TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Thermodynamic theory of the equation of state of materials

M. F. Sarry

All-Russia Scientific-Research Institute of Experimental Physics, 607190 Sarov, Nizhegorod District, Russia

(Submitted April 4, 1995; resubmitted April 15, 1998)
Zh. Tekh. Fiz. **68**, 1–9 (October 1998)

A universal derivation of the thermodynamic equations on the basis of a combined analysis of the exact relations for any material — the virial theorem, the shock adiabat, and the differential thermodynamic identity relating the thermic and caloric equations of state of matter — is given. This combination makes it possible to reduce the fundamental problem of thermodynamics to a boundary-value problem of mathematical physics. Analytic relations $T_s = T(P_s, \rho_s)$ and $T_s = T(D, u)$ are obtained for classical systems. © *1998 American Institute of Physics.* [S1063-7842(98)00110-X]

INTRODUCTION

It is impossible to obtain on the basis of thermodynamics an equation for calculating the thermodynamic parameters of a material, even in the case of the simplest thermodynamic system — an ideal gas with constant specific heat, i.e., a perfect gas (PG), since thermodynamics gives only thermodynamic identities which relate some thermodynamic parameters of the system of interest with others. For example, the identity

$$dE_c(V,T=0) = -P_c(V,T=0)\,dV, \tag{1}$$

which is valid for any cold system, makes it possible to calculate, for example, the cold internal energy $E_c(V)$ (or the cold pressure $P_c(V)$) of a system only if its cold pressure function $P_c(V)$ (or, conversely, $E_c(V)$) is known, since then the identity (1) can be integrated and a solution for $E_c(V)$ can be found:

$$E_c(V) = E_c(V_0) - \int_1^2 P_c(V)\,dV; \quad 1 \equiv V_0, \quad 2 \equiv V,$$

$$E(V,T) \equiv E_c(V,T=0) + E_t(V,T),$$

$$P(V,T) \equiv P_c(V,T=0) + P_t(V,T). \tag{1a}$$

If, on the other hand, $E_c(V)$ is given, then $P_c(V) = -dE_c(V)/dV$ according to Eq. (1). Statistical physics most often needs to be invoked to obtain information about nonthermodynamic sources. Ultimately, the thermodynamic functions of a material turn out to be expressed in terms of exclusively statistical concepts, for example, in terms of two-particle distribution functions $n_2(r)$ (Refs. 1 and 2):

$$E(V,T) = (3/2)RT + (V/2)\int U(r)\,n_2(r)\,4\pi r^2 dr, \tag{2}$$

$$VP(V,T) = RT - (V/6)\int r[\partial U(r)/\partial r]\,n_2(r)\,4\pi r^2 dr. \tag{3}$$

These formulas are suitable for describing classical systems[1] (i.e., systems for which the kinetic energy can be identified with the kinetic energy of random motion of the constituent particles) if the two-particle interaction law operating between the structural units (electrons, atoms, ...) and the two-particle distribution function are known and can be expressed in the form

$$U(\mathbf{r}_1,\mathbf{r}_2) \to U(|\mathbf{r}_1 - \mathbf{r}_2|); \quad n_2(\mathbf{r}_1,\mathbf{r}_2) \to n_2(|\mathbf{r}_1 - \mathbf{r}_2|).$$

Therefore it is sufficient to know only $U(r)$ and $n_2(r)$ to construct the thermodynamics of such systems. In practice, however, it is very difficult to calculate $n_2(r)$ even very approximately. An approach leading to purely thermodynamic differential or integral (as desired) equations for the pressure $P(V,T)$ and internal energy $E(V,T)$ of the thermodynamic system of interest is expounded below, i.e., the problem of calculating $P(V,T)$ and $E(V,T)$ is reduced to solving a corresponding boundary-value problem of mathematical physics. The idea of this approach[3] is to convert the differential identity

$$T(\partial P/\partial T)_v = (\partial E/\partial V)_T + P(V,T), \tag{4}$$

relating two thermodynamic functions of a system — its internal energy and pressure — into a differential equation for one of them by using nonthermodynamic relations.

GENERAL EQUATIONS FOR THE PRESSURE AND INTERNAL ENERGY

Of the relations which are widely used in the theory of the equation of state of matter, one is a purely phenomenological relation, proposed by Mie and Grüneisen,[4,5] between the thermal pressure and the thermal internal energy of the system

$$P_t(V,T) = \Gamma(V)[E_t(V,T)/V]. \tag{5}$$

This relation can be used to obtain expressions for the derivatives $(\partial E_t/\partial V)_T$ and $(\partial P_t/\partial T)_v$, which can then be substituted (in turn!) into the thermal analog of the identity (4):[3]

$$T(\partial P_t/\partial T)_v = (\partial E_t/\partial V)_T + P_t(V,T) \tag{4a}$$

to obtain an equation for $P_t(V,T)$ or $E_t(V,T)$, respectively. These equations have the form (notation: $\varkappa(V) \equiv (\Gamma + \Gamma^2 - d\Gamma/d\ln V)/\Gamma^2$)

1063-7842/98/43(10)/8/$15.00

$$(\partial \ln P_t/\partial \ln T)_v - [1/\Gamma(V)](\partial \ln P_t/\partial \ln V)_{T\neq 0} = \varkappa(V),$$

$$(\partial \ln E_t/\partial \ln T)_v - [1/\Gamma(V)](\partial \ln E_t/\partial \ln V)_{T\neq 0} = 1. \quad (6)$$

To solve these equations uniquely it is necessary to prescribe boundary conditions $P_t(V,T_0)=f(V)$ or $P_t(V_0,T) \equiv \varphi(T)$, where $f(V)$ and $\varphi(T)$ are known functions (for example, taken from experimental data or other considerations), and $\Gamma(V)$; similar conditions must also be prescribed in order to solve the equation for $E_t(V,T)$. Equations (6) can be written in the form of integrodifferential equations:

$$P_t(V,T) = P_t(V,T_0) + [\varkappa(V) + (1/\Gamma)\partial/\partial \ln V]\int_1^2 P_t(V,x)\,dx/x,$$

$$E_t(V,T) = E_t(V,T_0) + [1 + (1/\Gamma)\partial/\partial \ln V]\int_1^2 E_t(V,x)\,dx/x;$$

$$1 \equiv T_0; \quad 2 \equiv T. \quad (6a)$$

For $E_t(V,T)$ one can also write an equation for a different boundary curve:

$$E_t(V,T) = E_t(V_0,T) - [1 - \partial/\partial \ln T] \times \int_1^2 E_t(x,T)[\Gamma(x)/x]\,dx;$$

$$1 \equiv V_0; \quad 2 \equiv V. \quad (6b)$$

The analogous formula for the thermal part of the pressure is more complicated.

To obtain the complete functions $P(V,T)$ and $E(V,T)$ the cold parts $P_c(V)$ and $E_c(V)$ must be added to the parts $P_t(V,T)$ and $E_t(V,T)$ found above:

$$P(V,T) = P_c(V,T=0) + P_t(V,T);$$

$$E(V,T) = E_c(V,T=0) + E_t(V,T). \quad (7)$$

The virial theorem[3] (see Eq. (A4) in the Appendix) could be used as another nonthermodynamic relation, but in the general case it is inconvenient because besides the entire internal energy it also includes the kinetic part, i.e., in the general case the virial theorem has the form $F(P,E,E_{\text{kin}},V)=0$, and therein lies its inconvenience — separate equations will have to be written out for the kinetic and potential parts of the pressure and energy, and this engenders additional difficulties in prescribing the boundary conditions. For this standpoint the most convenient nonthermodynamic relation is the shock adiabat (SA)[4,5]

$$2(E_s - E_0) = (P_s + P_0)(V_0 - V_s). \quad (8)$$

Here P_s, E_s, and V_s are thermodynamic parameters of the post-shock state of the substance and must certainly satisfy the identity (4). The derivatives required to transform Eq. (4) into an equation for $P(V,T)$ or $E(V,T)$ (for a solid P_0 can be set equal to zero) can be obtained from Eq. (8):

$$2(\partial P/\partial \ln T)_v = P - P_0 - [\partial P/\partial \ln(V_0 - V)]_T \quad (9)$$

and a similar equation for the internal energy

$$2(\partial E/\partial \ln T)_v = 2(E - E_0) - [\partial E/\partial \ln(V_0 - V)]_T - (V_0 - V)P_0. \quad (10)$$

These equations can be converted into integrodifferential equations

$$P(V,T) = P(V,T_0) + (1/2)[1 - \partial/\partial \ln(V_0 - V)] \times \int_1^2 P(V,x)\,dx/x - (1/2)P_0 \ln(T/T_0)$$

$$1 \equiv T_0 \quad 2 \equiv T, \quad (9a)$$

$$E(V,T) = E(V,T_0) + [1 - (1/2)\partial/\partial \ln(V_0 - V)] \times \int_1^2 E(V,x)\,dx/x - [E_0 + (1/2) \times (V_0 - V)P_0]\ln(T/T_0),$$

$$1 \equiv T_0, \quad 2 \equiv T. \quad (10a)$$

Equations of the type (9a) and (10a) but with integrals over the volume have computational peculiarities. For example, the equation of the type (9a) for the pressure $(1 \equiv V_0;\ 2 \equiv V)$ is

$$P(V,T) = P(V_0,T) + [1 - 2\partial/\partial \ln T] \times \int_1^2 P(T,t)\,dt/(t - V_0) - P_0 \times \ln[(V - V_0)/(V_0 - V_0)]$$

and similarly for the internal energy of the material. Equations (9) and (10) or their integrodifferential forms are most useful, since they give the complete functions, and for their unique solution one also requires boundary curves for the complete functions $P(V,T_0)$ and $E(V,T_0)$.

However, there exist systems for which the virial theorem permits obtaining the complete equations for their pressure and internal energy. These are the classical systems already mentioned above.[1] Their kinetic energy can be written in the form

$$E_{\text{kin}}(T) = RT/(\gamma - 1), \quad (11)$$

and then the virial theorem (A4) for them assumes the form

$$3PV = 3RT + [E(V,T) - RT/(\gamma - 1)]\Sigma_n n. \quad (12)$$

Hence the isothermal derivative required to transform the identity (4) into an equation is found immediately:

$$(\partial E/\partial V)_T = (3/\Sigma_n n)[P + V(\partial P/\partial V)_T]. \quad (13)$$

Substitution of this expression into the identity (4) gives a first-order linear differential equation for the pressure of the system:

$$(\partial \ln P/\partial \ln T)_v - \varkappa_1(\partial \ln P/\partial \ln V)_T = \varkappa_2,$$

$$\varkappa_1 \equiv (3/\Sigma_n n); \quad \varkappa_2 \equiv (3 + \Sigma_n n)/\Sigma_n n. \quad (14)$$

A first-order equation for the internal energy of the system can be derived completely analogously, eliminating from the identity (4) the pressure and its isochoric derivative by means of the virial theorem (12):

$$(\partial \ln E/\partial \ln T)_v - \varkappa_1(\partial \ln E/\partial \ln V)_T = 1. \tag{15}$$

Here it is worth mentioning that a relation between $P_c(V)$ and $E_c(V)$ sufficient for closing the relation (1) (see Appendix 2 for the dimensional method of calculating $P_c(V)$ and $E_c(V)$) can be extracted from expression (12) for the virial theorem.

ANALYTICAL SOLUTION OF THE EQUATION FOR THE PRESSURE

As an illustration it could be helpful to solve analytically, at least in part, Eq. (14) for the pressure, for example, under the condition that the function $P(V,T_0)\equiv f(V)$ is given. Such equations can be solved by two methods: the method of characteristics and the method of separation of variables. For solving Eq. (14) by the method of characteristics it is convenient to rewrite this equation in the equivalent form

$$T(\partial P/\partial T)_v - \varkappa_1 V(\partial P/\partial V)_T = \varkappa_2 P. \tag{14a}$$

The equations of the characteristics of this equation are $dT/ds = T$ and $dV/ds = -\varkappa_1 V$. Their solutions are $T(s) = C_1\exp(s)$ and $V(s) = C_2\exp(-\varkappa_1 s)$. To determine C_1 and C_2 one can assume that $s=0$, and, introducing new coordinates s and t, require s to vary along the characteristics and t to vary along the volume axis (the plane with abscissa V and ordinate T will be called the (V,T) plane). These conditions can indeed be satisfied, since in the case of first-order linear equations the initial condition given at some point on the V axis will go over into the (V,T) plane along only one line — the characteristic. These conditions on s and t are met by the functions $T(s) = T_0\exp(s)$ and $V(s) = t\exp(-\varkappa_1 s)$. By eliminating the parameter s from these equations, it is easy to obtain t — a parametric equation of the characteristics in the (V,T) plane

$$T(V;T) = T_0(t/V)^{1/\varkappa_1}. \tag{16}$$

An equation for P is obtained along the characteristics (16):

$$(dP/ds) = \varkappa_2 P \rightarrow P(s;t) = P(s=0;t)\exp(\varkappa_2 s),$$

and therefore the solution of the initial equation (14a) in the (V,T) plane is

$$P(V,T) = f[V(T/T_0)^{\varkappa_1}](T/T_0)^{\varkappa_2}. \tag{17}$$

To solve Eq. (14) by the method of separation of variables it is convenient to rewrite this equation in a different, equivalent form:

$$(1/\varkappa_2)(\partial \ln P/\partial \ln T)_v - (\varkappa_1/\varkappa_2)(\partial \ln P/\partial \ln V)_T = 1. \tag{14b}$$

If the solution of this equation is sought in the special form $P(V,T) = \Phi_1(V)\Phi_2(T)$, then the preceding equation assumes the form $T\Phi_2'/\Phi_2 = \varkappa_2 + \varkappa_4 V\Phi_1'/\Phi_1 \equiv \lambda$. Hence it is easy to obtain

$$\Phi_1(V) = \Phi_1(V_0)(V/V_0)^{\varkappa_5}; \quad \Phi_2(T) = \Phi_2(T_0)(T/T_0)^{\lambda},$$

where $\varkappa_2 = 1+\varkappa_1$; $\varkappa_3 \equiv \varkappa_1/\varkappa_2$; $\varkappa_4 \equiv \varkappa_2\varkappa_3$; $\varkappa_5 \equiv (\lambda - \varkappa_2)/\varkappa_4$.

Then the complete analytical solution of Eq. (14b) has the form

$$P(V,T) = P_0(V/V_0)^{\varkappa_5}(T/T_0)^{\lambda}, \quad P_0 \equiv P(V_0,T_0). \tag{18}$$

The boundary condition $P(V,T_0) = f(V)$ makes it possible to calculate the constants λ and P_0. Indeed, the boundary condition in the form

$$P_0(V/V_0)^{\varkappa_5} = f(V) \equiv A(V/V_0)^q \tag{19}$$

directly yields the two equalities $P_0 = A$ and $\varkappa_5 = q$, since relation (19) should hold for arbitrary V. The equality $\varkappa_5 = q$ in turn determines the separation parameter $\lambda = \varkappa_1 q + \varkappa_2$. Then the complete solution assumes the form

$$P(V,T) = f(V)(T/T_0)^{q\varkappa_1+\varkappa_2}. \tag{20}$$

In this solution, which can be compared with the solution (17), the temperature dependence is singled out explicitly. This fact can be used to obtain from the identity (4) an explicit expression for the internal energy also:

$$E(V,T) = E(V_0,T) + (q+1)\varkappa_1(T/T_0)^{q\varkappa_1+\varkappa_2}\int_{v_0}^{v} f(V)\,dV. \tag{21}$$

Here only $E(V_0,T)$ is unknown, and once again it must be taken from experiment. Thus, to calculate the pressure $P(V,T)$ and internal energy $E(V,T)$ of a classical system uniquely it is necessary to prescribe two boundary conditions of the type $P(V,T_0)\equiv f(V)\equiv A(V/V_0)^q$ and $E(V_0,T) \equiv \chi(T)$.

TEMPERATURE OF AN ADIABATICALLY COMPRESSED SUBSTANCE

In adiabatic compression processes $(dQ = dE + PdV = 0)$ the question of the final temperature of the material has a special place. In the case of a perfect gas there exist exact theoretical formulas for the gas temperature: gradual (static) adiabatic compression of a perfect gas,

$$TV^{\gamma-1} = \text{const} \rightarrow T = T_0(\rho/\rho_0)^{\gamma-1} = T_0\sigma^{\gamma-1} \tag{22}$$

and shock (dynamic) adiabatic compression of a perfect gas,

$$T = T_0P'(\varkappa + P')/(\varkappa P' + 1);$$

$$\varkappa = (\gamma+1)/(\gamma-1); \quad P' = P/P_0. \tag{23}$$

In the case of gradual adiabatic compression of an arbitrary body, however, there is only a quadrature formula for calculating the gas temperature exists,[3,4]

$$T = T_0\exp\left[-\int_1^2(\partial P/\partial E)_v\,dV\right] = T_0\exp\left[-\int_1^2(\partial P_t/\partial E_t)_v\,dV\right]; \quad 1\equiv V_0; \quad 2\equiv V, \tag{24}$$

and to calculate the temperature, in practice, it is necessary to have the pressure as a function of volume and energy $P=P(V,E)$ in a static process (along the Poisson adiabat). This function is known only in the case of a perfect gas:

$$E(V,P)=VP/(\gamma-1)\rightarrow P=(E/V)(\gamma-1), \tag{25}$$

where $\gamma-1$ is the Grüneisen coefficient for a perfect gas.

Then the isentrope (24) goes over to Eq. (22). However, the isentrope (24) can be somewhat simplified, if the Mie–Grüneisen relation (5) is used:

$$T=T_0\exp\left[-\int_1^2\Gamma(V)dV/V\right];\quad 1\equiv V_0;\quad 2\equiv V. \tag{26}$$

It is important that the combination of the virial theorem (12) of classical systems with the SA (8), which is valid for any substance and in any aggregate state of the substance, makes it possible to calculate purely analytically the temperature of a shock-compressed classical system also, if, of course, P_s and V_s are known. Here it should be noted that in the Feynman formulas (2)–(3) the expression $E_{\text{kin}}=(3/2)RT$ is used for E_{kin} instead of expression (11), i.e., E_{kin} of a system consisting of structural units for which $\gamma=5/3$. According to Feynman this is because in the case of condensed classical systems it is ''very difficult'' in practice to take into account the internal degrees of freedom of the structural units. For this reason, Feynman assumes that such systems consists of inert-gas particles and, setting $\gamma=5/3$, that the structural units of the systems possess only translational degrees of freedom. If $\gamma=5/3$ in the virial theorem (12) also and a purely Coulomb potential is used (in this case one must set $\Sigma_n=1$), then the virial theorem (12) assumes the very simple form

$$6PV=2E+3RT. \tag{12a}$$

Then, combining this relation with the SA (8) yields

$$T_s-T_0=(7V_s-V_0)P_s/3R, \tag{27}$$

where P_s and V_s must be taken from shock experiments, $V_0=10^{-6}(\mu/\rho)m^3$ is the molar volume of the substance, μ is the molar mass (in grams) of the substance (m equals numerically the molecular mass of the material), ρ is the normal density of the substance (in $\text{g}\cdot\text{cm}^{-3}$), $R=8.31$ J/mole$\cdot$K, $T_0=300$ K, and the difference T_s-T_0 is in K or °C .

Similarly, for the case of the virial theorem (12)

$$\begin{aligned}T_s'={}&1+\{(\gamma-1)\Sigma_n n/2[3(\gamma-1)-\Sigma_n n]\}\\&\times(P_0V_0/RT_0)\{(P_s'V_s'-1)\\&\times[(6+\Sigma_n n)/\Sigma_n n]-(P_s'-V_s')\}.\end{aligned} \tag{28}$$

In Eq. (28) $P_s'\equiv P_s/P_0$ and $V_s'=V_s/V_0$. Switching in Eq. (28) to a perfect gas (which requires taking the limit $\Sigma_n\rightarrow\infty$) yields the well-known form[4]

$$T_s'=1+[(\gamma-1)P_0V_0/2RT_0][(1-P'V')+(P'-V')]. \tag{23a}$$

For a strong shock wave Eq. (28) assumes the form

$$\begin{aligned}T_s'={}&(1/2)\{(\gamma-1)/[3(\gamma-1)-\Sigma_n n]\}\\&\times(P_0V_0/RT_0)[(6+\Sigma_n n)V_s'-\Sigma_n n]P_s'.\end{aligned} \tag{29}$$

For the case of a perfect gas Eq. (29) gives the well-known result[4]

$$T_s'=[(\gamma-1)/(\gamma+1)]P_s', \tag{30}$$

since in this case $V'\approx(\gamma-1)/(\gamma+1)\rightarrow 1-V'$. For a solid (in which case $P_0\approx 0$), Eq. (29) assumes the form

$$\begin{aligned}T_s={}&(1/2R)\{(\gamma-1)/[3(\gamma-1)-\Sigma_n n]\}\\&\times[(6+\Sigma_n n)V_s-V_0\Sigma_n n]P_s.\end{aligned} \tag{31}$$

Finally, a different (kinematic) expression can be obtained for the excess temperature produced by shock compression of a solid (D and u are, respectively, the velocity of the shock wave and the mass velocity of the matter behind the shock wave; see next section):

$$\begin{aligned}&T_s-T_0=(1/\nu R\Sigma_n n)\{3D-[(6+\Sigma_n n)/2]u\}u,\\&\nu\equiv[3(\gamma-1)-\Sigma_n n]/(\gamma-1),\end{aligned} \tag{32}$$

which is similar to the thermodynamic expression (28). Switching in Eq. (32) to $\gamma=5/3$ and a Coulomb potential ($\Sigma_n=1$) yields

$$T_s-T_0=(1/R)[2D-(7/3)u]u. \tag{32a}$$

Using the Lennard–Jones potential[2] in Eq. (28)

$$U(|\mathbf{r}_j-\mathbf{r}_{j'}|)=A_1/r^{12}-A_2/r^6, \tag{33}$$

where A_1, $A_2>0$, $\Sigma_n n=18$, Eq. (28) becomes

$$\begin{aligned}T_s'={}&1+\{3(\gamma-1)/[(\gamma-1)-6]\}(P_0V_0/RT_0)\\&\times[(4/3)(P'V'-1)-(P'-V')]\\&\approx 1+[(\gamma-1)/(\gamma-7)](P_0V_0/RT_0)[4P'V'-3P']\\&\rightarrow T_s-T_0=[4(\gamma-1)/(\gamma-7)R]\\&\times[(V/V_0)-0.75]V_0P_s.\end{aligned} \tag{34}$$

This formula makes it possible to estimate the compression at which the Lennard–Jones potential comes into play — $(V/V_0)<0.75$, if $\gamma<7$. However, expression (28) also admits a negative difference $(T_s-T_0)<0$, if a certain relation holds between the parameters γ and $\Sigma_n n$. Although, due to its unusual nature, this effect is reminiscent of the situation of a rapidly stretched rubber band,[6] it should be used, most likely, as an additional condition on the actually possible form of the relation between the internal parameters $\Sigma_n n$ and γ (if the actual central two-particle potential $U(|\mathbf{r}_1-\mathbf{r}_2|)$ can indeed be represented precisely in the form $U(|\mathbf{r}_1-\mathbf{r}_2|)\equiv U(r)=\Sigma_n A_n/r^n$). Of course, this effect does not occur under shock compression of a perfect gas, where there is no parameter $\Sigma_n n$ associated with the presence of interaction in the system.

