Y}(\sigma)$.

For a vortex line, the direction of the normal vector is opposite that of the radial unit vector $\mathbf{n}=-\mathbf{e}_r$. In this system of coordinates the external field has the form $\mathbf{B}_0=B_0(\mathbf{b}\sin\alpha+\mathbf{t}\cos\alpha)$.

Let the ratio of transverse dimensions of the beam to its radius of curvature $r_c=1/k\sin\alpha$, and also to the radius of torsion $r_t=1/k\cos\alpha$, be small quantities. To accuracy up to first-order terms, the longitudinal velocity of the beam is constant; therefore, setting $\mathbf{V}=u\mathbf{t}+\Gamma\mathbf{e}_r+\Lambda\mathbf{b}$ and neglecting second-order terms, we find from the Euler equation for a charged gas the following equations for the functions Γ,Λ:

$$M\Gamma+f\rho=F_\rho,\quad M\Lambda=F_\zeta,$$

$$M=\frac{\partial}{\partial t}+u\frac{\partial}{\partial\sigma}+(\Gamma-w\zeta)\frac{\partial}{\partial\rho}+(\Lambda+w\rho)\frac{\partial}{\partial\zeta}, \tag{12}$$

where $f=u^2/r_c^2$, $w=u/r_t$, F_ρ, F_ζ are terms that give rise to the self-field and the beam emittance.

Expressions for these terms when the beam has an elliptic cross section are most simply displayed in a system of coordinates q_1, q_2 connected with the axes of symmetry of the beam cross section,[8]

$$F_1=\frac{hq_1}{a(a+b)}+\frac{Hq_1}{a^4},\quad F_2=\frac{hq_2}{b(a+b)}+\frac{Hq_2}{b^4}.$$

Here $h=4Ic^2/I_A\gamma^2$, I is the beam current, $I_A=\gamma umc^2/e_0$ is the Alfvén current, a, b are semi-axes of the beam cross section, and $H=u\varepsilon$, where ε is the beam emittance. The self-field of the beam is approximated by the electromagnetic field of a rectilinear beam with elliptic cross section, since corrections due to the beam curvature will, as in the case of an annular beam,[9] be terms of second order in smallness.

Because the orientation of the beam cross section changes as time passes, the coordinate axes for the system q_1, q_2 will rotate with respect to the unit vectors $\mathbf{e}_r$, **b** through a certain angle ψ

$$q_1=\rho\cos\psi+\zeta\sin\psi,\quad q_2=\zeta\cos\psi-\rho\sin\psi.$$

Accordingly, Γ, Λ should be written in terms of components of the gas velocity V_i in the new system of coordinates:

$$\Gamma=V_1\cos\psi-V_2\sin\psi-\zeta\Omega,$$

$$\Lambda=V_2\cos\psi+V_1\sin\psi+\rho\Omega,$$

where $\Omega=\dot{\psi}$ is the angular velocity of rotation of the beam with respect to the Frenet triad.

If σ is replaced by the variable $\tau=\sigma-ut$, then the derivative with respect to τ in Eq. (12) disappears. For this reason the dependence of the beam characteristics on τ is

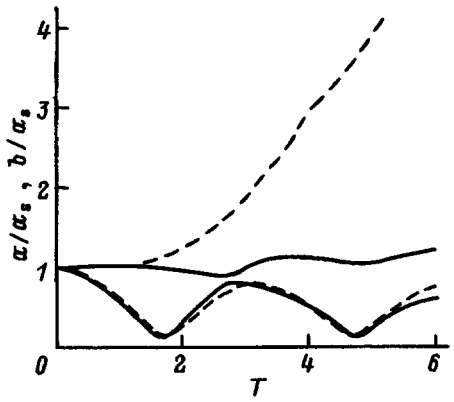

FIG. 3. Variation in the transverse dimensions of the beam.