SHOCK ADIABAT AND THE EQUATION OF STATE OF A MATERIAL

At present, the construction of different theoretical models of the equation of state (mainly in the single-particle approximation) is a very complicated and internally not particularly consistent scheme: The total energy, internal or free, must be constructed from separate ''pieces,'' for which, moreover, different approximations are used in part. The only advantage of such a scheme is the formal thermodynamic consistency, since everything else is obtained only by the appropriate differentiation. Of course, the theoretically constructed equations of state must be compared with experiment. But experiments on gradual self-similar compression (they refer to temperatures of the order of tenths of an eV and pressures of the order of several Mbar; 1 eV = 11600 K; 1 bar = 10^6 dynes/cm^2 = 10^5 Pa ≈ 1.02 kgf/cm$^2 \equiv 1.02$ at ≈ 0.98 atm) do not yet cover the temperature and pressure ranges which are now required (tens of eV and tens of Mbar). For this reason, the main source of reference data are experimental results on shock compression of materials,[7,8] which, of course, it is difficult to treat as self-similar (rather, it is one-dimensional). Such compression will be close to self-similar if the shock pressure is so high that the contribution of shear stresses (stiffness contribution) becomes negligibly small. In this case, the degree of volume compression of the material is determined mainly by the Coulomb interaction of its electrons, while its hardness (the shear moduli) is determined by the Coulomb interaction of its nuclei. Everyone uses the points of the experimental SA as a test. For this, the theoretical SA is constructed from the expressions, found theoretically, for the internal energy and pressure by solving the SA Eq. (8) numerically (substituting the theoretically found expressions for $E=E(\rho,T)$ and $P=P(\rho,T)$ the equation for the SA assumes the form $F(\rho,T)=0$); the solution of the equation $F(\rho,T)=0$ is sought for T_i, given ρ_i. Then $P_i=P(\rho_i,T_i)$ is calculated according to the theoretically found equation of state $P=P(\rho,T)$. The point of intersection of the straight lines $\rho=\rho_i$ and $P=P_i$ gives the ith point of the theoretical SA. The points of the experimental SA are plotted on the SA constructed in this manner and the acceptability of the theoretically calculated $E=E(\rho,T)$ and $P=P(\rho,T)$ is judged. The equation of state itself can also be extracted directly from the experimental SA. Moreover, initially, the main goal of producing and measuring high pressures is to establish the actual equation of state of the material at high pressures, temperatures, and densities. This approach to using the experimental SA is described in detail in Ref. 4 and 5.

The experimental data on the shock compression of a material are usually presented in the form of individual points in the (ρ,P) or (σ,P) plane, where $\sigma\equiv\rho/\rho_0$, and these points are calculated using only the two relations

$$P=P_0+uD\rho_0;\quad \sigma=D/(D-u)$$

$$(\text{for a solid}\quad P=uD\rho_0), \tag{35}$$

following from the formulas[4,5,9]

$$(V_0-V)D^2=(P-P_0)V_0^2;$$

$$u=[(V_0-V)(P-P_0)]^{1/2}=(P-P_0)/D\rho_0, \tag{36}$$

which relate the kinematic quantities D and u of the process of shock compression of the material with its thermodynamic quantities in the same process. Here D is the shock-wave velocity in the as yet unperturbed material and u is the velocity at which the material is dragged the shock wave behind the shock front (the total specific energy of the entrained matter $=E+u^2/2$). In experiments on shock compression of matter, only these two kinematic quantities are measured: D directly and u indirectly. The equation $f(D,u)=0$ itself is called the kinematic SA of the material. In the general case, this relation is unknown, in contrast to its thermodynamic SA (8). Relation (8) is meaningful only when the function $E=E(V,P)$ is known, in which case the SA assumes the explicit form $P=P(V)$. Actually, however, the exact form of the function $E=E(V,P)$ is known only for a perfect gas, for which $E=VP/(\gamma-1)$. For this reason, the explicit form of the thermodynamic SA for perfect gases is[4,5,9]

$$P(V;T)=P_0(GV_0-V)/(GV-V_0);\quad G\equiv(\gamma+1). \tag{37}$$

For some reason the literature does not contain the equation of the kinematic SA even for a perfect gas, even though, as we have already mentioned, the function $E=E(V,P)$ for it is known exactly. This is especially strange, since experimenter almost always present their results graphically in the form of a kinematic SA, i.e., in the form $D=D(u)$, and not in the form of the thermodynamic SA $P=P(V)$. The kinematic SA, in contrast to its thermodynamic SA, cannot be expressed in an explicit form even for a perfect gas:

$$D^2-D(\gamma+1)u/2=\gamma P_0V_0. \tag{38}$$

This formula follows from Eqs. (35) and (8) (or from the SA (37) for a perfect gas), if the internal energy, pressure, and volume of the gas are expressed only in terms of kinematic quantities ($E=[P_0+uD/V_0](D-u)V_0/(\gamma-1)D$ only for a perfect gas; $P_0+P=2P_0+uD/V_0$, $V_0-V=uV_0/D$ — these relations hold for any solid) and substituted into Eq. (37).

In contrast to the thermodynamic SA (8), which is known for an arbitrary system, the kinematic SA cannot be obtained even in an implicit form, i.e., in the form $f(D,u)=0$, and even for a classical system. However, it will certainly be helpful to express it using relation (35), the virial theorem (12), and the SA (8), at least in the form $f(D,u,T)=0$ (of course, the same thing can also be done for the thermodynamic SA of a classical system). This expression is

$$(6/\Sigma_n n)D^2u+\{2\nu R(T-T_0)-[(6+\Sigma_n n)/\Sigma_n n]u^2\}D$$

$$-2P_0V_0[(3+\Sigma_n n)/\Sigma_n n]u=0,$$

$$\nu\equiv[3(\gamma-1)-\Sigma_n n]/(\gamma-1). \tag{39}$$

Now it is evident that for a solid $(P_0\approx 0)$ the kinematic SA (39) assumes an almost explicit form $D=D(u,T)$

$$D(u)=[(6+\Sigma_n n)/6]u-(1/3)\nu R(T-T_0)\Sigma_n n(1/u). \tag{40}$$

If the second term in Eq. (40) is small for some reason, then $D(u)$ will be a linear function of the velocity u. Such a dependence has indeed been found in experiments on shock compression of the solids, and it holds fairly well for many of them away from their phase transitions (PTs).

I shall now say a few words about why P_0, which is the ordinary atmospheric pressure, need not be taken into account for a solid. From the formula for the pressure of a perfect gas $P=(N/V)kT$ it is evident that the thermal pressure in condensed media should be greater than in a gas by approximately the same factor as the density N/V of their structural units is greater. But a gas develops a pressure of one atmosphere (atm) at room temperature ($T=300$ K) and with N/V of the order of 10^{19} cm^3 (i.e., of the order of the Loschmidt number 2.69×10^{19} cm^{-3}). In condensed media, however, the number of structural units (for example, atoms and molecules) is of the order of 10^{23} cm^{-3} (of the order of Avogadro's number $[(N/V)=N_A/\mu]$). For this reason, the thermal pressure here is approximately 10^4 times greater at the same temperature.

ON THE TWO-TERM REPRESENTATION OF THERMODYNAMIC QUANTITIES

In questions involving theoretical calculations of the equation of state of matter, the internal energy and pressure of the material are often separated into cold and warm parts[4,5] as done in Eq. (1a). This separation is most widely used in the theory of the thermodynamic SA. This is because, as already mentioned above (see Eq. (8)), this SA is known generally only in the implicit form $f(E,P,V)=0$ (see Eq. (8)), since the function $E=E(V,P)$, required for obtaining the explicit form $P=P(V)$ of the SA, is known exactly only for a perfect gas. For this reason, in the general case one still strives to express E, appearing in Eq. (8), in terms of P to obtain a relation of the type $E=E(V,P)$ but independent of $f(E,P,V)=0$, i.e., of Eq. (8), to close the latter relative to P and V only. For this, the Mie–Grüneisen relation $P_t=\Gamma(V)/E_t/V$ (see Eq. (5)) is usually used for lack of anything more suitable (the virial theorem, unfortunately, gives not $E=E(V,P)$ but rather $E=(V,P,E_k)$). It is then necessary to introduce a separation of the type (1a), which is entirely correct only in gradual processes and in shock relations. In using such separations, P is separated from E, appearing in Eq. (8), according to the scheme

$$\begin{aligned}E_s&=E_c^s+E_t^s=E_c^s+[V_s/\Gamma(V_s)]P_t^s\\&=E_c^s+[V_s/\Gamma(V_s)](P_s-P_c^s),\end{aligned}$$

and upon substituting this expression into Eq. (8) it is easy to obtain the SA in an explicit form, in which, however, three functions only of the volume will now appear: $E_c^s(V_s)$, $P_c^s(V_s)$, and $\Gamma(V_s)$. The first two functions should not present any difficulties. For them, the dependences on V and the relation between E_c^s and P_c^s, which hold in gradual thermodynamic processes, should be used: Since the internal energy and pressure of the system are functions of the state of the system, $E_c^s(V_s)=E_c^{\mathrm{fl}}(V_{\mathrm{fl}})$ if $V_s=V_{\mathrm{fl}}$ and similarly for P_c^s and P_c^{fl}, and therefore the relation between P_c^s and E_c^s should be the same as in the case of gradual processes, i.e., Eq. (1). In these arguments one is immediately alerted by the fact that the A_c^s, in contrast to A_c^{fl}, are not complete quantities, and therefore A_c^s and A_t^s need not be functions of the state of the system and the equation $A_c^s(V_s)=A_c^{\mathrm{fl}}(V_{\mathrm{fl}})$ with $V_s=V_{\mathrm{fl}}$ need not hold. Here only the complete quantity $A_s=A_c^s+A_t^s$ will be certainly a function of state. This is confirmed by an analytical investigation also. Indeed, for simplicity, consider the case of shock compression of a solid (then $P_0=0$), which is initially at $T=0$ (and therefore one can also set $E_0=0$). Then the SA (8) at first can be written as

$$2E(V,P)=(V_0-V)\cdot P, \tag{8a}$$

i.e., in this case the total specific energy $E_{\mathrm{tot}}=(V_0-V)\cdot P=E+E_{\mathrm{kin}}$ imparted to the material by the shock wave is distributed equally between its internal energy E and its entrainment energy $E_{\mathrm{kin}}=u^2/2$, which is very interesting in itself. Thus, after the substitution $A_s=A_c^s+A_t^s$ the SA (8a) itself becomes

$$2E_c^s(V_s)+2E_t^s(V_s,P_t^s)=(V_0-V_s)P_c^s+(V_0-V_s)P_t^s. \tag{8b}$$

Here V_0 is fixed and V_s assumes arbitrary values. But then, in order for Eq. (8b) to hold identically with respect to V_s two identities must be satisfied:

$$2E_x^s(V_s)=(V_0-V_s)P_c^s, \tag{41}$$

$$2E_t^s(V_s,P_t^s)=(V_0-V_s)P_t^s. \tag{42}$$

These relations are completely analogous to the complete relation (8a), but they do not admit a similar interpretation. It is much more important, however, that relation (41) differs from the cold isentrope (1a). This attests to the fact that the functions A_c^s and A_t^s are not functions of state of the system — they are exclusively artificial constructs. There is one other substantial difference in this question: Although state functions should not depend on the method by which the system reaches the final state (here fixed here by the volume V_s), initially they are determined strictly only for gradual (static) thermodynamic processes, where one talks about only diverse continuous paths of a transition from the initial into the final state, and in the process the system must pass through all intermediate states. In the case of shock compression, however, the system jumps from the initial into the final state, bypassing all of its intermediate states, and this jump can actually transfer the material from the SA of its initial phase to the SA of some other phase. This phase need not be the phase following in an order corresponding to gradual compression. The difference, analytically noted, between relations (41) and (1a) (quantitatively, this difference (expressed by the area of a triangle) is negligible, since only one curvilinear side of the triangle, being the zeroth isentrope, need be replaced by a straight line connecting the same initial and final points) could be attributable to the fact that in the case of gradual compression the independent variables in the functions $E_c(V)$ and $E_t(V,S)$ can be V and S, while in the case of shock compression V and S can no longer be independent; here only V (in the analytical coordinates V, $P(V)$) or only P (in the natural coordinates P, $V(P)$) can vary arbitrarily. For this reason, the separations of the type $A_s=A_c^s+A_t^s$ with a static relation between E_c^s and P_c^s and a static dependence of E_c^s on V_s are incorrect.

Using a separation of the type (1a), one can attempt to write all thermodynamic identities (they interrelate the thermodynamic parameters of materials in static processes performed on them) separately for the cold and warm parts. Thus, the identity

$$dE(V,S)=TdS-P(V,T)dV, \tag{43}$$

written using separations of the type (1a), assumes the form

$$dE_c(V)+dE_t(V,S)=-P_c(V)dV+TdS-P_t(V,T)dV. \tag{44}$$

Since relation (44) remains an identity, it should split into two relations:

$$dE_c(V)=-P_c(V)dV,$$

$$dE_t(V,S)=TdS-P_t(V,T)dV. \tag{44a}$$

Other identities for the free energy, the enthalpy, and the Gibbs potential can be separated in exactly the same manner, while the identity (4) will assume the form[3]

$$dE_c(V)/dV=-P_c(V),$$

$$T(\partial P_t/\partial T)_\nu=(\partial E_t/\partial V)_T+P_t(V,T). \tag{4b}$$

It is now evident from expressions (44a) and (4b) that in gradual processes both the individual parts of the internal energy and the pressure can be state functions of the system, i.e., for example, the Mie–Grüneisen relation (5) is not meaningless in static thermodynamic processes.

APPENDIX [3]

1. *Virial theorem.* A relation between the pressure, the kinetic and potential parts of the internal energy, and the volume of a system is usually referred to as the virial theorem.[1,10] A derivation of this relation for the particular case of a classical system with an interaction potential between the structural particles in the form of functions of their coordinates with a fixed degree of homogeneity is contained in Refs. 1 and 10. In Ref. 3 a simpler, quantum-mechanical derivation of this relation is given on the basis of the variational principle of quantum mechanics.[11] Moreover, this derivation also holds for the case, which is sometimes forced, that the interaction potential energy of the structural particles of the material is expressed in the form of a function of the relative coordinates with different degrees of homogeneity $U(r)\approx\Sigma_n A_n/r^n\equiv\Sigma_n U_n(r)$. Since Ref. 3 is not readily accessible[3] and to achieve complete the presentation, it probably makes sense to present here a brief derivation of the expression for the virial theorem. It is based on the variational principle and perturbation theory in the form of virtual work: If the linear dimensions of the volume of a body change isotropically under infinitesimal uniform and hydrostatic compression $L\rightarrow L'=(1-\varepsilon)L$, then this compression will lead to the following change in the potential: $U_n(r)\equiv A_n/r^n=(1-n\varepsilon)A_n/r'^n=(1-n\varepsilon)U'_n$, where the prime here and below marks parameters of the compressed material. Now, the variational principle[11] must be taken into account, and the fully primed equation $H'|m'\rangle=E'_m|m'\rangle$ must be replaced by a partially primed equation $H'|m\rangle=E'_m|m\rangle$, which introduces an error, proportional to ε^2, in the quantity E'_m-E_m. For this reason, we can write

$$-P\equiv\langle\partial H/\partial V\rangle=\lim(\nu'\rightarrow\nu)\langle(H'-H)/(V'-V)\rangle. \tag{A1}$$

If $H-H_{\mathrm{kin}}\equiv\Sigma_n U_n(r)$, then $(H-H_{\mathrm{kin}})'=\Sigma_n(1+n\varepsilon)U_n(r)$, and $(H-H_{\mathrm{kin}})'-(H-H_{\mathrm{kin}})=\varepsilon\Sigma_n nU_n(r)$; $V'-V=-3\varepsilon V$. Now relation (A1) gives

$$3VP_{\mathrm{pot}}(V,T)=\Sigma_n nE_n^{\mathrm{pot}}(V,T). \tag{A2}$$

The average $\langle\ \ldots\ \rangle$ in Eq. (A1) can be interpreted as a static average over a Gibbs distribution or a quantum-mechanical average over the ground state of the system. In the case that there is no interaction between the structural particles of the system, classical or quantum, the virial theorem of the system must evidently have the form

$$3VP_{\mathrm{kin}}(V,T)=3(\gamma-1)E_{\mathrm{kin}}(V,T), \tag{A3}$$

where the thermodynamic parameter γ is always the exponent in the Poisson equation $PV^\gamma=$ const for a given ideal system and simultaneously $\gamma=C_p/C_\nu$ only in the case of a classical ideal (but necessarily monatomic!) gas.

For this reason, the most general form of the virial theorem is

$$\begin{aligned}3VP(V,T)&=3(\gamma-1)E_k(V,T)+\Sigma_n nE_n^p(V,T)\\&=3(\gamma-1)E_k+[\Sigma_n nE_n^p/\Sigma_n n]\Sigma_n n\\&\approx3(\gamma-1)E_{\mathrm{kin}}(V,T)+E_{\mathrm{pot}}(V,T)\Sigma_n n.\end{aligned} \tag{A4}$$

If the material is treated as a collection of interacting electrons and nuclei (then one must set $\gamma=5/3$ and $\Sigma_n n\rightarrow1$), Eq. (A4) assumes its standard form[10]

$$3VP(V,T)=2E_{\mathrm{kin}}(V,T)+E_{\mathrm{pot}}(V,T). \tag{A5}$$

It now remains to make two remarks. First, contrary to the often encountered assertions that the virial theorem (A5) can be used to calculate the pressure directly, it is not suitable for this purpose even when E_{kin} and E_{pot} are known, since the use of this equation always results in the calculation of a small difference of large quantities, which the sum $2E_{\mathrm{kin}}+E_{\mathrm{pot}}$ is. The same is also true of the virial theorem in the form (A4). Second, since the specific internal energy and the equation of state of a perfect gas have the form[10,12]

$$E_{\mathrm{kin}}(T)=i(RT)/2\mu;\quad P_{\mathrm{kin}}(V,T)=RT/\mu V, \tag{A6}$$

where i is the number of degrees of freedom of a structural particle of this system, we have

$$C_\nu\equiv(\partial E/\partial T)_\nu=iR/2\mu;$$

$$C_p\equiv C_\nu+P(dV/dT)_p=(i+2)R/2\mu.$$

Then $\gamma\equiv C_p/C_\nu=(2+i)/i$, and since i can assume only integer values $i=0,1,2,3,\ \ldots,\infty$, γ can assume the values $\gamma=2/0, 3/1, 4/2, 5/3, 6/4,\ \ldots\ , 1$. This leads to a unique ''quantization'' of the maximum compression of such a gas by a single shock wave:

$$\rho_{\mathrm{lim}} \equiv [(\gamma+1)(\gamma-1)]\rho_0 = (1+i)\rho_0$$
$$= \rho_0; 2\rho_0; 3\rho_0; 4\rho_0; 5\rho_0; \ldots; \infty. \qquad \text{(A7)}$$

2. *Cold pressure of a simple substance.*[3,13] Dimensional considerations can also be used to calculate the cold energy of a substance. Consider an electrically neutral (on the average) cell of a cold simple substance. The total energy of the cell can be represented as a sum of the kinetic and Coulomb energies. The dependences $E_c^{\mathrm{kin}}(n)$ and $E_c^{\mathrm{coul}}(n)$ of the densities of these parts of the energy of a cell on the electron density in the cell (the Coulomb part contains the electron-electron interaction energy and the interaction of electrons in the cell with the nucleus of the cell) can be easily found from dimensional analysis of the operators $\hbar^2\Delta/2m$ and e^2/r; they should be proportional to $n^{5/3}$ and $n^{4/3}$, respectively, because $E_c^{\mathrm{kin}}(n)$ can depend only on the three parameters $\hbar$, m, and n, i.e., $E_c^{\mathrm{kin}}/|\omega| \sim \hbar^s m^t n^p$. But the energy density has the dimensions $[E/V] = ML^{-1}T^{-2}$. Then $[\hbar^2 m^t n^p] \equiv (ML^2T^{-1})^s M^t (L^{-3})^p = ML^{-1}T^{-2} \rightarrow s=2$, $t=-1$, and $p=5/3$. The Coulomb energy density can depend only on two parameters: e and n. Then one obtains analogously $E_c^{\mathrm{coul}}/|\omega| \sim e^2 n^{4/3}$. Thus the energy of a cold cell (it is proportional to the cell volume $|\omega|$) can be written in the form of a function of its electron density n as

$$E_c^{\mathrm{cell}}(n) = A_1 n^{2/3} + A_2 n^{1/3}. \qquad \text{(A8)}$$

The specific energy of this substance can be obtained hence by multiplying by N_A/A — the number of atoms per unit mass of the substance (A is the mass number of the substance). Since the average volume per atom of a simple substance with density ρ is $V_{av} = A/\rho N_A$, the average electron density in the cell will be $n_{av} \equiv z^*/V_{av}$, where z^* is a certain effective number of free electrons per atom of the substance; of course, this number depends on the density of the substance, but between PTs it can be regarded as fixed. Hence one obtains the relation $n_{av} = (z^*/A)\rho N_A$ and the formula for the specific energy is

$$E_c(\rho) = [A_1 \varkappa^{1/3}\rho^{1/3} + A_2]\varkappa^{1/3}\rho^{1/3}, \quad \varkappa \equiv z^* N_A/A. \qquad \text{(A9)}$$

The derivative of this expression gives the pressure in the cold substance

$$P_c(\rho) \equiv -\partial E_c/\partial V = \rho^2 \partial E_c/\partial\rho$$
$$= \rho^{5/3}\varkappa^{1/3}[2A_1\varkappa^{1/3} + A_2\rho^{-1/3}]. \qquad \text{(A10)}$$

The equilibrium density of the material (for a given phase) can be determined from the condition that the pressure equals zero in the substance in this state:

$$P_c(\rho_0) = 0 \rightarrow 2A_1\varkappa^{1/3} + A_2\rho_0^{-1/3} = 0.$$

Hence A_2 can be expressed in terms of A_1 and ρ_0: $A_2 = -2A_1(\varkappa\rho_0)^{1/3}$. Then the specific energy of the cold substance and the pressure in it assume the form

$$E_c = 2A_1\varkappa^{2/3}\rho^{1/3}[(\rho^{1/3}/2) - \rho_0^{1/3}]$$
$$= 2A_1\varkappa^{2/3}\sigma^{1/3}[(\sigma^{1/3}/2) - 1]; \quad \sigma \equiv \rho/\rho_0;$$

$$P_c = (2A_1/3)\varkappa^{2/3}\rho^{4/3}[\rho^{1/3} - \rho_0^{1/3}]$$
$$= (2A_1/3)\varkappa^{2/3}\rho_0^{5/3}\sigma^{4/3}[\sigma^{1/3} - 1].$$

The remaining free parameter A_1 can likewise be expressed in terms of the characteristics of the substance in its equilibrium (for a given phase) state, keeping in mind the expressions for the square of the isothermal sound velocity and isothermal bulk modulus in the substance (Ref. 4, p. 554)

$$C^2 = -V^2 dP_c/dV = dP_c/d\rho = dP_c/\rho_0 d\sigma,$$

$$K \equiv -V dP_c/dV = \rho dP_c/d\rho = \sigma dP_c/d\sigma.$$

Simple calculations using the expression for P_c give for the sound velocity

$$C^2 = (2A_1/9)\varkappa^{2/3}\rho^{1/3}[5\rho^{1/3} - 4\rho_0^{1/3}] = C_0^2[5\sigma^{2/3} - 4\sigma^{1/3}],$$

$$C_0^2 \equiv (2A_1/9)(\varkappa\rho_0)^{2/3}.$$

Now the expressions for the parameters A_1 and A_2 are

$$A_1 = (9C_0^2/2)/(\varkappa\rho_0)^{2/3}; \quad A_2 = -9C_0^2/(\varkappa\rho_0)^{1/3}.$$

The expression for the bulk modulus is

$$K = (2A_1/9)\varkappa^{2/3}\rho^{4/3}[5\rho^{1/3} - 4\rho_0^{1/3}] = K_0[5\sigma^{5/3} - 4\sigma^{4/3}],$$

and therefore there is a simple relation $K = (2A/9)(\varkappa\rho)^{2/3}\rho = C^2\rho$. Now the expressions for E_c and P_c assume the very simple form

$$E_c = 9K_0\rho_0^{-1}[(\sigma^{2/3}/2) - \sigma^{1/3}],$$

$$P_c = 3K_0[\sigma^{5/3} - \sigma^{4/3}]. \qquad \text{(A11)}$$

These formulas describe the relations existing only within a single phase of the substance, for which ρ_0 and K_0 are the equilibrium values.

[1] R. P. Feynman, *Statistical Mechanics*, Benjamin, Reading, Mass., 1972 [Mir, Moscow, 1975].

[2] R. Kubo, *Statistical Mechanics* (North-Holland, Amsterdam, 1965; Mir, Moscow, 1967).

[3] M. F Sarry, Vopr. At. Nauk Tekh. Ser. TPF (Kharkov), No. 3, 10 (1992).

[4] Ya. B. Zel'dovich and Yu. P. Raĭzer, *Elements of Gas Dynamics and the Classical Theory of Shock Waves*, (Academic Press, New York, 1968; Russian original, Nauka, Moscow, 1966), pp. 10–13.

[5] V. N. Zharkov and B. A. Kalinin, *Equations of State of Solids at High Pressures and Temperatures* [in Russian], Nauka, Moscow (1968).

[6] R. Kubo, *Thermodynamics* (North-Holland, Amsterdam, 1968; Mir, Moscow, 1970).

[7] L. V. Al'tshuler, Usp. Fiz. Nauk **85**, 197 (1965) [Sov. Phys. Usp. **8**, 52 (1965)].

[8] L. V. Al'tshuler and A. A. Bakanova, Usp. Fiz. Nauk **96**, 193 (1968) [Sov. Phys. Usp. **11**, 678 (1968)].

[9] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon Press, New York, 1987; cited Russian original, Nauka, Moscow, 1988).

[10] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. rev. and enl., with L. P. Pitaevskiĭ (Pergamon Press, New York, 1980; Russian original, Nauka, Moscow, 1976).

[11] F. Seitz, *Modern Theory of Solids* (McGraw-Hill, New York, 1940; GITTL, Moscow, 1949).

[12] I. K. Kikoin and A. K. Kikoin, *Molecular Physics* [in Russian], Fizmatgiz, Moscow, 1963.

[13] B. A. Nadykto, Usp. Fiz. Nauk **163**, 37 (1993).

Translated by M. E. Alferieff

Vibrational pumping of H_2 molecules in a hydrogen stream flowing through a cesium–hydrogen discharge

F. G. Baksht and V. G. Ivanov

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

(Submitted June 2, 1997)
Zh. Tekh. Fiz. **68**, 10–19 (October 1998)

The H_2 molecular vibrational distribution function that is formed in a flow of molecular hydrogen is calculated. It is assumed that the hydrogen stream flows in a planar channel and passes through two sections in succession. In the first section a cesium–hydrogen discharge and preliminary vibrational pumping of H_2 molecules occur. In the second section, where there is no discharge, the $v-v$ exchange process results in a considerable increase in the density of vibrationally excited molecules in a certain upper part of the vibrational spectrum. The possibility of using the vibrational distribution function produced in this manner to generate negative hydrogen ions as a result of subsequent dissociative attachment of electrons to vibrationally excited molecules is discussed. © *1998 American Institute of Physics.* [S1063-7842(98)00210-4]

1. The generation of vibrationally excited hydrogen molecules is now of very considerable interest in connection with a variety of applications (ion sources, plasma chemistry, and others). As a rule, such molecules are generated in different types of discharges containing either a pure hydrogen plasma or a hydrogen plasma with an easily ionized additive, specifically, cesium. Depending on the specific applications, different schemes are used to produce vibrational excitation of H_2. For example, in existing bulk sources of negative hydrogen ions, where H^- ions are generated in the process of dissociative attachment (DA) of thermal electrons to vibrationally excited H_2 molecules in the ground electronic state $X^1\Sigma_g^+$ (Ref. 1) vibrational excitation is often accomplished by a cascade method using radiative deexcitation of singlet electronically excited states $B^1\Sigma_u^+$, $C^1\Pi_u^*$, and others.[2] In H^- sources employing such a two-step scheme for generating vibrationally excited molecules $H_2(X^1\Sigma_g^+)$, the vibrational excitation and dissociative attachment processes are, as a rule, separated in space.[3–5] The separation is accomplished either by organizing a separate discharge chamber, where excited H_2 molecules are formed (two-chamber source), or by producing a so-called hybrid source,[3,4] where the separation occurs in a single discharge chamber where fast cathodic electrons which excite H_2 molecules are held near the chamber walls as a result of drift in crossed **E** and **H** fields. Here the H^- ions are generated in a dissociative attachment process at the center of the chamber, where there are no fast or hot thermal electrons which destroy H^- ions and the electron temperature $T_e \sim 1$ eV, i.e., the optimal value for dissociative attachment processes.[6] Among a number of works devoted to the theory of two-chamber and hybrid sources, we call attention to Ref. 4, where H^- and D^- sources are optimized with respect to the discharge current and the electron temperature, and the theory is compared with experimental data.

We note that spatial separation of H_2 vibrational excitation and H^- generation processes could be very desirable for another reason, on which we shall dwell here in connection with the possible use of a low-voltage (LV) cesium–hydrogen discharge for generating vibrationally excited H_2 molecules.[7] Since such a discharge can be produced with a comparatively high density of both electrons and H_2 molecules (see, for example, Refs. 8 and 9), vibrational excitation processes can occur in it intensively directly in the ground electronic state $X^1\Sigma_g^+$ in both $e-v$ processes and $v-v$ exchange processes. At the same time, in a number of applications, specifically, in H^- sources for controlled thermonuclear fusion, it is desirable to have a comparatively low hydrogen pressure in the region from which H^- ions are extracted. For this reason, it is also advantageous to separate in space the generation of vibrationally excited molecules (the first, relatively high-pressure chamber) and the formation and subsequent extraction of H^- ions (the second, low-pressure chamber). The required pressure differential between the chambers can be produced by efflux of vibrationally pumped hydrogen at the sound velocity V_s from the first chamber into the second chamber.

In the present paper it will be shown that if the flow is properly organized, the vibrational distribution function (VDF) of H_2 molecules flowing out of the first chamber can be appreciably improved, specifically, in the range of vibrational numbers v that is important for dissociative attachment processes. Preliminary results have been published in Ref. 10.

2. The flow model considered here is illustrated in Fig. 1a. The molecular hydrogen stream flows in a channel formed by two parallel walls and passes successively through sections I and II (of length h_1 and h_2, respectively). A low-voltage cesium–hydrogen discharge between two planar electrodes is produced in section I. The initial vibrational pumping of the hydrogen occurs here. Next, the hydrogen flows into the section II, which is insulated from section I. Hydrogen flows out from the end of the channel ($x = h_2$) of section II at the sound velocity V_s. The walls of section II

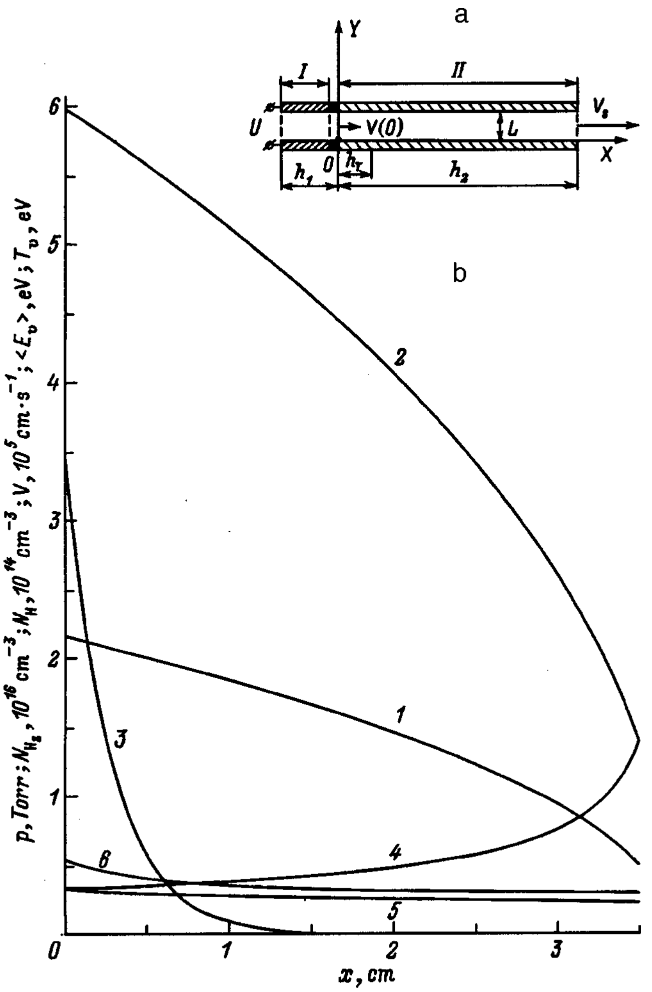

FIG. 1. a — Diagram of hydrogen flow in a channel: I — Discharge zone, II — flow of vibrationally pumped gas in the channel; b — distribution of the flow parameters along the channel: *1* — p, *2* — N_{H_2}, *3* — N_{H}, *4* — V, *5* — $\langle E_v\rangle$, *6* — $T_v=(E_1/k)/\ln(N_0/N_1)$. Discharge parameters: $L=0.3$ cm, $h_1>1$ cm, $N^{(0)}_{\mathrm{H}_2}=3\times10^{16}$ cm^{-3}, $N^{(0)}_{\mathrm{H}}=1.62\times10^{14}$ cm^{-3}, $T_e=0.88$ eV, $N^{(0)}_{\mathrm{Cs}}=7\times10^{13}$ cm^{-3}, $n_e=2.07\times10^{13}$ cm^{-3}, $N^{(0)}_{\mathrm{H}^-}=5.1\times10^{12}$ cm^{-3}, $\langle E_v\rangle=0.366$ eV, $T_v=0.59$ eV. Gas temperature in the discharge $T_0=0.08$ and in the channel $T=0.04$ eV.

are maintained at close to room temperature. Intensification of the vibrational pumping of H_2 molecules in a definite, upper part of the vibrational spectrum occurs in a nonresonant $v-v$ exchange process in cold gas in this section.

The variation of the pressure $p(x)$ and density $N_{\mathrm{H}_2}(x)$ of molecular hydrogen along the channel in section II can be determined approximately by analogy to a viscous gas flow in a circular pipe (cf. Ref. 11). For a sufficiently long channel ($h_2/L\gg1$), on averaging the gas-dynamic velocity (Ref. 12, p. 81) over the transverse cross section, we obtain the following expression for the average flow velocity V in a planar channel:

$$V=-\frac{L^2}{12\eta}\frac{dp}{dx}, \tag{1}$$

where η is the viscosity of molecular hydrogen.

Assuming that $V=V_s$ at the channel cutoff, we obtain the molecular hydrogen pressure distribution along the channel

$$p(x)=[p_0^2-(p_0^2-p_s^2)x/h_2]^{1/2}. \tag{2}$$

Here $p_0=p(0)$ is the hydrogen pressure at the channel entrance, i.e., approximately the pressure in the discharge, and p_s is the pressure at the cutoff. The length h_2 and width L of the channel are related by the relation

$$h_2/L=(\mathrm{R}_s/24)(c_p/c_v)^{-1}[(p_0/p_s)^2-1], \tag{3}$$

where $\mathrm{R}_s=\rho_s L V_s/\eta$ is the Richardson number, calculated with respect to the transverse section L of the channel and the flow parameters at the cutoff; $\rho_s=M_{\mathrm{H}_2}p_s/kT$ is the hydrogen density at $x=h_2$; and, T is the gas temperature in the channel.

The parameters of the plasma in the initial low-voltage discharge (section I) and the gas parameters in the channel (section II) were assumed to be uniform over the cross section. As an example, Fig. 1b shows the distribution of the main flow parameters along the channel in section II: the pressure p, the molecular hydrogen density N_{H_2}, and the flow velocity V, as well as the average vibrational energy $\langle E_v\rangle$ and vibrational temperature T_v of the molecules, defined in terms of the populations N_0 and N_1 of the ground and first excited vibrational levels, and the density N_{H} of atomic hydrogen (see below).

3. Let us now discuss the main aspects of the calculation of the component composition of the plasma and the VDF in the initial low-voltage discharge. The parameters of the plasma in a low-voltage cesium–hydrogen discharge were calculated by the method of Ref. 13, which was developed for a uniform gas-discharge gap. The plasma studied was a quite dense discharge plasma, in which the cathode beam relaxes in the plasma on electron–electron pair collisions and the beam energy is expended on heating of the thermal electrons.[14] The calculation was performed in two steps.[9] First, the system of equations determining the electronic-vibrational kinetics in the discharge was solved at a fixed electron temperature T_e. All plasma parameters, including the VDF $f_v^{(0)}$ of H_2 molecules and the ion fluxes j_1 extracted from the plasma on each electrode, were found from the solution of this system. Second, the system of boundary conditions which describes the balance of charged particles and the energy of the electrons in the near-cathode and near-anode boundaries of the plasma was solved with the known plasma parameters. The near-cathode φ_1 and near-anode φ_2 potential barriers in the Langmuir sheaths and the electron fluxes j_{e1} and j_{e2} extracted from the plasma to the cathode and anode were determined from the solution of this system. As a result, the voltage $U=\varphi_1-\varphi_2$ and the discharge current $j=j_{e2}-j_i$, which correspond to the set of plasma parameters determined at the first stage, were found. We call attention to an important circumstance, which distinguishes the burning regimes investigated in the present work, those of a low-voltage cesium–hydrogen discharge in a dense, collisional plasma, from similar regimes which were previously investigated theoretically in Ref. 13, in particular. The burn-

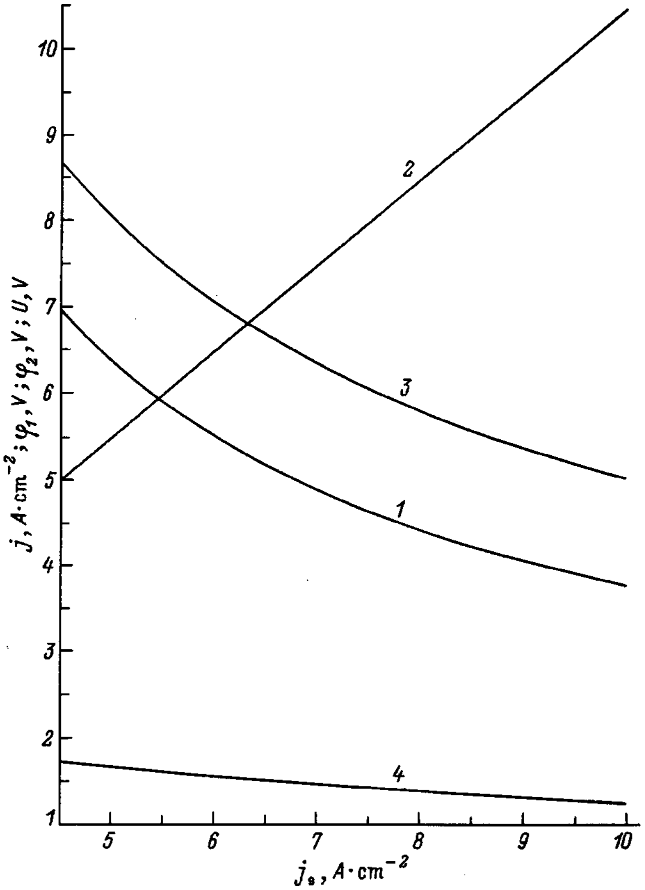

FIG. 2. Discharge voltage and current and near-electrode potential barriers in the Langmuir sheaths versus the cathode emission current: *1* — U, *2* — j, *3* — φ_1, *4* — φ_2. Discharge parameters: $L=0.3$ cm, $h_1>1$ cm, $N_{\mathrm{H}_2}^{(0)}=3\times 10^{16}$ cm^{-3}, $N_{\mathrm{H}}^{(0)}=1.8\times 10^{14}$ cm^{-3}, $T_e=0.65$ eV, $N_{\mathrm{Cs}}^{(0)}=10^{14}$ cm^{-3}, $n_e=3.26\times 10^{13}$ cm^{-3}, $N_{\mathrm{H}^-}^{(0)}=3.8\times 10^{12}$ cm^{-3}, $\langle E_v\rangle=0.335$ eV, $T_v=0.55$ eV.

ing regimes of a low-voltage Cs–H$_2$ discharge that were studied in the present work are characterized by comparatively low emission currents $j_s<10$ A/cm^2, which makes it much easier to produce the regimes in practice. This is illustrated in Fig. 2, which shows φ_1, φ_2, U, and j as functions of j_s. One can see that a typical low-voltage regime, where at the channel cutoff $\varphi_1<E_d/q$ (E_d is the threshold for direct dissociation of H$_2$ by electron impact), is produced for moderate values of the cathode emission current $5<j_s<10$ A/cm^2. The required vibrational pumping of H$_2$ molecules to high vibrational levels is now achieved not only in the discharge but also as a result of the subsequent increase in the density N_v of vibrationally excited molecules due to an appropriate organization of the flow in the channel.

As a result of the quite low flow velocity V_0 of the gas in the discharge zone, the low-voltage discharge was calculated by a method developed for a gas-discharge gap with a plasma at rest. At the same time, there are two circumstances, specific to a discharge in a moving plasma, that must be analyzed. The first one is the relatively long vibrational relaxation time of molecules, which in a number of cases leads to comparatively slow formation of the VDF $f_v^{(0)}$ of H$_2$ molecules in the discharge. To take this circumstance into account a calculation of the establishment of the VDF of the molecules was carried out for the typical working plasma parameters of our cesium–hydrogen discharge. The computational method is presented in Ref. 15. In addition to Ref. 16, in the present calculations the time dependence of the density N_{H} of hydrogen atoms and negative hydrogen ions N_{H^-}, which are established comparatively slowly, was taken into account in the vibrational relaxation process. Figure 3 displays the computational results in the form of ratios of the instantaneous VDF $f_v(t)$ to its value established in the discharge $f_v^{(0)}\equiv f_v(\infty)$. The calculations were performed for two values of the translational temperature T_0 of the gas in the discharge. One can see that the character of the vibrational relaxation depends on T_0 (cf. Figs. 3a and 3b). At higher temperatures T_0, characteristic maxima appear in the dependences $f_v(t)/f_v(\infty)$. These peaks show that the relaxation of the populations $N_v(t)$ of the upper levels to their steady-state values $N_v^{(0)}\equiv N_v(\infty)$ is nonmonotonic. The nonmonotonic behavior is due to $v-t$ exchange, which at high gas temperatures T_0 is more pronounced and starts to influence the populations of the upper vibrational levels substantially, decreasing them appreciably at the end of the vibrational relaxation process. We note that for this reason at a higher gas temperature ($T_0=0.08$ eV) the steady-state populations $N_v^{(0)}$ of the upper vibrational levels and, together with them, the atomic densities $N_{\mathrm{H}}^{(0)}$ are appreciably lower than at a low temperature ($T_0=0.026$ eV). The total vibrational relaxation time in both examples is $\tau_v^{(0)}\simeq 15$ μs, and the corresponding total vibrational relaxation length is $l_v^{(0)}=V_0\tau_v^{(0)}\simeq 1$ cm.

Second, the specific nature of a discharge in a moving plasma is manifested as a quite slow establishment of the translational temperature T_0 of the gas in the discharge. This is an important, since, as indicated above, the translational temperature T_0 appreciably influences the vibrational relaxation process and the character of the steady-state vibrational distribution in the discharge. Under the conditions considered, the heating of the gas in the discharge gap is mainly due not to heat release in the discharge plasma (in elastic collisions of electrons with heavy particles or in $v-t$ and nonresonant $v-v$ exchange processes) but rather to contact of the gas with the hot emitter, i.e., the formation of an expanding thermal boundary layer near the emitter. Estimating the emitter length h_T for which the thickness δ_T of the thermal boundary layer equals the width L of the interelectrode gap,[16] we obtain $h_T=\theta\cdot\mathrm{Pr}\cdot\mathrm{R}_0 L$, where $\mathrm{Pr}=\eta_0 c_p/\kappa_0$ is the Prandtl number, $\mathrm{R}_0=\rho_0 L V_0/\eta_0$ is the Reynolds number calculated from the flow parameters in the discharge zone (κ_0, η_0, and ρ_0 are the thermal conductivity, viscosity, and density of hydrogen in the gas-discharge gap, and c_p is the specific heat at constant pressure), and $\Theta\sim 0.1$ is a numerical factor. As a result, for typical discharge and flow parameters considered here, the gas temperature T_0 characteristic for a stationary discharge is established in the discharge zone at a distance $h_T\simeq 0.2$ cm, i.e., more rapidly than the VDF. The remaining times and corresponding characteristic lengths (Maxwellization of the electrons, ionization of Cs atoms,[17] and others) are appreciably shorter than the times considered above. Therefore, in order for there to be

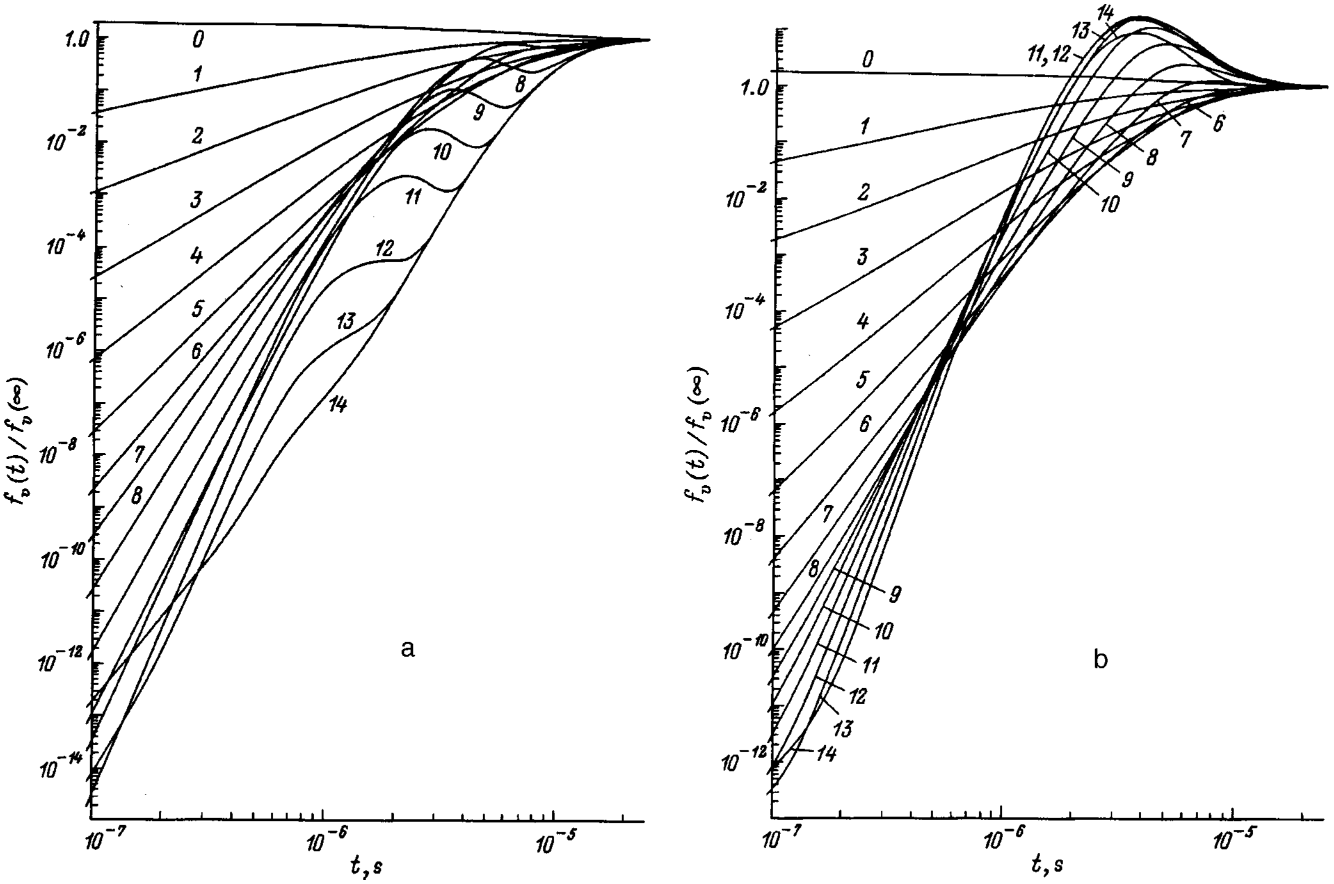

FIG. 3. Vibrational relaxation of H_2 molecules in the plasma of a low-voltage discharge (the numbers of the vibrational levels are indicated on the curves): $N^{(0)}_{H_2}=3\times10^{16}$ cm^{-3}, $N^{(0)}_{Cs}=10^{14}$ cm^{-3}, $T_e=0.8$ eV, $n_e=5.8\times10^{13}$ (a), 5.4×10^{13} cm^{-3} (b); $N^{(0)}_{H}=7.8\times10^{14}$ (a), 3.3×10^{14} cm^{-3} (b); $N^{(0)}_{H^-}=6.6\times10^{12}$ (a), 7×10^{12} cm^{-3} (b); $T_0=0.026$ (a), 0.08 eV (b).

enough time for the plasma to relax completely in the discharge zone to a state characteristic for a stationary low-voltage discharge, under the conditions considered the length of the discharge zone must satisfy $h_1>1$ cm. This condition is assumed to hold below.