now parametric in nature and is determined by the initial conditions when a beam with a given cross section is injected. As a result, Eq. (12) for the function V_i leads to the following equations:

$$NV_1-2\Omega V_2-q_2\dot{\Omega}-\lambda q_1+f\times\left(q_1\cos^2\psi-\frac{q_2}{2}\sin 2\psi\right)=F_1, \tag{13}$$

$$NV_2+2\Omega V_1-q_1\dot{\Omega}-\lambda q_2+f\times\left(q_2\sin^2\psi-\frac{q_1}{2}\sin 2\psi\right)=F_2, \tag{14}$$

where

$$N=\frac{\partial}{\partial t}+V_i\frac{\partial}{\partial q_i}+w\left(q_1\frac{\partial}{\partial q_2}-q_2\frac{\partial}{\partial q_1}\right),$$

$$\lambda=\Omega(w+\Omega).$$

The transverse motion of a gas in the new system of coordinates can also include translation along with the elliptic current lines; therefore, the following starting expression should be used for V_i in terms of the self-similar variables $\xi=q_1/a$, $\eta=q_2/b$:

$$V_1=\dot{a}\xi-\omega a\eta, \quad V_2=\dot{b}\eta+\omega b\xi, \tag{15}$$

where ω is some function of time.

The particle density that satisfies the equation of continuity has the form

$$n(\mathbf{x},t)=\frac{I}{\pi abue_0D}H(1-\xi^2-\eta^2),$$

where $D=1+k\rho\cos\alpha$.

Substituting Eq. (15) into Eqs. (13), (14) eventually leads to a system of ordinary differential equations for the time-dependent functions a, b, ω, Ω:

$$\ddot{a}=2\omega\Omega b+\left(\mu+w\omega\frac{a}{b}-f\cos^2\psi\right)a+\frac{h}{a+b}+\frac{H}{a^3}, \tag{16}$$

$$\ddot{b}=2\omega\Omega a+\left(\mu+w\omega\frac{b}{a}-f\sin^2\psi\right)b+\frac{h}{a+b}+\frac{H}{b^3}, \tag{17}$$

$$\dot{\Omega}+2\Omega\frac{\dot{b}}{b}+\dot{\omega}\frac{a}{b}+2\omega\frac{\dot{a}}{b}+w\frac{\dot{a}}{a}+\frac{f}{2}\sin 2\psi=0, \tag{18}$$

$$\dot{\Omega}+2\Omega\frac{\dot{a}}{a}+\dot{\omega}\frac{b}{a}+2\omega\frac{\dot{b}}{a}+w\frac{\dot{b}}{b}-\frac{f}{2}\sin 2\psi=0, \tag{19}$$

where $\mu=\lambda+\omega^2$.

Adding Eqs. (18), (19) and integrating the resulting equation gives the function Ω in the form:

$$\Omega=\frac{1}{2ab}\left[2a_0b_0\left(\Omega_0+\frac{w}{2}\right)+\omega_0(a_0^2+b_0^2)-\omega(a^2+b^2)\right]-\frac{w}{2}. \tag{20}$$

This relation between the angular velocity of rotation of the beam relative to the Frenet triad and the angular velocity of the internal motion of the gas can also be obtained from the condition of conservation of the longitudinal component of the vector $(\mathbf{W}+e\mathbf{B}_0/mc)/n$, where $\mathbf{W}=\mathrm{curl}\,\mathbf{V}$ is the vorticity and n is the gas density.

Figure 3 shows the results of calculating the transverse beam dimensions by numerically solving the system of Eqs. (16)–(19). The following values for the parameters were chosen: $\alpha=3\pi/8$, $\omega_0=0$, $h/(\omega_c a_s)^2=0.05$, $a_0=b_0=a_s$, where a_s is the equilibrium radius of a Brillouin flow propagating along the magnetic field with intensity $B_0\cos\alpha$. This radius is determined from the expression $w^2/4a_s=h/2a_s+H/a_s^3$, since, as is clear from Eq. (20), the ''effective'' Larmor frequency of a helical beam equals $w/2$. The lower curves correspond to a/a_s, the upper b/a_s; the variable $T=\omega_c t$. In these calculations two initial values of angular velocity of the rotating beam were used: $\Omega_0=0$ (dashed curves) and $\Omega_0=-w/2$ (solid curves). These results show that when a rotating beam is injected at an angle to the magnetic field, the spreading of the beam under the influence of space charge is decreased.