For what follows, it is important to set the temperature T of the gas in the channel, i.e., in section II. Since the probabilities for the highly excited vibrational levels to be populated by nonresonant $v-v$ exchange increase considerably (compared with the corresponding vibrational deexcitation probabilities) as the translational temperature T decreases, it is desirable for the gas temperature in the channel to be quite low. In specific calculations, we shall assume that the channel walls are at room temperature, while the gas temperature T in the channel is somewhat higher than the room temperature; this approximately takes into account some heating of the gas in the channel by means of $v-t$ and nonresonant $v-v$ exchange. Most of the calculations below are for $T=0.03$–0.04 eV. Since the relaxation length h_T of the gas temperature is very small compared with the channel length h_2, we shall neglect the change in pressure p over the distance $\sim h_T$ (Fig. 1b) and we shall introduce a jump in the gas temperature $\Delta T=T_0-T$ at the boundary ($x=0$) of the gas-discharge gap (section I) with the channel (section II). For continuous pressures p and neutrals fluxes at the boundary, this results in corresponding jumps in the densities N_{H_2} and N_H and flow velocity V at the boundary. The vibrational distribution function f_v and the degree of dissociation of hydrogen remain continuous. The values of N_{H_2}, N_H, and V in the discharge zone, i.e., before the jump, are denoted by $N^{(0)}_{H_2}$, $N^{(0)}_{H}$, and V_0. The corresponding quantities after the jump in temperature are denoted by $N_{H_2}(0)$, $N_H(0)$, and $V(0)$.

4. The vibrational distribution function of H_2 molecules in the channel of section II was determined by solving the system of equations

$$\frac{d}{dx}(N_vV)=I^{(vv)}_v\{N_v\}+I^{(vt)}_{vM}\{N_v\}+I^{(vt)}_{vA}\{N_v\}+I^{(w)}_v\{N_v\}$$

$$(v=0,1,2,\ldots,14) \qquad (4)$$

where $N_v(x)$ is the density of vibrationally excited molecules in the level v in the channel.

The terms on the right-hand side of Eq. (4) take into account successively $v-v$ and $v-t$ exchange with hydrogen molecules and atoms and vibrational relaxation of molecules on the channel walls. The corresponding expressions are formulated in the same way for both the discharge section I and channel II. The terms describing $v-v$ and $v-t$ exchange are expressed in the same matter as in Refs. 8 and 13.

We shall discuss separately the terms $I_v^{(w)}\{N_v\}$, describing the interaction of molecules with the channel walls. In contradistinction to the discharge section I, in channel II, where there is no $e-v$ exchange, the specific form of the terms $I_v^{(w)}$ is very important, since it strongly influences the VDF $f_v(x)$ formed there. It is convenient to describe the losses of vibrationally excited molecules at the walls by introducing a corresponding effective lifetime τ_v, so that $I_v^{(w)} \times\{N_v\}=-N_v/\tau_v$. For a gas at rest in a gap with parallel flat walls, the expression for τ_v can be written by analogy to Ref. 18 in the form

$$\tau_v=\frac{\Lambda^2}{D_{\text{sd}}}+\frac{L}{\bar{v}}\frac{2-\gamma_v}{\gamma_v}. \tag{5}$$

Here $\Lambda=L/\pi$ is the effective diffusion length for a planar geometry; $D_{\text{sd}}=(3\pi/16\sqrt{2})\bar{v}l_{\text{H}_2}$ is the self-diffusion coefficient of H_2 molecules;[19] $\bar{v}=\sqrt{8kT/\pi M_{\text{H}_2}}$; $l_{\text{H}_2}=1/N_{\text{H}_2}\overline{\sigma}^{(1)}$ is the mean free path of H_2 molecules; and, γ_v is the probability that a H_2 molecule excited to the level v vanishes at the wall. The first term on the right-hand side of Eq. (5) is much greater than the second term if $\gamma_v\gg(3\pi^2/8\sqrt{2})(l_{\text{H}_2}/\Lambda)$. In this case $\tau_v=\tau_d\equiv\Lambda^2/D_{\text{sd}}$, where τ_d is the time for molecules to diffuse to the wall, which does not depend on γ_v. Since $l_{\text{H}_2}/\Lambda\ll1$, this limit in any case obtains for $\gamma_v\sim1$. In the opposite limiting case of very small γ_v ($\gamma_v \ll(3\pi^2/8\sqrt{2})(l_{\text{H}_2}/\Lambda)$), $\tau_v=2L/(\gamma_v\bar{v})$ and the lifetime is much greater than the diffusion time τ_d. It is shown in the Appendix that when certain conditions are satisfied an expression of the type $I_v^{(w)}\{N_v\}=-N_v/\tau_v$, where τ_v is determined from Eq. (5), correctly describes the loss of particles (molecules, atoms, and so on) from the gas flow at the wall. Since the second term in Eq. (5) in the limit of very small γ_v is quite obvious and its use for a gas at rest or a moving gas requires no justification, the analysis is made in the Appendix for the example of the calculation of particle losses from the flow in the opposite, diffusion limit, for which $\gamma_v\sim1$ and $\tau=\tau_d$. The specific calculations were performed for the losses of atomic hydrogen diffusing from the flow to the channel wall in section II. The diffusion coefficient D_{12} of atomic hydrogen in molecular hydrogen is taken from Ref. 20.

5. We shall now consider the calculation of the probabilities γ_v of surface deactivation of molecules on the channel walls. The calculations performed in the present paper, both taking into account and neglecting the surface deactivation of H_2 molecules, using expression (5) show that the VDF $f_v(h_2)$ of H_2 molecules in high vibrational levels depends strongly on γ_v. This makes it necessary to take account of the specific mechanism of the interaction of H_2 molecules with the walls in the theoretical analysis of the flow of vibrationally excited hydrogen in the channel.

It is obvious that to decrease losses of vibrationally excited molecules at the walls the channel walls must be made of a material with the largest possible potential barrier for adsorption of molecular hydrogen on the surface. In what follows, as an example of such material, we consider copper, for which values of the barrier in the range 0.8–1.5 eV have been found.[21–24] The most complete data known to us on the probabilities of surface deactivation of H_2 molecules on copper are contained in Ref. 23. In Ref. 23 it was shown that for the comparatively low kinetic energies of H_2 molecules interacting with a wall ($E_{\text{kin}}<0.1$ eV, as studied in the present work), the surface deactivation is due to the tunneling of H_2 molecules through the surface potential barrier, and the probability $w_v(E_{\text{kin}})$ of such tunneling was determined as a function of the vibrational quantum number v for $v\leqslant10$ and molecular kinetic energy $E_{\text{kin}}>0.1$ eV. In the present work we set $\gamma_v=w_v(E_{\text{kin}})$ in the calculations. The data of Ref. 23 were extrapolated to vibrational numbers $v=11$–14. Since it is difficult to extrapolate the data of Ref. 23 into the region $E_{\text{kin}}<0.1$ eV, in the present work the values of w_v calculated for $E_{\text{kin}}=0.1$ eV have been used in the calculations. For some values of v this appreciably increases the probability w_v of surface deactivation of the molecules. It was assumed that in the case of stationary gas flow the interaction with the wall does not change the total number of H_2 molecules in the flow, i.e., the decrease in the number of vibrationally excited molecules as they relax on the wall is compensated by desorption of ground-state molecules ($v=0$) from the wall. In Ref. 23 two sets of data were obtained, corresponding to two different orientations of the molecular axis relative to the surface — parallel to and perpendicular to the surface — with the molecule located at the top of the potential barrier at the solid–gas interface. The calculations showed that the results are essentially independent of the orientation of the molecule. We employed the data referring to the case in which a molecule located at the top of a potential barrier is oriented perpendicular to the surface.

6. We shall now examine the computational results. In the calculations, the density $N_{\text{H}}(x)$ of hydrogen atoms along the channel (see Appendix) was found simultaneously with the VDF of the H_2 molecules from the system of equations (4). In accordance with Eq. (A8) $N_{\text{H}}(x)$ was determined from the equation

$$\frac{d}{dx}(N_{\text{H}}V)=-N_{\text{H}}/\tau_d^{(\text{H})}. \tag{6}$$

As initial conditions (at $x=0$) for Eqs. (4) and (6) we used the known VDF $f_v^{(0)}$ and the known molecular density $N_{\text{H}_2}(0)=N_{\text{H}_2}^{(0)}T_0/T$ and atomic density $N_{\text{H}}(0)=N_{\text{H}}^{(0)}T_0/T$ established after the temperature jump in the gas at the channel entrance. The computational results are displayed in Figs. 1b and 4, 5, and 6.

The ratio of the VDF $f_v(x)$ formed during the hydrogen flow in the channel, to the initial VDF $f_v(0)=f_v^{(0)}$ formed in the discharge is presented in Figs. 4a, 5a, and 6a. One can see that the process leading to the establishment of the VDF $f_v(x)$ in the channel can be divided into two stages. At the first stage the populations of the highly excited vibrational states decrease rapidly as a result of the rapid $v-t$ relaxation on the hydrogen atoms, which is a consequence of the very large values of the constants for $v-t$ exchange of H_2 molecules with H atoms. At the second stage a very strong (for certain levels by many order of magnitude) increase of the populations occurs as a result of vibrational pumping by non-

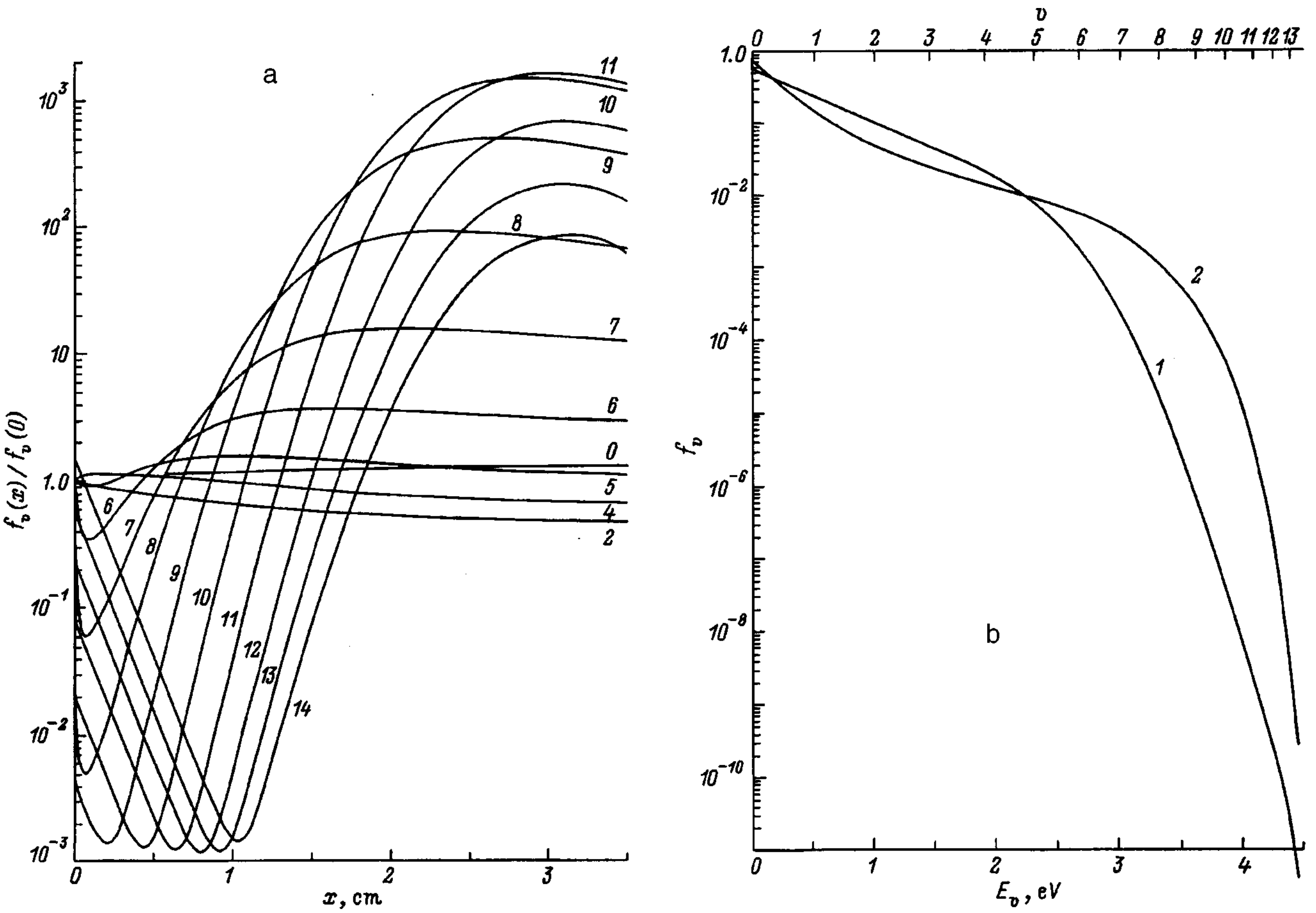

FIG. 4. Variation of the vibrational distribution function during hydrogen flow in a channel: a — Along the channel (the numbers of the vibrational levels are indicated on the curves); b — vibrational distribution function at the beginning and end of the channel: *1* — $f_v(0)$, *2* — $f_v(h_2)$; $h_2=3.5$ cm; the discharge parameters and the gas temperature in the channel are the same as in Fig. 1b.

resonant $v-v$ exchange in the cooled gas. We note that the increase in the populations of the vibrationally excited H_2 molecules accompanying $v-v$ exchange was observed in Refs. 25–27 in the afterglow of discharges containing vibrationally excited hydrogen. Figure 5c displays separately on an enlarged scale the initial stage of deformation of the VDF of the molecules at the channel entrance. Figure 5c refers mainly to several upper levels and illustrates the initial increase (cf. Fig. 5a) in the populations of the highest levels due to their population by recombination processes when the gas temperature decreases by an amount ΔT.

The results displayed in Figs. 4 and 5 were obtained taking account of the vibrational deactivation of H_2 molecules on the channel walls. The results presented in Fig. 6 were obtained for the same parameters of the discharge and flow as in Fig. 5 but neglecting the vibrational deactivation at the walls, i.e., for $\gamma_v=0$. Comparing Figs. 5a and 6a shows that in the example considered taking account of H_2 deactivation on the channel walls decreases the VDF $f_v(h_2)$ by approximately two orders of magnitude for the vibrational levels whose populations are most intensively enhanced in the channel. In a variety of other examples the difference between the computational results obtained taking into account and neglecting surface deactivation is much smaller. For example, for computational parameters corresponding to Fig. 4 this difference does not exceed one order of magnitude.

Figures 4b and 6b display the initial $f_v(0)=f_v^{(0)}$ and final $f_v(h_2)$ molecular VDFs. Figures 5b and 6b show in addition plots of the quantities

$$\Gamma_v(h,T_e')=\frac{f_v(h)K_v(T_e')}{\langle K_{\mathrm{DA}}(h,T_e')\rangle}, \tag{7}$$

where

$$\langle K_{\mathrm{DA}}(h,T_e')\rangle=\sum_v f_v(h)K_v(T_e') \tag{8}$$

it is the effective dissociative attachment constant, which gives the total rate of H^- formation due to electron attachment to all vibrational levels.

The quantity Γ_v is the probability (normalized to 1) of a H^- ion being produced in the process of dissociative attachment of electrons having temperature T_e' to molecules excited to the level v and possessing a VDF $f_v(h)$. Here $K_v(T_e)$ is the corresponding dissociative attachment constant[6] and h is the channel length. The quantity Γ_v characterizes the relative contribution of different vibrational levels to the dissociative attachment process. It is assumed that this process occurs in the chamber into which the vibra-

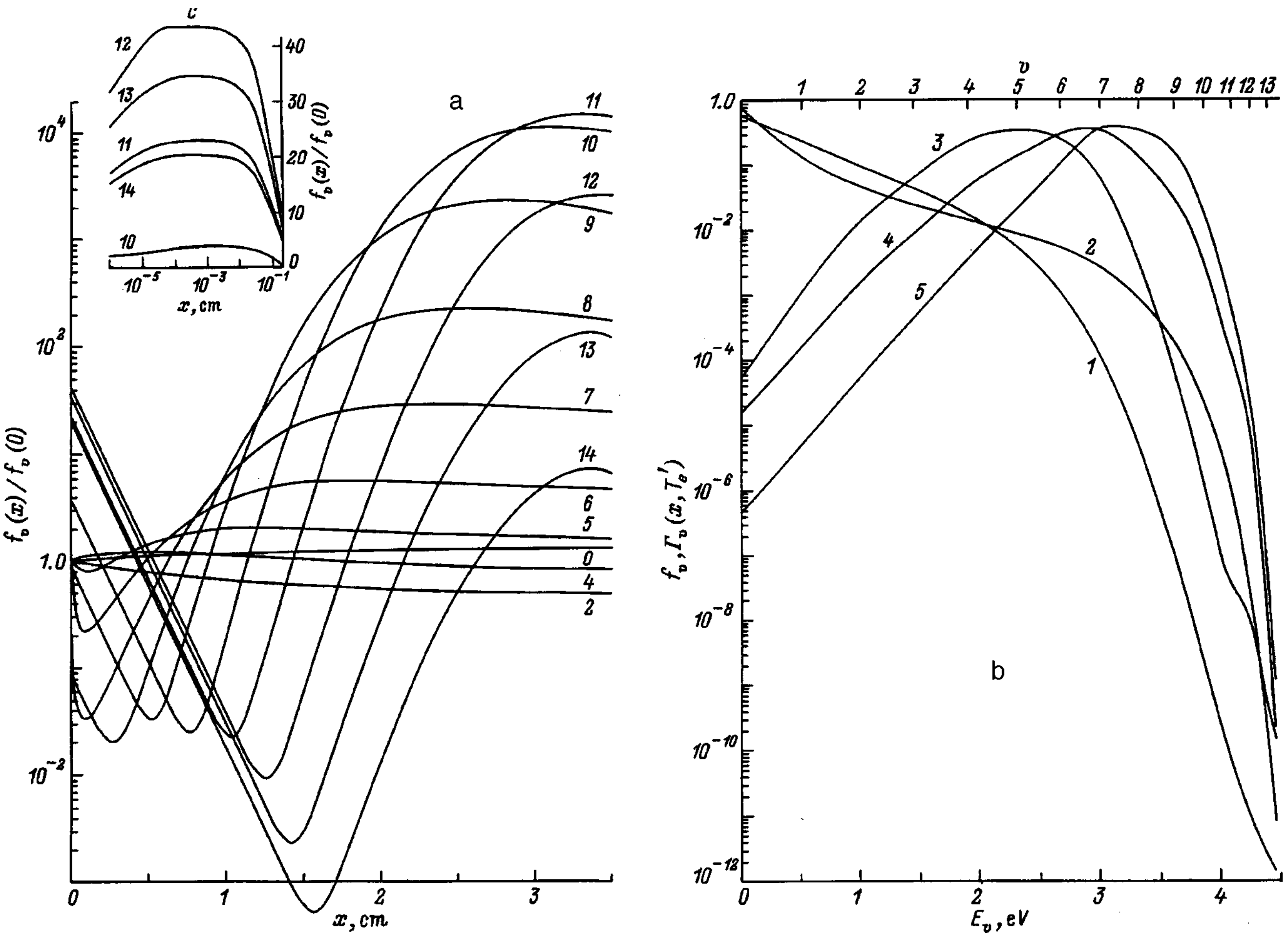

FIG. 5. Same as Fig. 4: a — Along the channel (the numbers of the vibrational levels are indicated on the curves); b — vibrational distribution functions and relative contribution of different levels to dissociative attachment at the beginning and end of the channel: *1* — $f_v(0)$, *2* — $f_v(h_2)$, *3* — Γ_v(0.07 eV), *4* — $\Gamma_v(h_2$, 0.7 eV), *5* — $\Gamma_v(h_2$, 0.2 eV); c — initial stage of vibrational relaxation in the channel: the parameters of the discharge plasma are the same as for Fig. 2; the gas temperature in the discharge $T_0=0.06$ eV and in the channel $T=0.03$ eV; $j_s=4.5$ A/cm^2, $j\simeq 5$ A/cm^2, $U\simeq 7.9$ V, $\varphi_1=8.65$ V, $\varphi_2=0.75$ V.

tionally pumped hydrogen flows and in which the electrons have a Maxwellian distribution with temperature T'_e. One can see how after passage through the channel the maximum of Γ_v shifts to higher values of v, i.e., to higher values of the dissociative attachment constants.[6] In the present paper the channel and flow parameters were not optimized for the purpose of maximizing the effective constant $\langle K_{\mathrm{DA}}(h,T'_e)\rangle$ as a result of the passage of vibrationally pumped hydrogen through the channel. A special paper will be devoted to this subject. Here we shall examine only the choice of the optimal electron temperature T'_e for dissociative attachment in the chamber into which the vibrationally pumped hydrogen flows. In Fig. 7 (cf. Fig. 8 in Ref. 4) the effective dissociative attachment constant is plotted as a function of T'_e for several computational variants: without enhancement in the channel (curve *1*) and after "enhancement" of the vibrational distribution in the channel (curves *2* and *3*, for which the vibrational deactivation of molecules on the channel walls is and is not taken into account, respectively). One can see that the greatest gain in $\langle K_{\mathrm{DA}}\rangle$ is obtained at low temperature T'_e, when comparatively high vibrational levels ($v\simeq 7$–9), whose populations are appreciably enhanced as a result of vibrational pumping in the channel, make the main contribution to dissociative attachment. The largest increase in vibrational pumping of H_2 molecules in the channel is achieved on the levels $v\simeq 9$–12. Although these levels make a relatively small contribution to the dissociative attachment, their vibrational pumping can be very substantial in a variety of plasma chemical applications.

In summary, it was shown that a very considerable increase in the populations of vibrational states of the hydrogen molecule in a quite high part of the vibrational spectrum can be obtained by appropriately organizing a flow of vibrationally pumped hydrogen in a channel.

This work was supported by Grant INTAS No. 94-316.

We thank Yu. Z. Ionikh and S. M. Shkol'nik for a helpful discussion.

APPENDIX

The diffusion of H atoms in a flow of molecular hydrogen is described by the equation

$$\mathbf{V}\cdot\nabla N_{\mathrm{H}}-D_{12}\nabla^2 N_{\mathrm{H}}=0, \tag{A1}$$

where $\mathbf{V}\equiv\mathbf{V}(0)$ is the flow velocity at the entrance into the channel of section II.

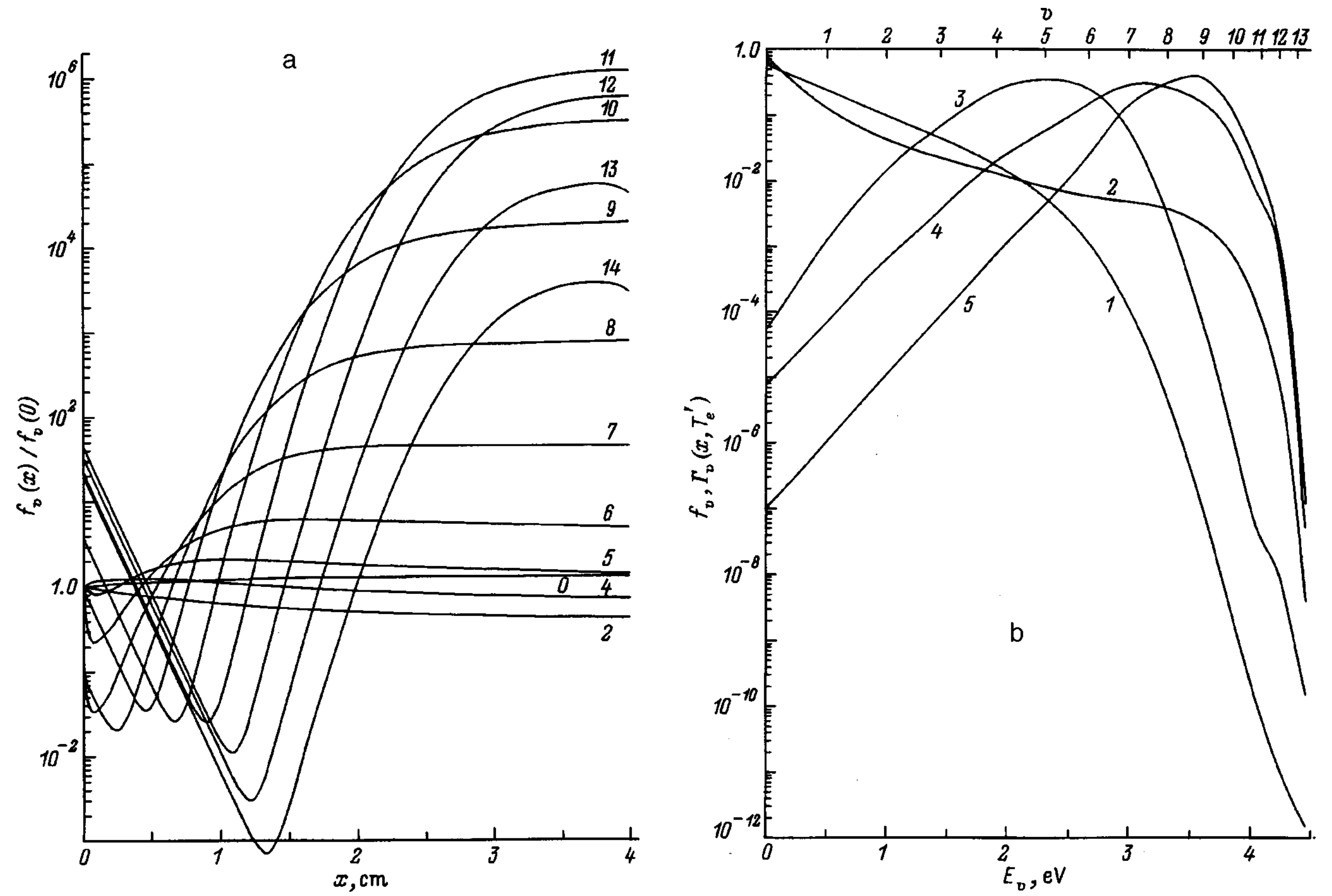

FIG. 6. Same as in Fig. 5 but neglecting vibrational deactivation of molecules on the channel walls.

In Eq. (A1) the recombination of atomic hydrogen into a H_2 molecule inside the channel[28] and the change in the flow velocity **V** on the initial section of the channel are neglected. The boundary conditions to Eq. (A1) correspond to prescribing the initial distribution $N_H(0,y)$ of the atoms at the channel entrance and setting the density N_H on the absorbing walls and at infinity equal to zero:

$$N_H(x,0)=N_H(x,L)=0, \tag{A2}$$

$$N_H(\infty,y)=0. \tag{A3}$$

Expanding $N_H(x,y)$ in a Fourier series

$$N_H(x,y)=\sum_{k=1}^{\infty}\tilde{N}_H(x)\sin\frac{\pi k y}{L} \tag{A4}$$

and substituting expression (A4) into Eq. (A1) we find

$$N_H(x,y)=\sum_{k=1}^{\infty}\tilde{N}_k(0)\times\exp\left\{-[\sqrt{1+(\delta k)^2}-1]\frac{x}{\delta\Lambda}\right\}\sin\frac{ky}{\Lambda}, \tag{A5}$$

where $\delta=2\Lambda/V\tau_d^{(H)}$ and $\tau_d^{(H)}=\Lambda^2/D_{12}$.

From Eq. (A5) we determine the atomic hydrogen density $\langle N_H\rangle$ averaged over the cross section of the channel

$$\langle N_H\rangle=\sum_{k=1}^{\infty}\tilde{N}_k(0)\times\exp\left\{-[\sqrt{1+(\delta k)^2}-1]\frac{x}{\delta\Lambda}\right\}\frac{1+(-1)^{k+1}}{\pi k}. \tag{A6}$$

Under the conditions considered here one has $\delta<1$ ($\delta\simeq0.5$), and the important values of x are those for which $x\sim V\tau_d^{(H)}$, i.e., $\Lambda/x\simeq\delta/2$. Then it is sufficient to take into account only the term with $k=1$ in the sum over k in Eq. (A6). This gives

$$\langle N_H\rangle\cong\frac{2\tilde{N}_1(0)}{\pi}\exp\left[-\frac{x}{V\cdot\tau_d^{(H)}}\right] \tag{A7}$$

and

$$\frac{dI_H}{dx}=-\frac{\langle N_H\rangle}{\tau_d^{(H)}}, \tag{A8}$$

where $I_H=V\langle N_H\rangle$ is the atomic hydrogen flux density averaged over the cross section.

One can see from Eq. (A8) that the quantities Λ and $\tau_d^{(H)}$ introduced above are indeed the effective length and effective diffusion time, respectively, of atomic hydrogen from the core of the flow to the channel wall. For $2\Lambda<V\tau_d$ an

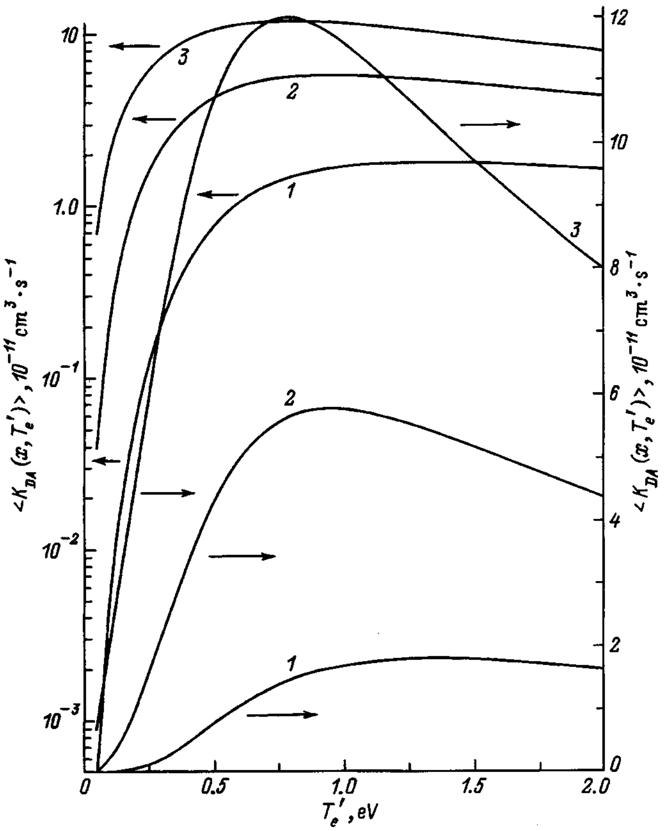

FIG. 7. Effective dissociative attachment constant $\langle K_{DA}(x,T'_e)\rangle$ versus the electron temperature T'_e (the discharge and flow parameters in the channel are the same as in Figs. 5 and 6): *1* — for $x=0$; *2,3* — for $x=h_2$, calculated with and without (respectively) allowance for the vibrational deactivation of hydrogen molecules on the channel walls. The units along the ordinates differ only in scale.

equation of the type (A8) also describes approximately the losses of other impurities present in the flow at the channel walls.

[1] M. Bacal and G. W. Hamilton, Phys. Rev. Lett. **42**, 1538 (1979).

[2] J. R. Hiskes, J. Appl. Phys. **51**, 4592 (1980).

[3] M. Bacal, Phys. Rev. B **37/38**, 28 (1989).

[4] D. A. Skinner, A. M. Bruneteau, P. Berlemont, and C. Courteille, Phys. Rev. E **48**, 2122 (1993).

[5] M. Bacal and D. A. Skinner, Comments At. Mol. Phys. **23**, 283 (1990).

[6] J. M. Wadehra, Phys. Rev. A **29**, 106 (1984).

[7] F. G. Baksht and V. G. Ivanov, Zh. Tekh. Fiz. **62**, 195 (1992) [Sov. Phys. Tech. Phys. **37**, 223 (1992)].

[8] F. G. Baksht, L. I. Elizarov, and V. G. Ivanov, Fiz. Plazmy **29**, 854 (1990) [Sov. J. Plasma Phys. **29**, 497 (1990)].

[9] F. G. Baksht, G. A. Djuzhev, and L. I. Elizarov, Plasma Sources Sci. Technol. **3**, 88 (1994).

[10] F. G. Baksht and V. G. Ivanov, Pis'ma Zh. Tekh. Fiz. **23**(1), 26 (1997) [Sov. Tech. Phys. Lett. **23**(1), 20 (1997)].

[11] S. Dushman, *Scientific Foundations of Vacuum Technique* (Wiley, New York, 1962, Mir, Moscow, 1964) 714 pp.

[12] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon Press, New York, 1987; Russian original, Nauka, Moscow, 1986).

[13] F. G. Baksht, L. I. Elizarov, V. G. Ivanov, and V. G. Yur'ev, Fiz. Plazmy **14**, 91 (1988) [Sov. J. Plasma Phys. **14**, 56 (1988)].

[14] F. G. Baksht and V. G. Ivanov, Fiz. Plazmy **12**, 286 (1986) [Sov. J. Plasma Phys. **12**, 165 (1986)].

[15] F. G. Baksht and V. G. Ivanov, Zh. Tekh. Fiz. **66**(9), 58 (1996) [Sov. Phys. Tech. Phys. **41**, 890 (1996)].

[16] H. Schlichting, *Boundary Layer Theory* (McGraw-Hill, New York, 1986; Nauka, Moscow, 1974) 711 pp.

[17] F. G. Baksht, G. A. Dyuzhev, A. M. Martsinovskiĭ *et al.*, *Thermionic Converters and Low-Temperature Plasma* [in Russian], Nauka, Moscow, 1973, 480 pp.

[18] Yu. Z. Ionikh, Opt. Spektrosk. **51**, 76 (1981) [Opt. Spectrosc. **51**, 39 (1981)].

[19] A. V. Eletskiĭ, L. A. Palkina, and B. M. Smirnov, *Transport Phenomena in Weakly Ionized Plasma* [in Russian], Atomizdat, Moscow, 1975, 336 pp.

[20] G. Blyth, J. Chem. Soc., Faraday Trans. 1 **83**, 751 (1987).

[21] P. Madhavan and J. L. Whitten, J. Chem. Phys. **77**, 2673 (1982).

[22] J. Harris and S. Anderson, Phys. Rev. Lett. **55**, 1583 (1985).

[23] M. Cacciatore and G. D. Billing, Surf. Sci. **232**, 35 (1990).

[24] C. T. Reitner, D. J. Auerbach, and H. A. Michelsen, Phys. Rev. Lett. **68**, 1164 (1992).

[25] C. Gorse, M. Capitelli, and M. Bacal, Chem. Phys. **117**, 172 (1987).

[26] F. G. Baksht, L. I. Elizarov, V. G. Ivanov *et al.*, Pis'ma Zh. Tekh. Fiz. **19**(22), 39 (1993) [Sov. Tech. Phys. Lett. **19**, 716 (1993)].

[27] F. G. Baksht, V. G. Ivanov, A. G. Nikitin, and S. M. Shkol'nik, Pis'ma Zh. Tekh. Fiz. **20**(22), 83 (1994) [Tech. Phys. Lett. **20**(11), 927 (1994)].

[28] F. G. Baksht, Zh. Tekh. Fiz. **52**, 3 (1982) [Sov. Phys. Tech. Phys. **27**, 1 (1982)].

Translated by M. E. Alferieff

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Numerical investigation of cross-flow separation in three-dimensional supersonic flows around circular cones

Yu. P. Golovachov and N. V. Leont'eva

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

(Submitted June 25, 1997)
Zh. Tekh. Fiz. **68**, 20–26 (October 1998)

The problem of supersonic three-dimensional flow over sharp cones is solved using the system of Navier–Stokes equations in a locally conical approximation. Numerical solutions are found using an implicit finite-difference scheme with second-order exponential approximation of the equations with respect to the cross-flow coordinate. The appearance and development of cross-flow separation on the leeward side of the cones in a laminar regime of gas flow in the shock layer are investigated. © *1998 American Institute of Physics.* [S1063-7842(98)00310-9]

INTRODUCTION

As the angle of attack increases, the three-dimensional supersonic flow over conical bodies is accompanied by cross-flow separation, which results in the formation of a very complicated vortex structure on the leeward part of the body. The formation of vortices has a substantial effect on the aerodynamic characteristics and the heat transfer between the gas and the surface in the flow. An important feature of flow over conical bodies which have a smooth contour of the cross section, in particular, circular cones, is that the position of the point of cross-flow separation is not known in advance. The appearance of cross-flow separation and further development of the structure of the gas-dynamic field on the leeward side of the body are determined by molecular transport and viscous–nonviscous interaction processes, which makes it necessary to use the Navier–Stokes model in the calculations. Numerical investigations of three-dimensional supersonic flow over sharp cones have been performed in this model, for example, in Refs. 1–6. Besides cross-flow separation, the computational results demonstrate characteristic features of the flow, such as the appearance of secondary vortices and internal shocks, increase in the heat flux to the surface of the body in the leeward part of the symmetry plane of the flow, change in the shape of the forward shock wave, destruction of the symmetry of the separated flow, and transition to a nonstationary periodic regime of flow over the body. However, the existing results do not contain systematic data on the conditions for the appearance of these characteristic features and the development of structure of flow in the shock layer. Even for circular cones, the computational results refer to a comparatively narrow range of flow conditions that corresponds to existing experimental data for Reynolds numbers $\mathrm{Re} \gtrsim 10^5$ and high supersonic Mach numbers.

In the present paper we present computational results for flow over sharp circular cones at incident-flow Mach numbers $1.5 \leqslant M_\infty \leqslant 8$, Reynolds numbers $10^2 \lesssim \mathrm{Re} \lesssim 10^5$, and relative angles of attack $0 \leqslant \alpha/\vartheta_c \leqslant 3$, where α is the angle of attack and ϑ_c is the cone half angle. The appearance of cross-flow separation in the laminar regime of gas flow in the shock layer is investigated for flow over circular cones. Information about the limiting values of the angle of attack for which cross-flow separation arises and their dependence on the cone half angle, the Reynolds number, the Mach number, and the temperature of the surface in the flow is obtained from the numerical solutions of the Navier–Stokes equations. The change occurring in the structure of the gas-dynamic field and the pressure, heat flux, and friction stress distributions on the surface of cones when the conditions of flow over the cones change is investigated.

STATEMENT OF THE PROBLEM

The numerical investigation of three-dimensional supersonic flow over sharp cones is performed assuming locally conical gas flow in the shock layer.[7] The Navier–Stokes equations are written first in a cylindrical coordinate system (x, r, ϑ), where the x axis is aligned along the longitudinal axis of the cone. Next it is assume that the values of the gas-dynamic functions remain unchanged along rays drawn from the vertex of the body. Under this assumption

$$\frac{\partial}{\partial x} = \frac{r}{x}\frac{\partial}{\partial r} \tag{1}$$

and the calculation of three-dimensional flow over the body reduces to solving a two-dimensional problem in the plane $x=$ const, whose distance from the vertex of the cone is determined by the value of the Reynolds number $\mathrm{Re} = \rho_\infty V_\infty x/\mu_\infty$. Here the subscript ∞ marks parameters of the unperturbed incident flow. The applicability of such an approximation for calculating three-dimensional supersonic flow over sharp conical bodies is confirmed by the satisfactory agreement of the results with the solutions of three-dimensional problems and with the experimental data.[4,7,8]

The boundary conditions of the problem are formulated as follows. All components of the gas velocity vanish on the cone surface. Moreover, the temperature T_w or the condition of thermal insulation $\partial T/\partial n = 0$ is prescribed. The outer boundary of the computational region is set at the forward shock wave. Here the Rankine–Hugoniot relations are used as boundary conditions. Under these conditions the flow in the shock layer is stationary and symmetric relative to a plane containing the velocity vector of the incident flow. Correspondingly, the computational region is bounded by the half plane $0 \leqslant \vartheta \leqslant \pi$ using the symmetry conditions at the boundaries.

In the calculations the model of a homogeneous perfect gas with specific heat ratio $\gamma = 1.4$ is used. The coefficient of viscosity is calculated using the Sutherland formula for air and the Prandtl number $\mathrm{Pr} = 0.72$.

NUMERICAL METHOD

A change to the new independent variables

$$n = \frac{r - r_w}{r_s - r_w}, \quad s = \frac{\vartheta}{\pi}, \tag{2}$$

transforms the computational region into a unit square. Here $r = r_w$ and $r = r_s(\vartheta)$ is the equation describing the contours of the cross section of the body and the forward shock wave in the section $x =$ const.

The stationary solutions are found by relaxation using a constant-direction numerical method.[7] After a change to dimensionless variables, the Navier–Stokes equations in the locally conical approximation assume the form

$$E \frac{\partial \mathbf{X}}{\partial t} - \frac{1}{\rho \,\mathrm{Re}} \frac{\partial}{\partial n}\left(A \frac{\partial \mathbf{X}}{\partial n}\right) + B \frac{\partial \mathbf{X}}{\partial n} + C \frac{\partial \mathbf{X}}{\partial s} + D\,\mathbf{X} + \mathbf{F} = 0, \tag{3}$$

$$e \frac{\partial \mathbf{X}}{\partial t} + b \frac{\partial \mathbf{X}}{\partial n} + c \frac{\partial \mathbf{X}}{\partial s} + d\,\mathbf{X} + \mathbf{f} = 0. \tag{4}$$

Here $\mathbf{X} = \{T u w v p\}^T$ is a column vector of the desired functions; T, p,and ρ are the temperature, pressure, and density of the gas, respectively; u, v, and w are the components of the velocity vector in the x, n, and s directions, respectively. The matrix equation (3) includes the equation of energy balance and the equations for the components of the momentum in the x and s directions, and the matrix equation (4) contains the equation of momentum balance in a projection on n and the continuity equation.

The initial equations are approximated by a two-layer implicit finite-difference scheme[9] in which an exponential approximation is used in the transverse coordinate n for Eqs. (3), while difference formulas of the type

$$\frac{\partial f}{\partial n} \sim \frac{f_{j+1} - f_j}{\Delta n}, \quad j = 1, \dots, J, \tag{5}$$

where f stands for any of the unknown functions, are used for Eqs. (4).

The first element of the vector f in Eqs. (4) contains a repeated derivative of the normal component of the gas velocity. This derivative can be approximated by a symmetric difference formula

$$\frac{\partial}{\partial n}\left(\mu \frac{\partial v}{\partial n}\right) \sim \frac{\mu_{j+1}(v_{j+2} - v_j) - \mu_j (v_{j+1} - v_{j-1})}{2(\Delta n)^2},$$

$$j = 2, \dots, J-2. \tag{6}$$

At the mesh points near the boundaries $(j = 1, j = J-1)$ asymmetric, approximations are used instead of Eq. (6).

Repeated derivatives with respect to s and mixed derivatives in all equations are approximated by central difference relations. At large Reynolds numbers ($\mathrm{Re} \gtrsim 10^4$) the use of central difference formulas for derivatives with respect to s in the convective terms of the equations caused the numerical solution to oscillate in the circumferential direction. The oscillations were eliminated by switching to one-sided difference approximations using $(\pm)$ splitting of the corresponding matrix of coefficients[10]

$$C \frac{\partial \mathbf{X}}{\partial s} = C^{+}\left(\frac{\partial \mathbf{X}}{\partial s}\right)^{L} + C^{-}\left(\frac{\partial \mathbf{X}}{\partial s}\right)^{R},$$

$$C^{+} = R \frac{D + |D|}{2} R^{-1}, \quad C^{-} = R \frac{D - |D|}{2} R^{-1}. \tag{7}$$

Here D is a diagonal matrix of eigenvalues of the matrix C; R is a matrix consisting of the right-hand eigenvectors of the matrix C; the superscripts L and R correspond to left and right one-sided difference formulas. In the method employed, the calculation of each step in time is performed by iteration. All of the coefficients of the equations, terms containing derivatives with respect to s, and free terms are calculated from the results of the preceding iteration. As a result, the system of difference equations in the current iteration decomposes into independent linear subsystems on the rays $s =$ const. In the above-indicated method of difference approximation of the equations in the transverse coordinate, these subsystems have tridiagonal block matrices of coefficients and can be solved by vector elimination. The position of the forward shock wave is determined at the same time that the field of the gas-dynamic functions in the shock layer is calculated. For this, the condition that the surface in the flow is impenetrable is used; this condition is transported to the outer boundary of the computational region and attached to the Rankine–Hugoniot relations.

The difference scheme employed is of second-order accuracy in both spatial coordinates. Depending on the conditions of flow over the body, several tens to several hundreds of time steps are required to establish a stationary solution. The number of iterations in the calculation of a new time layer varied from 1 to 4. A large part of the calculations was performed on a mesh containing 31 rays $s =$const with 51 mesh points on each ray. The coordinate transformations described in Ref. 7, which increase the density of mesh points in regions where the gradients of the unknown functions are large, were used to construct the difference mesh.

COMPUTATIONAL RESULTS

The computational results in dimensionless form are presented in the figures. The pressure and heat flux are scaled to $\rho_\infty V_\infty^2$ and $\rho_\infty V_\infty^3$, respectively. The coefficient of friction is determined as the ratio of the friction stress on the cone surface to $\rho_\infty V_\infty^2$. The temperature T_w of the cone surface is given as a fraction of the stagnation temperature T^* of the incident flow. In the cases where the temperature T_w is not indicated, the results correspond to a thermally insulated cone surface.

We shall examine first the computational results referring to the appearance of a cross-flow separation. Figure 1 shows the limiting angle of attack α^* at which the cross-flow separation of flow appears versus the cone half angle for three values of the Reynolds number. The limiting angle of

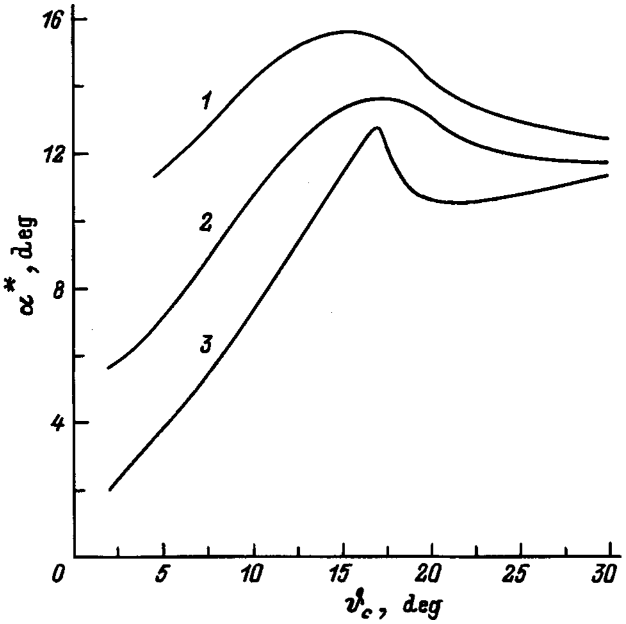

FIG. 1. Limiting angle of attack versus the cone half angle. $M_\infty=2$; Re: *1* — 10^3, *2* — 3770, *3* — 10^5.

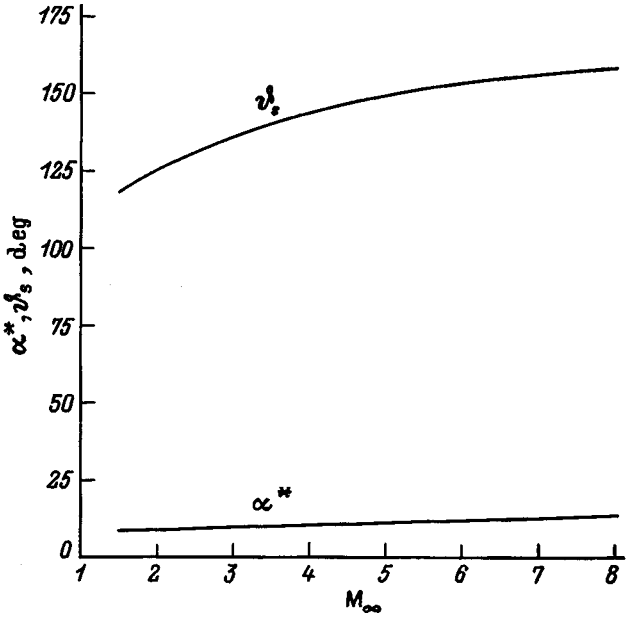

FIG. 2. Limiting angle of attack and the coordinate of the separation point versus the Mach number. $\mathrm{Re}=10^4$, $\vartheta_c=10°$, $\alpha=20°$, $T_w=0.25T^*$.

attack was determined from the change in sign of the derivative of the circumferential component of the coefficient of friction C_{f2} at the point $s=1.0$. One can see in the figure that, in the first place, the limiting angle of attack decreases as the Reynolds number increases. In the second place, the dependences are nonmonotonic: this indicates that the regime of nonseparated flow over the body lasts longer for cones with half angles $12\lesssim\vartheta_c\lesssim 18°$. As the Reynolds number increases, the maximum in the dependence $\alpha^*(\vartheta_c)$ becomes sharper. For $\vartheta\gtrsim 16°$ in all cases studied one has $\alpha^*<\vartheta_c$, i.e., separation occurs before the generator of the cone, $\vartheta=\pi$ moves into the ''shadow'' region. For $\vartheta\lesssim 16°$ the limiting angle of attack can, depending on the Reynolds number, be less than or greater than the cone half angle. The dependence of the limiting angle of attack on the Mach number of the incident flow, shown in the next figure (Fig. 2), is close to linear. As the temperature of the surface in the flow decreases, α^* decreases somewhat. The temperature factor becomes less influential as the Reynolds number increases.