[1] I. M. Kapchinsky and V. V. Vladimirsky, in *Proceedings of the International Conference on High Energy Accelerators and Instrumentation*, CERN, Geneva, 1959, p. 274.
[2] O. I. Yarkovoĭ, Zh. Tekh. Fiz. **36**, 988 (1966) [Sov. Phys. Tech. Phys. **11**, 731 (1966)].
[3] E. P. Lee, Phys. Fluids **19**, 60 (1976).
[4] E. P. Lee and R. K. Cooper, Part. Accel. **7**, 83 (1976).
[5] N. D. Naumov, Fiz. Plazmy **19**, 1406 (1993) [Plasma Phys. Rep. **19**, 738 (1993)].
[6] N. D. Naumov, Zh. Tekh. Fiz. **67**(7), 103 (1997) [Tech. Phys. **42**, 814 (1997)].
[7] G. V. Lizunov and A. A. Silivra, Geomagnetism and Aeronomy **28**, 980 (1988) [in Russian].
[8] J. D. Lawson, *The Physics of Charged-Particle Beams* [Clarendon Press, Oxford (1977); Mir, Moscow (1980)].
[9] V. P. Sarantsev and E. A. Perel'shteĭn, *Collective Acceleration of Ions by Electron Rings* [in Russian], Atomizdat, Moscow, 1979.

Translated by Frank J. Crowne

A study of the microstructure of thin *a*-C:H layers that orient liquid crystals

E. A. Konshina and T. S. Turovskaya

S. I. Vavilov State Optical Institute All-Russia Science Center, 199034 St. Petersburg, Russia
(Submitted November 21, 1996)
Zh. Tekh. Fiz. **68**, 108–110 (January 1998)

The surface structure of thin layers of *a*-C : H is studied by the method of island decoration, using silver films. The changes in the dimensions and density of silver particles is tracked by an electron microscope as a function of the thickness of the *a*-C : H layers. The electrical microprofile is discussed, along with the nature of the silver crystallization centers on the surface of the *a*-C : H layers. © *1998 American Institute of Physics.* [S1063-7842(98)02001-7]

The surface microstructure of layers of amorphous hydrogenated carbon (*a*-C : H) is being studied as background for the investigation of the orientation-sensitive interaction of these layers with liquid crystals. Previous papers have shown that thin layers of *a*-C : H obtained by chemical deposition of toluene vapors in a glow-discharge plasma give rise to planar orientation of liquid crystals based on cyanobiphenyls.[1,2] *a*-C : H has a spatially irregular structure in which the average ordering is determined by π-bonded clusters consisting of carbon atoms in the sp^2 state linked by a common system of conjugated multiple bonds. Along with polycyclic aromatic groups, polyene chains can also enter into the composition of these clusters.[3] Individual carbon clusters can be observed by scanning tunnel microscopy in amorphous carbon films obtained by deposition of graphite, which do not contain hydrogen.[4,5] In contrast to this, studies of orienting *a*-C : H layers deposited on polycrystalline electrode layers (based on indium and tin oxides) by electron and tunnel microscopy have not revealed any distinctive features of their structure. These microscopic investigations show that the thin layer of *a*-C : H obtained from a carbon-hydrogen plasma simply repeats the relief of the surface on which it is deposited.[6] This hinders the study of distinctive features of the surface structure of *a*-C : H by direct methods.

The goal of this paper is to investigate the morphology of structures at the surface of thin *a*-C : H layers by the method of island decoration using silver films.

The method of decoration, which is based on selective crystallization of a decorating material at a surface of a solid body, makes it possible to identify local active centers of the surface and its electrical microprofile.[7] In this work the decoration material we used was island-like films of silver with thicknesses of $\sim$50 Å, which were deposited by thermal evaporation in vacuum. In order to eliminate the influence of the evaporation rate and substrate temperature on the shape and size of particles of the metal condensate,[8] the silver was crystallized at room temperature with all other conditions kept the same. This allowed us to relate the change in the shape and size, and also density distribution, of the silver particles to the structural features and electrical microprofile of the surfaces of our samples. We used a luminescence electron microscope and the method of replicas to observe the morphological structure of the silver films.

The *a*-C : H layers investigated in this work were obtained by chemical deposition of hydrocarbon vapors in a dc glow-discharge plasma. The substrates were polished in order to eliminate the influence of their structure on the decoration pattern, and oriented obliquely at an angle of 5° to the axis of the apparatus in accordance with the standard technology for depositing orienting *a*-C : H layers. The deposition of the *a*-C : H layers was carried out at the temperature of the surrounding medium from toluene vapor at a discharge power of 1.8 W and pressure 0.08 Pa.