The onset of cross-flow separation is illustrated in Fig. 3, which shows lines of constant values of the vorticity of the cross flow on the leeward side of a cone with half angle $\vartheta_c=12°$ and angle of attack $\alpha=10°$. The values of the vorticity are scaled to V_∞/x. As follows from Fig. 1, in the case at hand the angle of attack is somewhat larger than the critical value. The figure shows that cross-flow separation starts near the intersection of the contour of the body with the symmetry plane. As the angle of attack increases and separated flow develops, the center of flow moves away from the surface of the body and the symmetry plane.

The structure of the velocity field of the cross flow and its dependence on the temperature factor and the Reynolds number at large angles of attack are shown in Figs. 4a–4c. Here one can see the presence of a developed separated flow, whose intensity increases as the surface cools and the Reynolds number increases. Figure 5 displays for one variant of the flow conditions the lines of constant values of the Mach number of the cross flow $\mathrm{M}=\sqrt{v^2+w^2}/c$, where c is the local value of the sound velocity. The dotted curves correspond to M = 1.0. The unique form of the sonic line near the leeward part of the symmetry plane is interesting. It is also

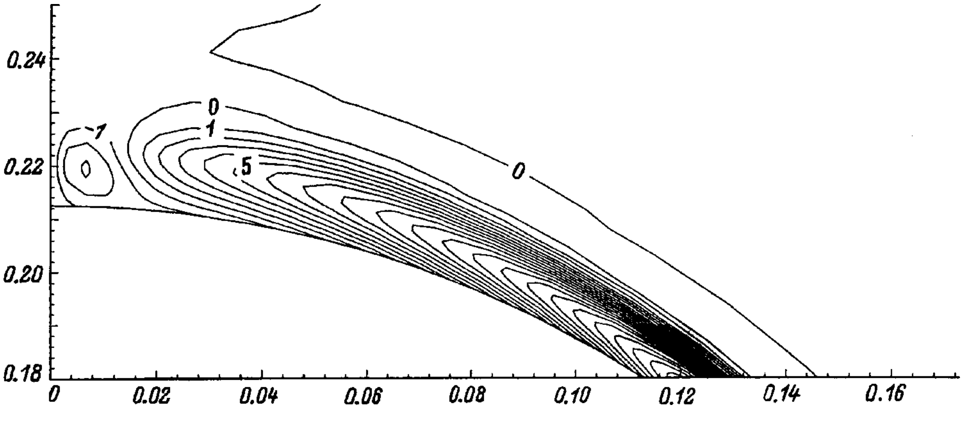

FIG. 3. Isolines of the cross-flow vorticity. $M_\infty=2$, $\mathrm{Re}=10^5$, $\vartheta_c=12°$, $\alpha=10°$.

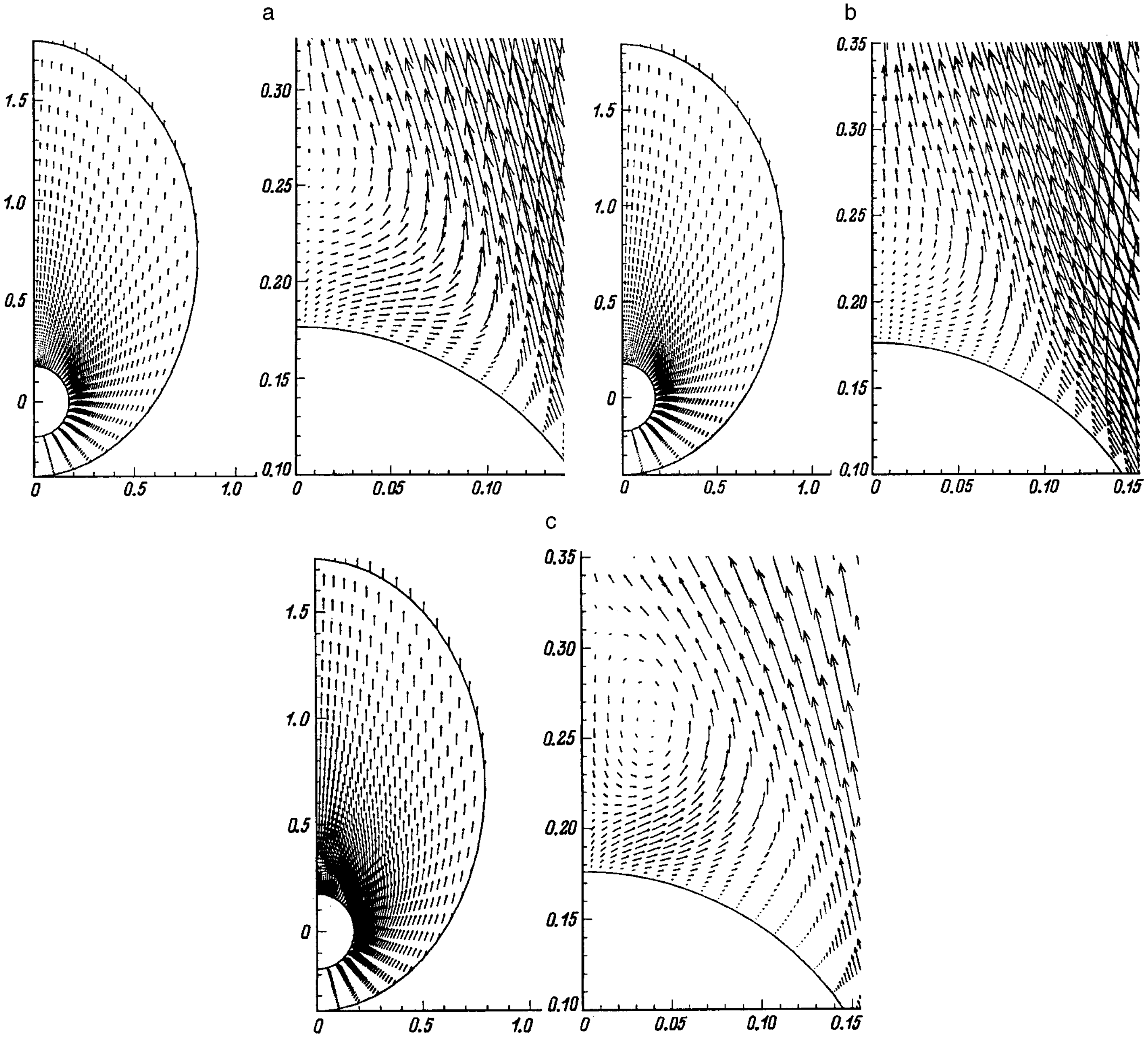

FIG. 4. Vector lines of the cross-flow velocity field. $M_\infty = 2$, $\vartheta_c = 10°$, $\alpha = 30°$; a — Re$=377$, $T_w = 0.25T^*$; b — Re$=377$; c — Re$=3770$.

seen in the figures that under the indicated conditions the forward shock wave has the normal form with maximum departure from the surface of the body in the leeward part of the symmetry plane of the flow. This shape of the forward shock wave was also obtained for other flow conditions.

Figure 6 shows the position of the point of cross-flow separation of the flow as a function of the angle of attack scaled to the cone half angle. One can see that for not too large values of the ratio α/ϑ_c, as the angle of attack α increases with ϑ_c held fixed, the region of separated flow becomes wider, and, conversely, as the half angle ϑ_c increases with α held fixed, this region contracts. This behavior persists up to a value of the relative angle of attack that decreases as the cone half angle increases. Further increase of the angle of attack has virtually no influence on the size of the separated region. It follows from the results presented that for $\alpha/\vartheta_c \lesssim 2.3$ the size of the region of separated flow is determined more by the change in the angle of attack than the cone half angle. This is evident from the fact that the contraction of the region of separated flow due to a decrease in α is greater than the expansion of the region due to the same decrease in ϑ_c. For $\alpha/\vartheta_c \gtrsim 2.3$ the dependence of the size of the separated region on the cone half angle with the ratio α/ϑ_c held fixed becomes nonmonotonic. Comparing the results for the cases of a thermally insulated cone surface and a surface at a fixed temperature (dashed and solid curve *2*) demonstrates the expansion of the region of separated flow when the surface in the flow is cooled. The results presented in Fig. 6 are in qualitative agreement with existing theoretical and experimental data[1–4,11] for large Reynolds numbers.

For large angles of attack, as the Reynolds number increases, a secondary separated flow arises. This effect is demonstrated in Fig. 7, which displays lines of constant vor-

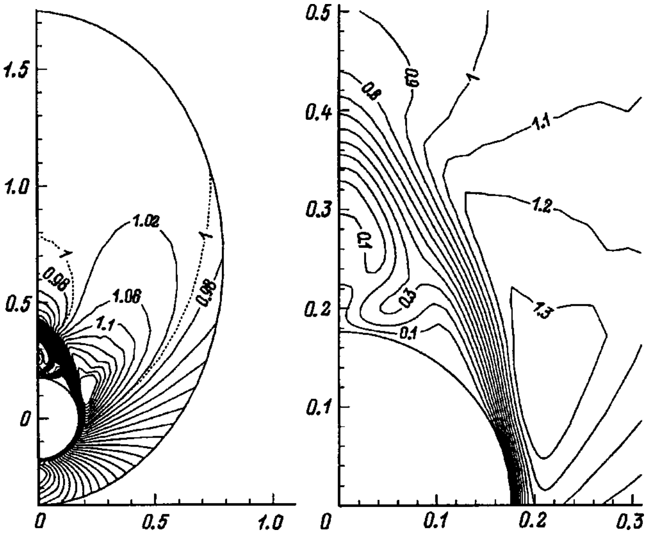

FIG. 5. Isolines of the cross-flow Mach number. $M_\infty = 2$, $Re = 3770$, $\vartheta_c = 10°$, $\alpha = 30°$.

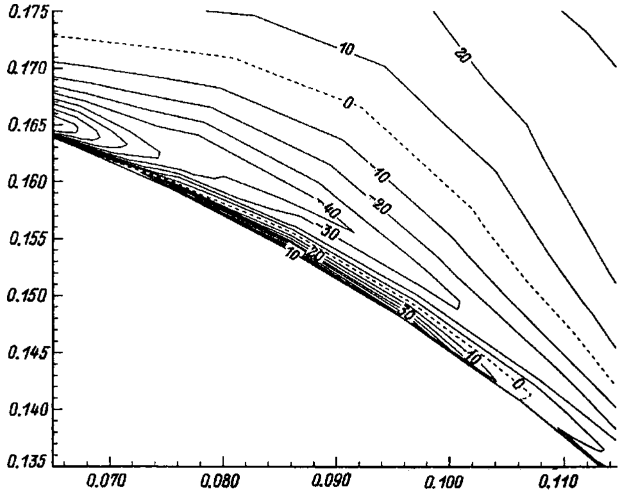

FIG. 7. Isolines of the cross-flow vorticity. $M_\infty = 2$, $Re = 0.9 \times 10^5$, $\vartheta_c = 10°$, $\alpha = 20°$, $T_w = 0.25T^*$.

ticity of the cross flow near the wall on the leeward side of the cone. Figure 8 shows the change in the boundaries of the main (curves *1*) and secondary (curves *2,2'*) separated flows as the Reynolds number increases. The solid lines correspond to $T_w = 0.25T^*$, and the dashed line corresponds to $T_w = 0.50T^*$. The figure also shows the results of measurements.[11] It is evident from the results presented that the influence of the temperature factor on the size of the main separated flow decreases as the Reynolds number increases. The dependence of ϑ_s on the Mach number of incident flow is shown in Fig. 2. Attachment of the main separated flow to the cone surface occurs at $\vartheta = \pi$. As one can see from Fig. 7, for sufficiently large Reynolds numbers a secondary separated flow arises inside the main separated flow. The variation of the coordinates of the points of separation and attachment of the secondary separated flow is shown in Fig. 8 (lines *2* and *2'*, respectively). For large Reynolds numbers the computed flow structure on the leeward side of the cone corresponds to the experimental results.[4,11,12]

Results demonstrating the influence of the Reynolds number on the change in pressure, the longitudinal component of the coefficient of friction, the angle of deflection of the limiting streamlines, and the heat flux on the cone surface are displayed in Figs. 9a–9d. The angle of deflection of the limiting streamlines from the direction of the generator of the cone was calculated from the formula $\Theta = \tan^{-1}(C_{f2}/C_{f1})$, where C_{f1} and C_{f2} are the components of the coefficient of surface friction along the generator and in the circumferential direction.

The changes in the distributions of the above-indicated quantities reflect the change in the flow structure in the shock layer as the Reynolds number increases. This change consists in a decrease of the thickness of the boundary layer on the windward part of the cone, intensification of the separated flow in the leeward region, and appearance of a secondary vortex. The coefficient of friction and the heat flux in the windward region vary with increasing Reynolds number ac-

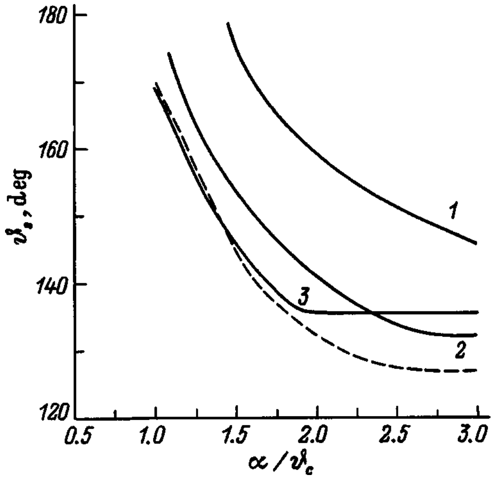

FIG. 6. Angular coordinate of the point of cross-flow separation versus the relative angle of attack. $M_\infty = 2$, $Re = 3770$, ϑ_c: *1* — 5, *2* — 10, *3* — 15°; dashed curve — $T_w = 0.25T^*$.

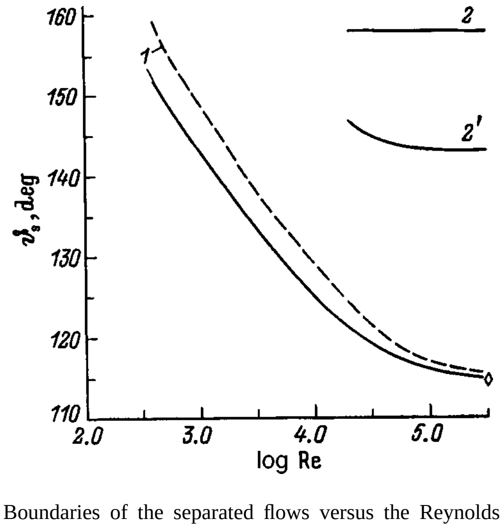

FIG. 8. Boundaries of the separated flows versus the Reynolds number. $M_\infty = 2$, $\vartheta_c = 10°$, $\alpha = 20°$, ◇ — experiment of Ref. 11.

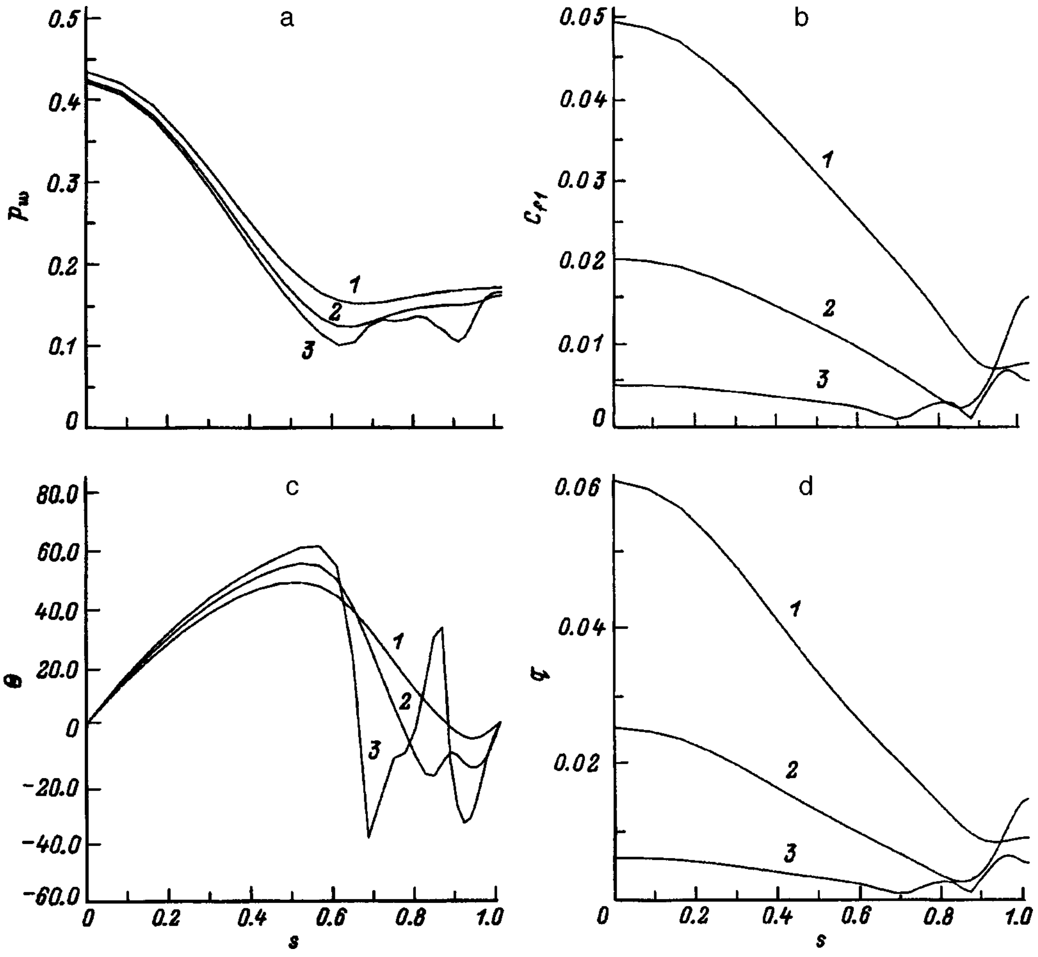

FIG. 9. Distribution of the pressure (a), coefficient of friction (b), directions of the limiting streamlines (c), and heat flux (d) over the surface of the cone. $M_\infty = 2$, $\vartheta_c = 10°$, $\alpha = 20°$, $T_w = 0.25T^*$, Re: *1* — 622, *2* — 3645, *3* — 0.6×10^5.

cording to the boundary-layered law, i.e., $\sim 1/\sqrt{\mathrm{Re}}$. The effect of a cross-flow separation on the distribution of these quantities is manifested as a breakdown of their monotonic decrease in the circumferential direction and the appearance of maxima at $\vartheta = \pi$. For $\mathrm{Re} = 0.6 \times 10^5$, because of the appearance of a secondary separated flow, the variation in all quantities, presented in Fig. 9, on the leeward part of the body becomes more complicated, and the values of the coefficient of friction and heat flux at the point $\vartheta = \pi$ exceed their values at $\vartheta = 0$. The increase in the heat flux in the leeward part of the symmetry plane of the flow was noted earlier in experimental and numerical investigations.[2,3,11,12] It is explained by the presence of a return cross flow, which carries gas from the high-temperature outer region of the shock layer into the region near the wall.

CONCLUSIONS

A numerical investigation of the appearance and development of a separated cross flow on the leeward side of sharp circular cones was performed using a locally conical approximation of the Navier–Stokes system of equations. The dependence of the structure of the gas-dynamic field and distributions of the resistance and heat-transfer parameters on the cone half angle, the Reynolds and Mach numbers of the incident flow, and the temperature factor was investigated.

This work was supported by the Russian Fund for Fundamental Research (Project No. 96-01-00596).

[1] A. Lin and S. G. Rubin, J. Fluid Mech. **59**, Pt. 3, 593 (1973).
[2] S. C. Lubard and W. C. Helliwell, AIAA J. **12**, 965 (1974).
[3] A. Lin and S. G. Rubin, AIAA J. **20**, 1500 (1982).
[4] D. J. Peake, D. F. Fisher, and D. S. McRae, AIAA J. **20**, 1338 (1982).
[5] R. W. Newsome, AIAA J. **24**, 552 (1986).
[6] O. A. Kandil, T. C. Wong, and C. H. Liu, in *Proceedings of the AGARD Conference* (1990), No. 0493.
[7] Yu. P. Golovachev, *Numerical Simulation of Viscous Gas Flow in a Shock Layer* [in Russian], Nauka, Moscow (1996), 376 pp.
[8] Yu. P. Golovachov and V. O. Chistyakov, in *Proceedings of the II Asian CFD Conference*, edited by H. Kubota and S. Aso, Tokyo (1996), Vol. 2, pp. 457–462.
[9] V. L. Belousov, Yu. P. Golovachev, and V. V. Zemlyakov, Matematicheskoe Modelirovanie **6**, 66 (1994).
[10] S. R. Chakravarthy, D. A. Anderson, and M. D. Salas, AIAA Paper No. 0268 (1980).
[11] V. S. Avduevskiĭ and K. I. Medvedev, Izv. Akad. Nauk SSSR Mekh. Zhidk. Gaza, No. 2, 117 (1966).
[12] J. Marcillat and B. Roux, AIAA J. **10**, 1625 (1972).

Translated by M. E. Alferieff

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Capillary oscillations of a thin film of viscous liquid on a solid substrate and its instability against surface charge

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

Yaroslavl State University, 150000 Yaroslavl, Russia

(Submitted June 16, 1997)
Zh. Tekh. Fiz. **68**, 27–31 (October 1998)

In a numerical analysis of the dispersion relation describing capillary motions in a thin film of a viscous, surface-charged liquid with fluctuation forces taken into account, it is found that the critical conditions of instability of the free surface of the liquid for a fixed thickness d of the liquid film in the region where the influence of the fluctuation forces is large ($d<100$ nm) depend strongly on the wave number and do not depend on the viscosity of the liquid, that the fluctuation forces strongly affect the wave number of the most unstable wavelength and decrease the instability growth rate, and that the capillary motions of the liquid admit an analogy with gravity–capillary motion and can be interpreted as fluctuation–capillary motions.

[S1063-7842(98)00410-3]

The instability of a charged surface (Tonks–Frenkel instability (TFI)) of a thin liquid film on a solid substrate is encountered in many academic, technical, and technological problems: the problem of the stability of a water layer on the surface of an ice core (melting hailstone) in a thundercloud or in a vacuum liquid mass spectrometer; the problem of the stability of a layer of liquid metal in liquid-metal ion sources, where electrospraying occurs from the lateral surface of the emitter tip along which the liquid metal is supplied (see, for example, Refs. 1–3). Capillary oscillations and the TFI of a liquid of finite depth have been investigated repeatedly.[1–7] However, the merits of existing results, especially the asymptotic transitions to a thin layer,[7] greatly diminish for films of thickness $d\leqslant 100$ nm. As shown in Ref. 5, for such a film the intermolecular interaction of the liquid with the solid substrate must be taken into account. This interaction leads to the appearance of an additional pressure, the so-called disjoining pressure due to fluctuation forces, which increases rapidly as the thickness of the liquid layer decreases: $\sim d^{-r}$ $(3\leqslant r\leqslant 4)$. It is reasonable to expect the critical conditions for the TFI in such a system to change.[8] In existing works on the TFI[3,6,7,9] the disjoining pressure was neglected or the electrohydrodynamic instability was analyzed in the model of an ideal liquid.[8]

1. We shall solve the problem of calculating the spectrum of capillary waves on a flat charged surface of an ideally conducting liquid layer, adjoining the vacuum, of thickness d, density ρ, viscosity ν, and surface tension γ in a gravitational field $\mathbf{g}$ and an electrostatic field. The top medium has a dielectric constant ε. The intensity $\mathbf{E}$ of the electric field at the surface of the liquid is determined by the potential difference between electrodes: the bottom electrode at $z=-d$, on which the liquid layer lies, at potential $\Phi_1=0$ and a parallel counterelectrode, separated from the surface of the liquid by distance b, at potential $\Phi_2=V$.

Let the x axis of a Cartesian coordinate system, whose z axis is directed vertically upwards $\mathbf{n}_z\| -\mathbf{g}$ ($\mathbf{n}_z$ is the unit vector in the direction of the Cartesian z axis), determine the direction of motion of a capillary plane wave $\sim\exp(st+ikx)$, and let the plane $z=0$ coincide the free unperturbed surface of the liquid (s is the complex frequency, k is the wave number, t is the time, and $i=\sqrt{-1}$). Let the function $\xi(x,t)=\xi_0\exp(st+ikx)$ describe a small perturbation of the equilibrium flat surface of the liquid by a thermal capillary wave motion of very small amplitude ($\xi_0\sim(kT/\gamma)^{1/2}$, where k is Boltzmann's constant and T is the absolute temperature). The fluid velocity field $\mathbf{U}(\mathbf{r},t)$ produced by the perturbation $\xi(x,t)$ is an infinitesimal of the same order.

Current ideas about the nature of the disjoining pressure do not have a rigorous theoretical justification and do not admit a universal analytical description of the phenomenon. Aspiring only to a qualitative estimate of the degree to which this phenomenon influences the TFI, we shall confine ourselves to a model where the disjoining pressure on the perturbed surface $\xi(x,t)$ is related with the thickness of the deformed layer as

$$P_d=\frac{A}{(d+\xi)^3},$$

which corresponds to the theoretical model of Ref. 5. The value of the constant $A\simeq 10^{-13}$ erg is determined only in order of magnitude. To linearize the problem, an expansion of P_d linear in ξ near $z=0$ must be used:

$$z=0:\quad P_d=P_{d0}+P_{d1},$$

$$P_d=\frac{A}{d^3},\quad P_{d1}=-\frac{3A}{d^4}\xi. \tag{1}$$

The linearized system of equations of hydrodynamics of a viscous liquid has the form [Note: here Δ denotes the Laplacian operator]

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

$$\operatorname{div}\mathbf{U}=0, \tag{3}$$

$$z=-d:\quad \mathbf{U}=0, \tag{4}$$

$$z=0:\quad -\frac{\partial\xi(x,t)}{\partial t}+U_z=0, \tag{5}$$

$$\mathbf{n}\cdot(\boldsymbol{\tau}\cdot\nabla)\mathbf{U}+\boldsymbol{\tau}\cdot(\mathbf{n}\cdot\nabla)\mathbf{U}=0, \tag{6}$$

$$-P(\mathbf{U})+\rho g\xi+2\rho\nu\mathbf{n}\cdot(\mathbf{n}\cdot\nabla)\mathbf{U}-P_{d1}(\xi)-P_E(\xi)+P_\sigma(\xi)=0, \tag{7}$$

where $P(\mathbf{U})$, $P_E(\xi)$, and $P_\sigma(\xi)$ are corrections to the hydrostatic, electrostatic, and surface tension pressure forces due to the perturbation ξ of the surface, which are all first-order infinitesimals in ξ;[10,11] P_{d1} is defined in Eq. (1); and, $\mathbf{n}$ and $\boldsymbol{\tau}$ are unit vectors normal and tangent, respectively, to the surface of the liquid.

The electric potentials Φ_1 in the liquid and Φ_2 in the external medium each satisfy Laplace's equation:

$$\Delta\Phi_i=0\quad (i=1,2) \tag{8}$$

with the boundary conditions

$$z=-d:\quad \Phi_1=0, \tag{9}$$

$$z=b:\quad \Phi_2=V, \tag{10}$$

$$z=\xi:\quad \Phi_1=\Phi_2. \tag{11}$$

2. The problem was solved by the method described in detail in Ref. 10 — separation of the velocity field $\mathbf{U}=\mathbf{U}(\mathbf{r},t)$ into two components: a potential component (with velocity potential $\varphi(\mathbf{r},t)$) and a rotational component (described by a stream function $\Psi(\mathbf{r},t)$)

$$U_x=\frac{\partial\varphi}{\partial x}-\frac{\partial\Psi}{\partial z}, \tag{12}$$

$$U_z=\frac{\partial\varphi}{\partial z}+\frac{\partial\Psi}{\partial x}. \tag{13}$$

Finally, the scalar equations for φ and Ψ are

$$\frac{\partial\Psi}{\partial t}-\nu\Delta\Psi=0, \tag{14}$$

$$\Delta\varphi=0 \tag{15}$$

and the expression for $P(\mathbf{r},t)$ is

$$P(\mathbf{U})=-\rho\frac{\partial\varphi}{\partial t}-\rho g z. \tag{16}$$

Substituting expressions (12) and (13) into Eqs. (2)–(7) gives the boundary conditions for φ and Ψ:

$$z=-d:\quad \frac{\partial\varphi}{\partial x}-\frac{\partial\Psi}{\partial z}=0, \tag{17}$$

$$\frac{\partial\varphi}{\partial z}+\frac{\partial\Psi}{\partial x}=0, \tag{18}$$

$$z=0:\quad \frac{\partial\xi}{\partial t}=\frac{\partial\varphi}{\partial z}+\frac{\partial\Psi}{\partial x}, \tag{19}$$

$$2\frac{\partial^2\varphi}{\partial x\partial z}+\frac{\partial^2\Psi}{\partial x^2}-\frac{\partial^2\Psi}{\partial z^2}=0; \tag{20}$$

$$\rho\frac{\partial\varphi}{\partial t}+\rho g\xi+2\rho\nu\left\{\frac{\partial^2\varphi}{\partial z^2}-\frac{\partial^2\Psi}{\partial x\partial z}\right\}-P_{d1}(\xi)-P_E(\xi)+P_\sigma(\xi)=0. \tag{21}$$

The system of equations (14)–(16) with the boundary conditions (17)–(21) comprises the hydrodynamic part of the problem in a scalarized form. The electrical part of the problem (8)–(11) can be solved by the methods of classical electrodynamics, and it is not difficult to show[12] that

$$P_E(\xi)=Wk\xi\coth(kb),\quad W=\frac{\varepsilon E_0^2}{4\pi},\quad E_0\equiv V/b.$$

The parameter W is proportional to the surface charge density on the unperturbed surface. The expression for the correction to the pressure P_σ due to surface tension in an approximation linear in ξ is[10]

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

3. We shall seek the solutions of the system (14) and (15) which are bounded and periodic in x in the chosen coordinate system in the form

$$\varphi=(C_1\sinh(mz)+C_2\cosh(mz))\exp(st-ikx), \tag{22}$$

$$\Psi=(C_3\sinh(qz)+C_4\cosh(qz))\exp(st-ikx), \tag{23}$$

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

Substituting expressions (22) and (23) into the boundary conditions (17)–(21) yields a homogeneous system of five linear algebraic equations for the unknown constants C_1, C_2, C_3, C_4, and ξ_0

$$GC_1+2\rho\nu qC_3+F\xi_0=0,$$

$$2\rho\nu kC_2+GC_4=0,$$

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

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

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

where

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

$$F\equiv\rho g+\frac{3A}{d^4}+\gamma k^2-Wk\coth(kb).$$

A necessary and sufficient condition for the existence of a nontrivial solution of the system of equations obtained is that the determinant consisting of the coefficients of the unknown quantities C_i and ξ_0 vanish:

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

This condition gives a dispersion relation for the spectrum of capillary motions of the liquid in the system being analyzed. In dimensionless variables, in which $\rho=\gamma=d=1$, the dispersion relation has the form

$$\begin{aligned} &4qk^2(k^2+q^2)+(k^2+q^2)^2 \\ &\times(k\sinh(k)\sinh(q)-q\cosh(k)\cosh(q)) \\ &+4k^3q(q\sinh(k)\sinh(q)-k\cosh(k)\cosh(q)) \\ &-\frac{Z(k,W)}{\nu^2}(q\cosh(q)\sinh(k)-k\cosh(k)\sinh(q))=0, \\ &q^2\equiv k^2+s/\nu, \\ &Z(k,W)\equiv kg+3Ak+k^3-Wk^2\coth(kb). \qquad (24) \end{aligned}$$

Here the previous notation is retained for all dimensionless quantities measured in units of their characteristic scales: $g_*=\gamma/\rho d$, $A_*=\gamma d$, $\nu_*=\sqrt{\gamma d/\rho}$, $k_*=1/d$, $s_*=\sqrt{\rho d/\gamma}$, $W_*=\gamma/d$, $b_*=d$.

4. Equation (24) differs from the dispersion relation for the analogous problem in which the disjoining pressure is neglected[12] only by the definition of the parameter $Z(k,W)$. It is easy to see that it has many similar properties. For example, from the properties of hyperbolic functions[13,14] it follows that if (k,q) is a solution of Eq. (22), then $(k,-q)$ and (k,q) are also solutions of Eq. (24) (**q** is the complex conjugate of q). The simplest obvious solutions of Eq. (24) are $s=0$ $(q=k)$ and $s=-\nu k^2$ $(q=0)$. All solutions of the dispersion relation (24) are defined on the top sheet of a Riemann surface and theoretically are observable.

5. The criterion for the onset of the TFI in the system is $Z(k,W)<0$ (Refs. 10,11)

$$Z(k,W)\equiv kg+3Ak+k^3-Wk^2\coth(kb)<0.$$

Evidently, the equation $Z(k,W)=0$ describes in the coordinates (k,W) the boundary between the stable $(Z(k,W)\geqslant 0)$ and unstable $(Z(k,W)<0)$ states of the system. Figures 1 and 2 display the curves corresponding to the boundary of stability for different values of the distance b to the top electrode, with and without the disjoining pressure taken into account, respectively. One can see that for wavelengths $\lambda<10\,d$ the critical conditions are essentially independent of the fluctuation forces. The presence of a disjoining pressure has a large effect on the critical conditions for instability of the free surface only in the range of wavelengths which are much greater than the thickness of the film, but as one can see from the figures, it is in this case that the critical conditions in the formulation presented depend strongly on the geometry of the system, characterized by the dimensionless parameter b. It follows from Figs. 1 and 2 that in the situation studied, the effect of the fluctuation forces on the critical conditions for the TFI should be investigated for wave numbers $k<0.01$.

Two cases are of interest for a qualitative investigation of the effect of the disjoining pressure on the instability of

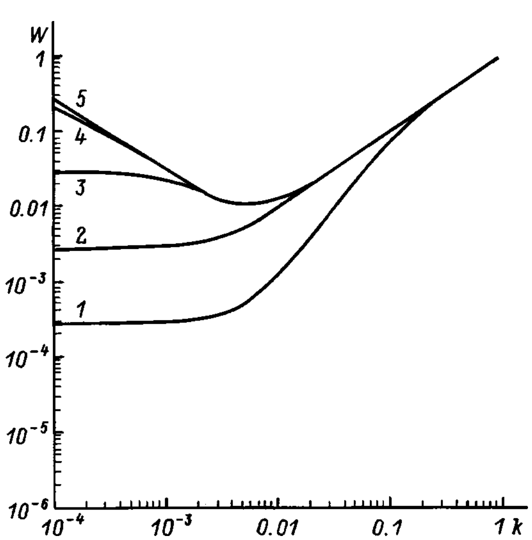

FIG. 1. Stability boundary for a charged free surface of a liquid $(Z(k,W)=0)$ with fluctuation forces taken into account: $A=10^{-5}$, $g=10^{-9}$; b: *1* — 10, *2*—10^2, *3*—10^3, *4*—10^4, *5*—10^5.

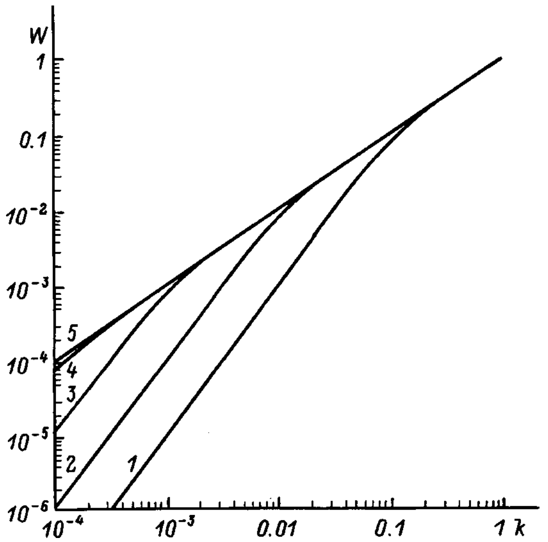

FIG. 2. Stability boundary for a charged free surface of a liquid $(Z(k,W)=0)$ neglecting fluctuation forces: $g=10^{-9}$: *1–5* — same as in Fig. 1.

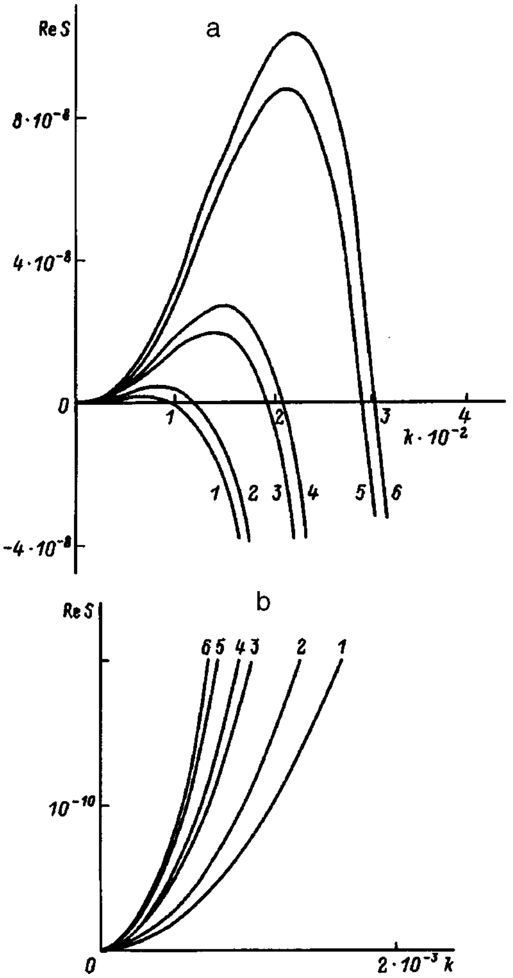

FIG. 3. Curves of the real component Re s of the dimensionless frequency versus the dimensionless wave number k for $\nu=0.3$, $b=100$, $A=10^{-5}$, $g=10^{-9}$, W: *1,2*—10^{-2}; *3,4*—2×10^{-2}; *5,6*—3×10^{-2}; a — on a small scale, b — left side on a larger scale.

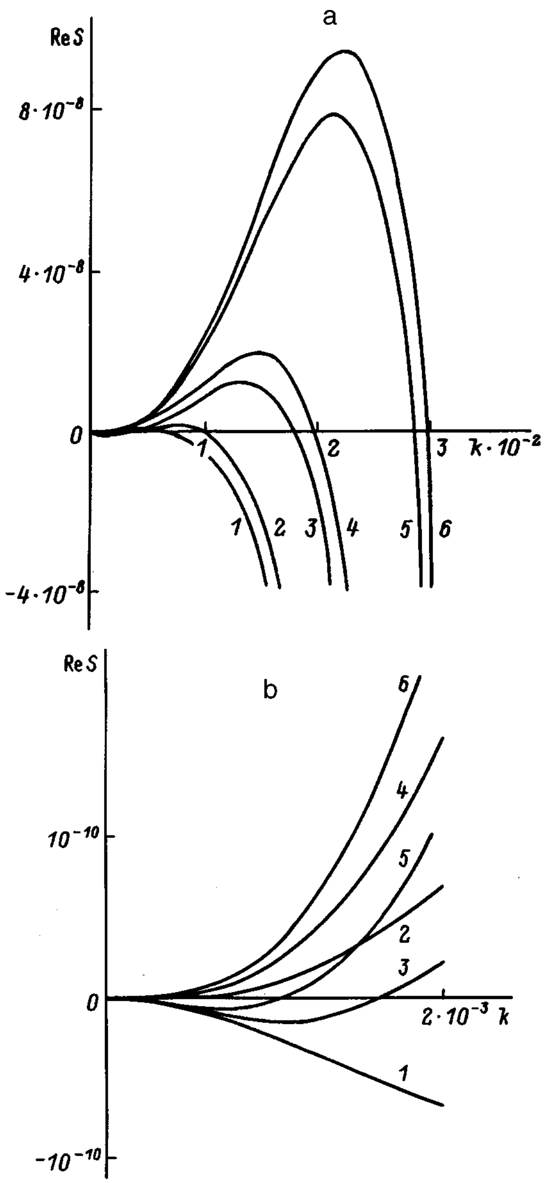

FIG. 4. Curves of the real component Re s of the dimensionless frequency versus the dimensionless wave number k for $b=10^5$, $A=10^{-5}$. The remaining parameters are the same as in Fig. 3; a — on a small scale; b — left side of a on a larger scale.

thin films: the case where the wavelength of the excited wave is comparable to the distance to the top electrode, $\lambda \simeq b$, and the case $b \gg \lambda$. For $k \simeq 0.01$ the values $b=10^2$ and 10^5, respectively, were chosen to model these two situations. The results of a numerical investigation of Eq. (24) are presented in Figs. 3 and 4 in the form of curves of the real part Re s of the dimensionless frequency versus the dimensionless wave number k for different values of the parameter W. The branches of the capillary motions presented in these figures characterize the behavior of the roots of the dispersion relation (24) that are associated with aperiodic instability of the liquid film. Odd-numbered branches were calculated with a large contribution from fluctuation forces and even-numbered branches were calculated without fluctuation forces $(A=0)$. Figures 3 and 4 make it possible to compare the character of the effect of a disjoining pressure on the TFI for $b=100$ and 10^5.

One can see from the figures that in both cases the fluctuation forces decrease the interval of wave numbers of unstable wavelengths on the side of large values of k by an amount determined by the parameter W. The wave number of the most unstable wavelength decreases when the disjoining pressure is taken into account. This is especially noticeable for the branches *1* and *2* in Figs. 3 and 4, which were constructed for lower values of W. The maximum value of the instability growth rate also decreases.

Figures 3b and 4a show in a larger scale the left-hand side of Figs. 3a and 4a, respectively, demonstrating that in a system with the parameter $b \gg \lambda$ the disjoining pressure also limits the range of k of unstable wavelengths on the side of small wave numbers. This limitation is especially noticeable for $W=2\times10^{-2}$ (branch *3* in Fig. 4a). For $W \leqslant 10^{-2}$ in a system with $b \gg \lambda$ fluctuation forces stabilize the motions of the liquid film in the entire range of values of k (Figs. 1 and 4a and 4b). This result was already mentioned in Ref. 7, where, however, the disjoining pressure was neglected, and the effect described was due to the presence of the term gk, which takes account of the gravitational forces, in the defi-

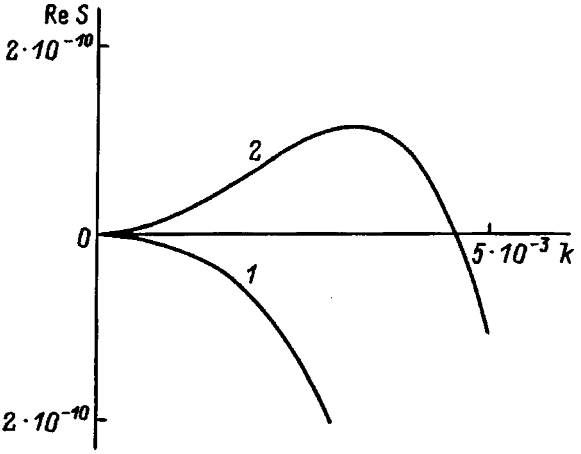

FIG. 5. Curves of the real component Re *s* of the dimensionless frequency versus the dimensionless wave number *k* with (*1*) and without (*2*) fluctuation forces taken into account, for $\nu=0.3$, $b=100$, $A=10^{-5}$, $g=10^{-9}$, and $W=2\times10^{-3}$.

nition of $Z(k,W)$. However, for the scaling chosen the dimensionless value $g\simeq10^{-9}$ for real liquids is several orders of magnitude smaller than the dimensionless value of the parameter $A=10^{-5}$, while the sum $kg+kA$ enters in the definition of $Z(k,W)$ from Eq. (24). It can be concluded that the effects caused in ''thick'' films (where fluctuation forces can be neglected) by the field **g** have analogs in the case of thin films, but now the effects are due not to gravitational but rather fluctuation forces. For this reason, the wave motion in thin films is more correctly termed fluctuation–capillary motion and it is more correct to talk about instability of fluctuation–capillary waves, just as in the case of gravity–capillary analogs.

It can be concluded from Fig. 1 that for $W=2\times10^{-3}$ and $\lambda\simeq b$ the fluctuation forces likewise can stabilize the motion in the system under study for all possible values of the wave numbers, in contrast to the situation where the disjoining pressure is neglected. This conclusion is confirmed by numerical calculations performed using Eq. (24), which are presented in Fig. 5. For the indicated value of W and taking account of the disjoining pressure only aperiodic motions (branch *2*) are observed, while the absence of fluctuation forces would have resulted in aperiodic instability (branch *1*).

CONCLUSIONS

Fluctuation forces have a large influence on the instability and capillary motions with wavelength $\lambda\gg d$ on a charged surface of a liquid film of thickness $d\leqslant100$ nm. The character of the manifestation of this influence depends on the dimensionless parameter b. For $b\simeq\lambda$ and fixed W the influence of the disjoining pressure is manifested as a decrease in the critical value of k at which the motion in the system becomes stable, a decrease in the value of k corresponding to the most unstable wavelength, and a decrease in the maximum instability growth rate. If $b\gg\lambda$ in addition to the effects indicated, then the range of values of k of the unstable harmonics is limited on the side of very small wave numbers. Fluctuation forces stabilize capillary motions in a liquid film for values of the parameter $W\leqslant10^{-2}$ in the case $\lambda\ll d$ and for $W\leqslant10^{-3}$ when $b\simeq\lambda$. The motion of the liquid in thin films should be thought of as fluctuation–capillary motion. The phenomenon of fluctuation–capillary motion is analogous to gravity–capillary phenomena.

[1] M. D. Gabovich, Usp. Fiz. Nauk **140**, 137 (1983) [Sov. Phys. Usp. **26**, 447 (1983)].

[2] A. G. Bailey, Atomization and Spray Technology **2**, 95 (1986).

[3] A. I. Grigor'ev and S. O. Shiryaeva, Uzv. Ross. Akad. Nauk, Mekh. Zhid. i Gaza, No. 3, 3 (1994).

[4] B. V. Deryagin, *Theory of the Stability of Colloids and Thin Films* [in Russian], Nauka, Moscow (1986), 205 pp.

[5] J. Frenkel, *Kinetic Theory of Liquids* (Clarendon Press, Oxford, 1946; Nauka, Leningrad, 1975), 592 pp.

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

[7] N. C. Surgy, J.-P. Chabrerie, O. Denoux, and J. E. Wesfreid, J. Phys. France. **3**, 1201 (1993).

[8] A. I. Grigor'ev, M. I. Munichev, and S. O. Shiryaeva, J. Colloid Interface Sci. **3**, 267 (1994).

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

[10] V. G. Levich, *Physicochemical Hydrodynamics* [in Russian], Fizmatgiz, Moscow (1959), 699 pp.

[11] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media* (Pergamon Press, New York; Russian original, Nauka, Moscow, 1992), 661 pp.

[12] D. F. Belonozhko, A. I. Grigor'ev, M. I. Munichev, and S. O. Shiryaeva, Pis'ma Zh. Tekh. Fiz. **22**(10), 84 (1996) [Tech. Phys. Lett. **22**(5), 425 (1996)].

[13] I. N. Bronshteĭn and K. A. Semendyaev, *Handbook of Mathematics for Engineers and Students at Institutes of Technology* [in Russian] (Nauka, Moscow 1980), 976 pp.

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

Translated by M. E. Alferieff

Parameters of the discharge plasmas of surface-plasma sources of negative hydrogen ions

V. V. Antsiferov

Institute of Nuclear Physics, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted May 21, 1997)
Zh. Tekh. Fiz. **68**, 32–38 (October 1998)

The main parameters of the plasma of high-current hydrogen–cesium glow discharges of surface-plasma (planotron and Penning) sources of negative hydrogen ions are determined using contact-free spectroscopic methods and compared for identical discharge current densities. The elemental and charge composition of the plasma is established. The temperature of the hydrogen atoms and the energy of the visible-range radiation of the plasma discharge are measured and estimates of the electron density in the plasma are made. The dynamics of the change in the parameters of the discharge plasma of a Penning source — the densities of hydrogen atoms, cesium atoms and ions, and molybdenum atoms — is tracked during a discharge pulse with spatial resolution along two coordinates. It is observed that cesium atoms and ions and molybdenum atoms are pent up near the cathode surface. © *1998 American Institute of Physics.* [S1063-7842(98)00510-8]

INTRODUCTION

Negative-hydrogen-ion sources employ mainly hydrogen–cesium discharges. Such discharges have small volumes and a complicated structure, and their parameters are strongly influenced by nonequilibrium processes occurring at the electrodes. The characteristics of the plasma in such discharges and the physical processes determining the parameters of the plasma have not been adequately studied because of the lack of reliable information about its parameters. The plasma parameters of such discharges have been measured, as a rule, by probe methods, whose use is strongly limited because of the small volumes of the plasma, the influence of the magnetic field of the source, and cesium adsorption. Essentially the only reliable methods for investigating the plasma parameters of such discharges are spectroscopic methods, which have a number of important advantages: contact-free measurement of the plasma parameters and no perturbation of the experimental object; high sensitivity, unique selectivity, and high rate of measurement of the plasma parameters; and, immunity to electromagnetic noise.