Microphotographs of the island-like silver films at the surfaces of germanium and *a*-C : H layers are shown in Fig. 1. The deposition of an *a*-C : H layer on the germanium changes the conditions for silver crystallization in an important way. Indications of this are the difference in density and dimensions of the decorating particles seen in Figs. 1(a) and 1(b) (at 50 000× magnification). The silver particles at the surface of a $\sim$600 Å thick *a*-C : H layer are spherical in shape [Fig. 1(b)]. When the thickness of the *a*-C : H layer is increased by a factor of two, a change is observed in the shape and size of the silver particles [Fig. 1(c), 50 000× magnification]. Particles with irregular shapes appear due to coalescence of neighboring spherical particles. In a number of cases, the observed changes in the silver particle density were associated with defects in the layer structure. Figure 1(d) (40 000× magnification) shows an example of a characteristic defect with a radial distribution of silver particles. The cause of this defect could be electrical breakdown at the substrate surface during condensation, which leads to a local change in the structure of the *a*-C : H layer.

Figure 2 shows the results of statistical processing of microphotographs of the silver island films at the surface of *a*-C : H layers. Depositing an *a*-C : H layer onto a germanium surface decreases the average size of the decorating silver particles from 300 ± 140 to 120 ± 30 Å. As the thickness of the *a*-C : H layer increases an increase in the silver particle dimensions is observed [Fig. 2(a)]. In this case there is a slight decrease in the density of the silver particle distribution on the *a*-C : H surface, from 1.9×10^{11} to $1.4\times10^{11}\text{cm}^{-2}$, due to the increase in particle size [Fig. 2(b)]. We observed no significant structural differences in the morphology of silver island films on the surfaces of *a*-C : H layers obtained by orienting the substrates obliquely and normally in the plasma. The films were isotropic, which is evidence that the deposition of hydrocarbon vapor in the plasma onto an obliquely oriented substrate does not lead to

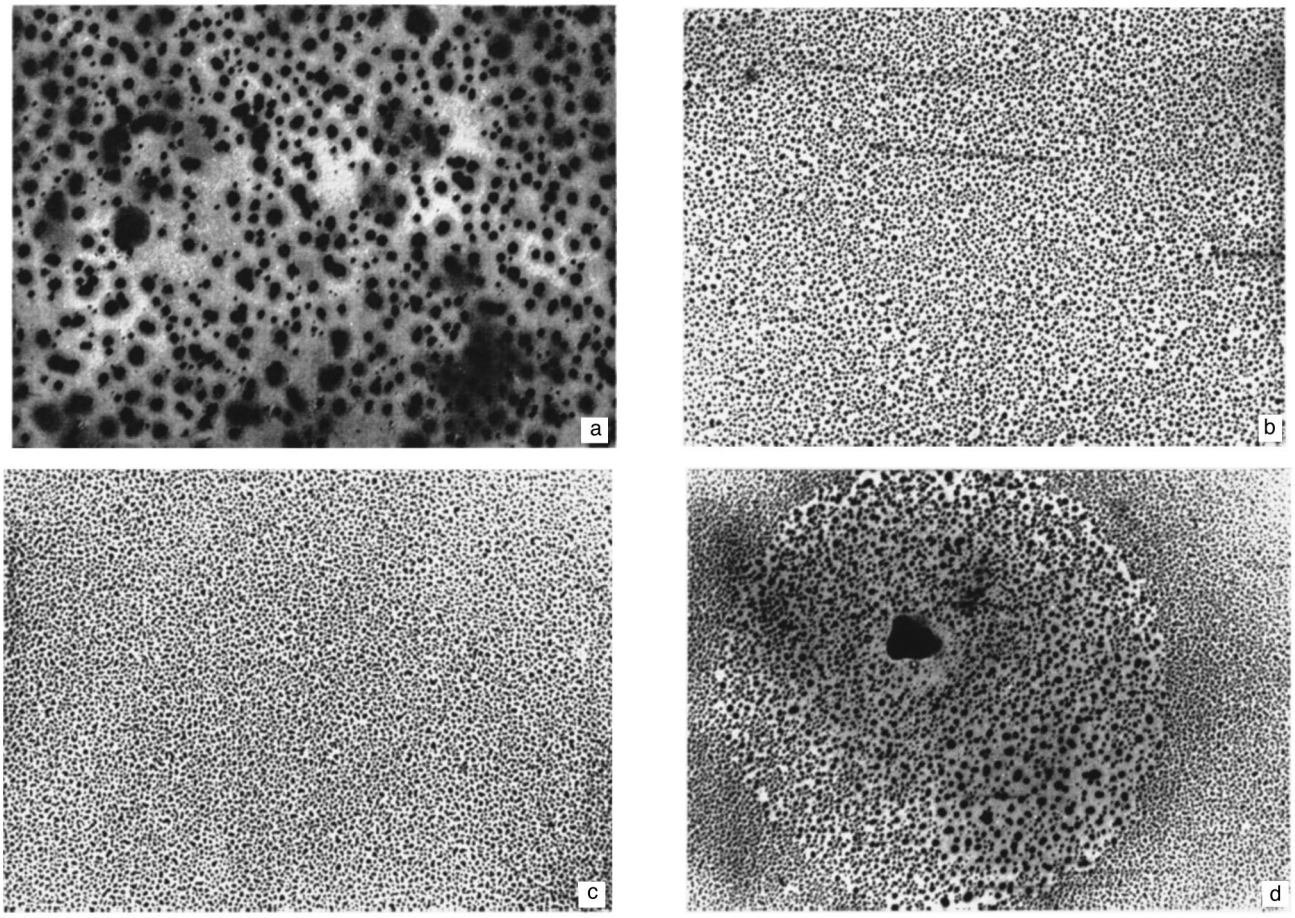

FIG. 1. Microphotographs of the surfaces of germanium (a) and a-C : H layers (b, c, and d) decorated by island-like silver films with thickness ~ 50 Å.

anisotropy of the surface of the a-C : H layers, as happens with condensates of inorganic materials.

The morphological structure of the island-like films of silver depends on the surface properties of the substrate material and the interaction between the metal particles and the solid surface. The authors of Ref. 9 explained the shape and size of silver spheroids at the surfaces of glassy carbon and pyrographite by invoking the influence of an electromagnetic enhancement mechanism based on surface plasmons. For pyrographite, increasing the film thickness results in an increase in the diameter of the silver spheroids, a property shared by a-C : H. However, it is noteworthy that the morphological structure of silver films at the surface of our samples of a-C : H different from that of the island-like silver films at the surface of pyrographite and glassy carbon. The silver particles at the surface of pyrographite and glassy carbon are irregular in shape. For these materials, the mean length and width of 60-Å-thick island-like films at the surface of glassy carbon were 360 and 230 Å, while at the surface of pyrographite they were 760 and 490 Å respectively. The dimensions of the silver particles we observed on the surface of a-C : H were smaller, with an average size in the range 110–220 Å. The size of the smallest of these silver particles (~ 20 Å) matches that of the small graphite domains (~ 15 Å) observed in amorphous carbon films by scanning tunneling microscopy.[5]

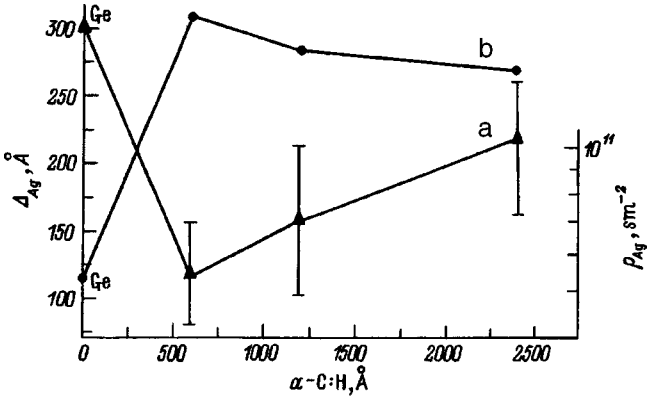

FIG. 2. Dependence of the average size (a) and average density (b) of silver particles at the surface of a-C : H layers on the thickness of the latter.

The electrically active centers of silver crystallization at the surfaces of pyrographite and glassy carbon could be graphite structures oriented parallel to the surface. However, in our view the current carrying regions of the a-C : H surface include not only graphitelike clusters but also isolated sets of several closely spaced clusters that share a common system of delocalized π electrons. These could explain the large scatter in the dimensions of silver particles at the surface of a-C : H from 20 to 400 Å.

The interstitial regions between silver particles at the

surface of a-C : H correspond to portions of the structure that do not carry an electric current. These could be hydrocarbon clusters containing CH, CH_2 and CH_3 groups in the sp^3 state,[10] in which delocalization of π electrons is impossible. In thin a-C : H films [Fig. 1(b)] the distances between individual silver particles range from 30 to 300 Å. When the thickness of the a-C : H layer is increased to $\sim$1200 Å, the distance between decorating particles decreases to 20–60 Å [Fig. 1(c)], which could be explained by the influence of the layer bulk properties on the electrical surface profile. The structure of the a-C : H surface can be viewed as an ''archipelago of islands'' made up of individual π clusters segregated into groups with delocalization of the π electrons and separated by ''channels'' of electrodeless space. The electron density should have its maximum value near the carbon atoms in the sp^2 state that enter into π-bonded clusters, and drop to zero at its boundaries.

By using the method of decoration, we can identify the electrical profile and establish the nature of active centers at the surface of orienting layers. The results we have obtained lead us to assert that the electrical microprofile of the surface of orienting a-C : H layers has an isotropic structure independent of the layer thickness or position of the substrate in the plasma. In our picture, the centers of silver crystallization at the surface of a-C : H are individual π clusters and segregated groups of these with a common system of delocalized π electrons that possess electrical activity. The observed increase in the size of the silver particles as the thickness of the a-C : H layer increases attests to the influence of clusters located in the bulk on the electrical profile of the surface. Our results allow us to formulate a hypothesis regarding the effect of the π-bonding interaction on the orientation of liquid crystals, whose molecules contain aromatic rings, at the phase boundary with an a-C : H layer. We claim that for a liquid crystal of the cyanobiphenyl class, the planar orientation observed is made possible by the interaction of π electrons from the biphenyl core of the molecule with π electrons of the polycyclic aromatic groups. The latter, as Baranov *et al.* have observed,[11] are oriented parallel to the a-C : H surface. Understanding of the interaction mechanism at the boundary of this solid-liquid crystal system will enable us to perfect the technology for orienting liquid crystal molecules, and thereby improve the characteristics of devices based on them.

The studies described in this paper were made possible thanks to partial support from the International Science Foundation (Grant No. NXQ300).

[1] E. A. Konshina, Kristallografiya **40**, 1074 (1995) [Crystallogr. **40**, 999 (1995)].
[2] E. A. Konshina, Proc. SPIE **2731**, 20–24 (1995).
[3] A. V. Baranov and E. A. Konshina, Opt. Spectrosc. **65**, 856 (1988) [Opt. Spectrosc. (USSR) **65**, 506 (1988)].
[4] S. Bhargava, H. D. Bist, S. B. Samanta *et al.*, Solid State Commun. **90**, 205 (1994).
[5] B. Marchon, M. Salmeron, W. Siekhaus *et al.*, Phys. Rev. B **39**, 12907 (1989).
[6] M. V. Isaev, E. A. Konshina, A. P. Onokhov, and T. S. Turovskaya, Zh. Tekh. Fiz. **65**(10), 175 (1995) [Tech. Phys. **40**, 1071 (1995)].
[7] *Decoration of the Surfaces of Solids* [in Russian], Nauka, Moscow, 1976.
[8] L. S. Palatnik, M. Ya. Fuks, and V. M. Kosevich, *Generation Mechanism and Substructure of Deposited Films* [in Russian], Nauka, Moscow, 1972.
[9] H. Ishida, H. Fukuda, G. Katagiri, and A. Ishitani, Appl. Spectrosc. **40**, 322 (1986).
[10] E. A. Konshina, Opt. Mekh. Promst., No. 2, p. 15 (1987) [Sov. J. Opt. Technol. **54**(2), 80 (1987)].
[11] A. V. Baranov, Ya. S. Bobovich , E. A. Konshina, and T. S. Turovskaya, Opt. Spectrosc. **67**, 456 (1989) [Opt. Spectrosc. (USSR) **67**, 267 (1989)].

Translated by Frank J. Crowne