In the present paper we report experimental results of spectroscopic investigations of the plasma parameters of discharges of H^- ion sources with planotron and Penning electrode geometries[1–5] and analyze the plasma parameters for the same discharge current densities.

SOURCE DESIGN AND PARAMETERS

The electrodes of the planar (Fig. 1a) and Penning (PIG) (Fig. 1b) H^- sources were fabricated from ultrapure molybdenum. The external magnetic field **B** provided for oscillations of electrons in crossed fields $\mathbf{E}\times\mathbf{B}$ for the planotron and in parallel fields $\mathbf{E}\parallel\mathbf{B}$ for the PIG sources. Purified hydrogen was fed into the discharge by means of an electromagnetic valve in the pulsed regime, while cesium was fed at a constant rate from an external heated container holding cesiated pyrographite. The main parameters of the gas-discharge chambers of the planotron and Penning sources were as follows: discharge voltage 100–600 V, discharge current 10–150 A, pulse duration 850–35 μs, pulsed repetition frequency 1–10 Hz, magnetic field 0.5–1.5 kGs, cathode temperature 400–800 °C, anode temperature 200–400 °C, planotron cathode area 5 cm^2, PIG cathode area 2.5 cm^2, H_2 density up to $10^{16}\,\mathrm{cm}^{-3}$, and cesium density up to $10^{13}\,\mathrm{cm}^{-3}$.

Several stable regimes (I–IV) of the hydrogen–cesium discharge were observed. The regimes differed by the discharge voltage and the form of the current–voltage characteristic (Fig. 1e). The voltage (Fig. 1c) and discharge current (Fig. 1d) pulses were nearly square. Regime IV was obtained in a cesium-free, purely hydrogen discharge on electrodes with were degassed by preconditioning. The hydrogen–cesium regimes I–III were distinguished by the amount of cesium accumulated on the electrodes and were determined by the temperature of the electrodes and the rate of introduction of cesium into the discharge.

The discharge-plasma radiation passing through the 0.2×10 mm diagnostic slit in the anode chamber, a collimating slit in a screen located inside the vacuum space, and quartz glass in the wall of the vacuum chamber was focused by a condenser onto the input slit of the detecting apparatus. An anode chamber with four diagnostic slits I–IV (Fig. 1b) was used to investigate the spatial characteristics of the discharge. The radiation of the discharge plasma near the cathode zone was investigated through slit I. This radiation was also investigated using slit IV with an angle of acceptance of the plasma volume near the cathode surface of the order of 2°. The total angular acceptance of each slit was 12°. The total angular acceptance of each slit was equal to 12°. The

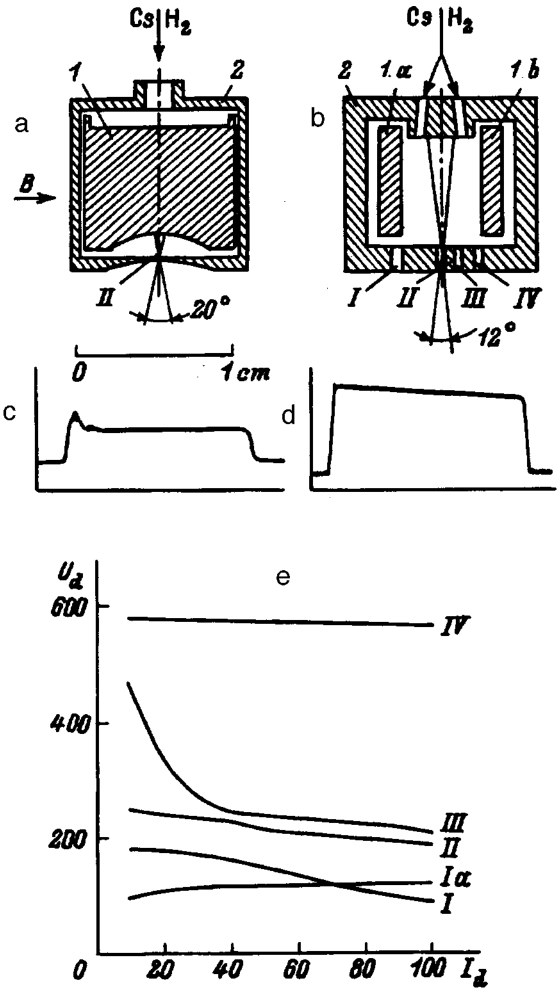

FIG. 1. Diagrams of the gas-discharge chambers of planotron (a) and Penning (b) sources of H^- ions (*1* — cathode, *2* — anode, *B* — magnetic field, I–IV — diagnostic slits); oscillograms of the voltage pulse (c) and current pulse (d) of discharges of the sources (the scale division along the ordinate is 200 (c) and 50 V (d)); dependences of the voltage on the discharge currents of surface-plasma sources of H^- ions in regimes I–IV (e).

sections of the plasma between the electrodes were observed through slits II and III. The use of a number of diagnostic slits and screens made it possible to determine the dynamics of the components and the main parameters of the PIG discharge plasma with spatial resolution in two coordinates. In the figures, the discharge voltage U_d and the discharge current I_d of the sources are given in units of V and A, respectively; the discharge current density J_d is given in A/cm^2; the radiation brightness B_λ is given in mW/sr·cm^2; the magnetic induction B is given in 10^{-2} T; and, the scale division along the abscissa in the oscillograms equals 200 μs.

DIAGNOSTICS APPARATUS

Four optical channels were used to detect the radiation from the discharge plasma. Spectrograms of the 230–900 nm plasma radiation, which were used to determine the elemental and charge composition of the plasma, were recorded with a STE-1 spectrograph. The absolute and relative intensities of the radiation in individual spectral lines of hydrogen, cesium, and molybdenum were recorded with a monochromator and photomultipliers, which were calibrated with a SI 8-200u tungsten lamp. The dynamics of the intensities of the spectral lines of the elements in the discharge was investigated on the edges of and during the discharge pulse. The general character of the time dependences of the spectral emission lines was determined using a RC filter, which was used to cut off the high-frequency noise of the discharge. A wider band detection system was used to investigate the discharge noise.

The temperature of the hydrogen atoms in the discharge was determined from the interferograms of the H_α line emission spectrum, and the electron density in the plasma was estimated from interferograms of the H_β line emission spectrum, which were recorded with Fabry–Perot interferometers. Interference filters with high transmittance were used to increase the detection efficiency. This made it possible to obtain interferograms of spectral lines in the time of one discharge pulse. The energy of the 400–800 nm radiation from the discharge plasma was measured with a RTR-30S thermocouple bolometer.

EXPERIMENTAL RESULTS AND DISCUSSION

1. Spectrum and energy of the radiation of the discharge plasmas

The energy of the visible-range (400–800 nm) radiation of the discharge plasma of the planotron in 4π steradians, measured in one 850 μs discharge pulse with discharge power 10 kW, was equal to 3.2 mJ.

The radiation spectra of the planotron and PIG discharge plasma were essentially identical. A small difference was observed only in the intensities of individual spectral lines. Besides the spectral lines of the main plasma components (hydrogen and cesium) a large number of lines of molybdenum sputtered from the cathode and lines of impurity oxygen were observed in the region of the spectrum 230–900 nm. The Balmer series (H_α, H_β, H_γ, H_δ) of hydrogen atoms, the lines of cesium atoms CsI, singly-charged cesium ions CsII (300–650 nm), doubly-charged cesium ions CsIII (250–290 nm), and the lines of molybdenum atoms MoI and singly-charged molybdenum ions MoII were also identified. For high discharge currents, transitions from highly excited levels ($n \sim 15$) of cesium atoms were detected. The line spectra of hydrogen molecules were also observed in the emission of the discharge plasma: Fulcher's α system (580–620 nm) and the electronic–rotational transitions $I\,^1\Pi_g \rightarrow B\,^1\Sigma_g^+$, $G\Sigma_g^+ \rightarrow B\,^1\Sigma_u^+$ in the region of the spectrum 420–490 nm.

2. Spectral line intensities of the plasma components

Hydrogen. In regime I with low molecular-hydrogen concentration in the gas-discharge chamber, the density of atomic hydrogen in the discharge at low discharge currents was proportional to the discharge current density. In this case, the brightnesses of the Balmer lines of hydrogen depended linearly on the discharge current (Figs. 2e and 2f) and on the density of the molecular hydrogen introduced into the discharge. The dependences of the brightnesses of the

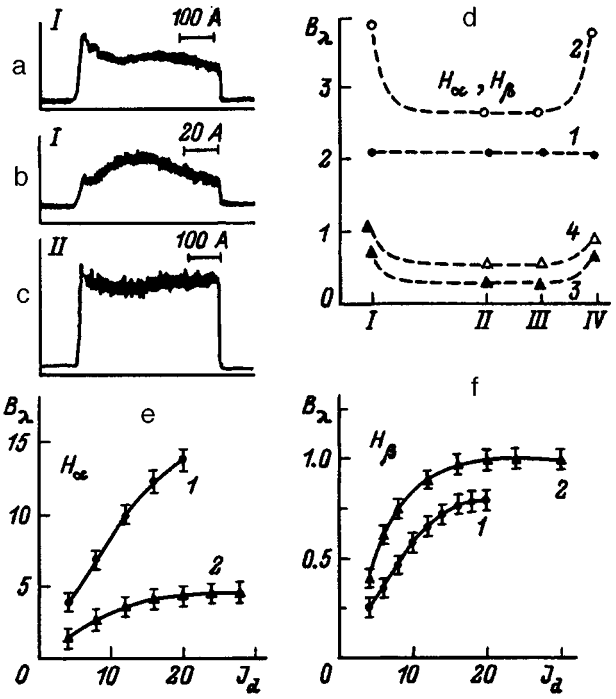

FIG. 2. Parameters of the Balmer lines of hydrogen in the plasma discharge: oscillograms of the emission intensities in regimes I (a,b) and II (c) (the scale division along the ordinate equals 5 (a), 1 (b), and 2 V (c)); spatial distribution of the brightnesses of the H_α line (*1,2*) and H_β line (*3,4*) over the diagnostic slits I–IV for PIG in regime I; brightnesses of the H_α line (e) and H_β line (f) as functions of the discharge current density for a planotron (*1*) and PIG (*2*).

spectral lines on the discharge current were recorded in a short time interval when the high currents were switched to low currents so that there would not be enough time for the temperature regime of the electrodes to change. The intensity readings were taken on the at the center of the pulse and adjusted for the spectral sensitivity of the photomultipliers. The kinetics of the radiation intensities of the Balmer lines at low discharge currents in all diagnostic slits was identical and repeated the shape of the gas pulse of molecular hydrogen (Fig. 2b). The spatial distribution of the brightness of the hydrogen emission lines was uniform along the entire discharge (Fig. 2d). It should be noted that the brightnesses in the plots were calculated taking account of the acceptance of the detection system for each diagnostic slit.

As the discharge current increased, the growth rates of the brightnesses of the Balmer lines slowed down and saturated at discharge current densities ~20 A/cm^2 (Fig. 2e and 2f). The magnitude of the saturation depended on the molecular hydrogen density and on the position of the hydrogen atoms in the discharge. The saturation of the H_α line brightness could be due both to partial displacement of hydrogen form the discharge zone into the side gaps of the discharge chamber and to the slowing down of the growth rates of the H^- and H^+ ion densities in the discharge plasma. Of the main channels leading to the formation of excited hydrogen atoms in a state with $n=3$, the process of mutual neutralization of the hydrogen ions $H^- + H^+ \rightarrow H^*(n=3) + H_0$ possesses the highest reaction rate $\langle \sigma v \rangle$. For this process, $\langle \sigma v \rangle = 3\times 10^{-7}\,\mathrm{cm}^3\cdot\mathrm{s}^{-1}$ (Ref. 6). Taking into account the probe measurements of the H^- and H^+ ion densities in the discharge of a surface-plasma source of H^- ions,[7] the effective reaction rate will be $N_{H^-}\cdot N_{H^+}\cdot\langle \sigma v \rangle = 10^{20}\,\mathrm{cm}^3\cdot\mathrm{s}^{-1}$. The kinetics of the Balmer line intensities at high discharge currents was essentially the same in regimes I and Ia (Fig. 2a), except for the pulse edges. The surge in the radiation intensity on the pulse edge in regime I attests to a relatively high contribution of atomic hydrogen desorbed from the surface of the electrodes in the gas-discharge chamber.

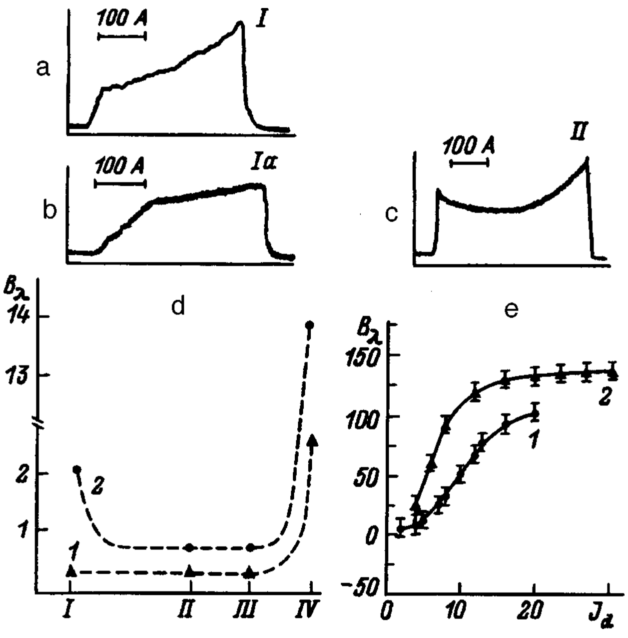

FIG. 3. Parameters of the CsI (852.1 nm) emission line of the discharge plasma: oscillograms of the emission intensities in regimes I (a), Ia (b), and II (c) (scale division along the ordinate 0.5 (a), 5 (b), and 0.2 V (c)); spatial distribution of the emission brightnesses over diagnostic slits I–IV in regime I for PIG; emission brightnesses versus the discharge current density J_d for a planotron (*1*) and PIG (*2*) in regime I (e).

In regimes II–IV, despite the high discharge voltage and higher hydrogen density in the gas-discharge chamber, the Balmer line intensities had approximately the same values as for the corresponding discharge currents of regime I, and their kinetics repeated the shape of the discharge current (Fig. 2c). The rates of growth and decay of the Balmer line intensities (leading and trailing edges of the radiation pulses) in these regimes were identical to the rates of growth and decay of the current in the discharge pulse.

Cesium. The intensities and kinetics of the emission lines of cesium atoms and ions depended considerably on the introduction of cesium into the discharge and the thermal regime of the electrodes. The atomic cesium line CsI (852.1 nm) is the brightest line in the emission spectrum at 230–900 nm. This line has the highest transition oscillator strength ($f=0.8$). The dependences of the CsI (852.1 nm) emission line brightnesses on the discharge current density in regime I for the planotron and PIG are presented in Fig. 3e. In this case, the readings of the radiation intensity was taken at the maximum at the end of the pulse. For very low discharge current densities, the growth of the electron and cesium densities in the volume of the discharge gives rise to a quadratic increase of the amplitude of the line brightness. As the discharge current increased, in a small range of values of its density the emission line brightness was observed to grow linearly. At high discharge current densities the rate of growth of the emission line brightness dropped sharply as a result of an increase in the cesium ionization rate and decrease in the cesium density in the interelectrode gap.

At high discharge current densities the intensities of the CsI (852.1 nm) line and of other cesium lines increased toward the end of the discharge pulse (Fig. 3a and 3b). The relatively long duration of the leading edges of the emission pulses in regimes I and Ia is due to the additional release of cesium into the discharge as a result of bombardment of the electrodes, whose sputtering coefficient in regimes I is decreased due to the increase in the coefficient of secondary ion-electron emission. The pulsed overheating of the surface by the discharge and the release of cesium from the colder, nonworking sections of the interelectrode gap increase the equilibrium density and lead to the accumulation of cesium in the discharge. The kinetics of the cesium line intensities had a completely different character for low discharge currents and in regimes II and III for high currents (Fig. 3c). The surge of the radiation intensity on the leading edge of the pulse is due to the excitation of cesium in the volume of the discharge gap, and the increase in intensity toward the end of the pulse is due to the release of cesium into the discharge as a result of the bombardment of the electrodes. The rates of growth of the cesium line intensities at the start of the radiation pulses increased substantially in regimes II and III as compared with regime I, but they were much lower than that of the Balmer lines. At the same time, the dropoff of the cesium emission line intensities at the end of the pulses was much more rapid. In regimes II and III, with hotter electrodes, the initial cesium density in the volume of the gas-discharge chamber was higher than in regime I, while the cesium concentration at the electrodes was low. The sputtering coefficient of cesium in this case is much higher, as a result of which the cesium emission line intensities reached a relatively high stationary level more rapidly.

The spatial distribution of the CsI (852.1 nm) emission line brightnesses over the diagnostic slits I–IV was strongly asymmetric (Fig. 3d), especially in the presence of high discharge current densities. This attests unequivocally to the fact that the emission surface of the cathode and the adjoining thin layer of plasma, the radiation from both of which is detected only through the slit IV, makes the determining the contribution to the emission in this line.

In contrast, for the CsI (455.5 nm) emission line the spatial distribution of the brightness over the diagnostic slits I–IV (Fig. 4c) was substantially symmetric. This means that the radiation from the interior of the discharge and not from the surface of the electrodes makes the main contribution to the CsI (455.5 nm) emission line. Moreover, a substantial difference was observed in the kinetics of the intensities of these two lines in the discharge near the cathodes. The intensity of the CsI (455.5 nm) emission line already became saturated at the start of the discharge pulse (Fig. 4b). At the same time, the kinetics of the intensities of these lines during the discharge pulse was qualitatively the same at the center of the discharge (Fig. 4a). The CsI (455.5 nm) emission line brightnesses saturated at much lower discharge current densities (Fig. 4d).

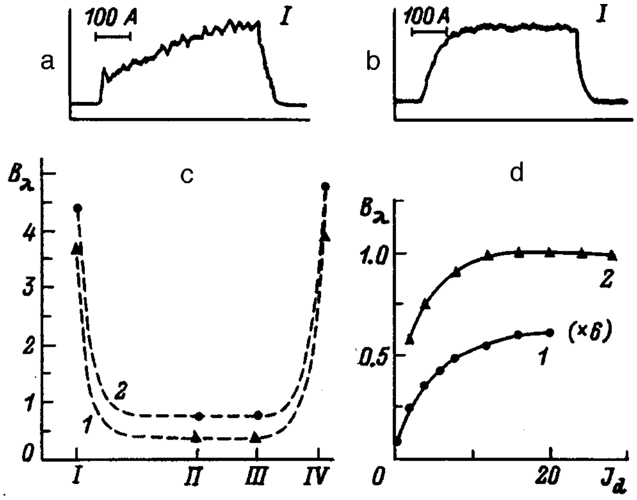

FIG. 4. Parameters of the CsI (455.5 nm) emission line of a discharge plasma: a,b — oscillograms of the emission intensities in regime I (scale division along the ordinate 0.05 (a) and 0.01 V (b)); c — spatial distribution of the emission brightnesses over the diagnostic slits I–IV in regime I for PIG; d — emission brightnesses versus discharge current density for a planotron (*1*) and PIG (*2*) in regime I.

Cesium ions. The kinetics of the cesium ion line intensities in regime I (Fig. 5a) was completely identical to the kinetics of the cesium atom emission line intensities. The two kinetics differed substantially in regime II (Fig. 5c). An increase in the electron and cesium ion densities in the dis-

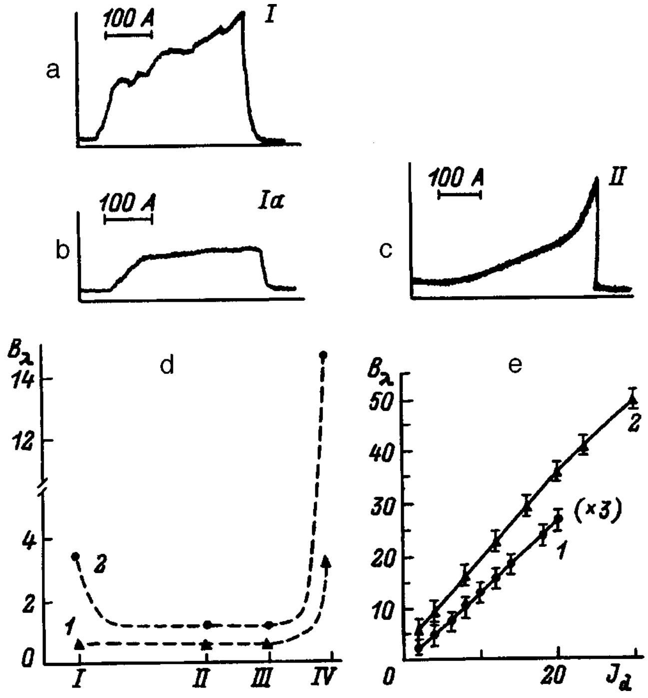

FIG. 5. Parameters of the CsII (460.4 nm) emission line of the discharge plasma: oscillograms of the emission intensities in regimes I (a), Ia (b), and II (c) (scale division along the ordinate 2 (a) , 10 (b), and 0.2 V (c)); d — spatial distribution of the emission brightnesses over the diagnostic slits I–IV in regime I for PIG; e — emission brightnesses versus the discharge current density for a planotron (*1*) and PIG (*2*) in regime I.

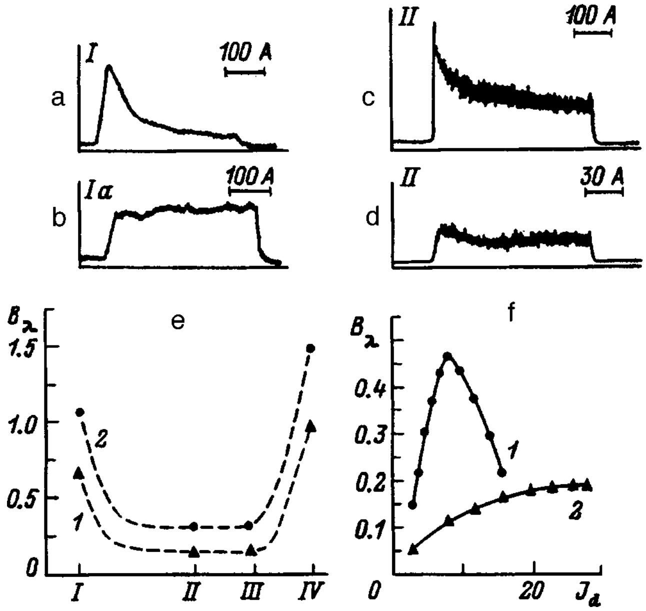

FIG. 6. Parameters of the MoI (553.3 nm) emission line of the discharge plasma: oscillograms of the emission intensities in regimes I (a), Ia (b), and II (c,d) (scale division along the ordinate 0.1 (a), 0.05 (b), and 0.01 V (c)); e — spatial distribution of the emission brightnesses over the diagnostic slits I–IV in regime I for PIG; f — emission brightnesses versus the discharge current density for a planotron (*1*) and PIG (*2*) in regime I (e).

charge gives rise to rapid quadratic growth in the amplitudes of the cesium ion line intensities with increasing discharge current density. The brightnesses of all cesium lines increased substantially in regime Ia with forced introduction of cesium into the discharge, when the discharge became essentially a cesium–hydrogen discharge. This gave rise to small changes in the kinetics of the cesium line intensities (Fig. 5b and 3b).

Just as for the CsI (852.1 nm) line, the spatial distribution of the brightnesses of the cesium ion lines over the diagnostic slits I–IV was sharply asymmetric (Fig. 5d). This attests to the fact that cesium atoms and ions are pent up near the cathode surface. The brightnesses of the cesium ion emission lines near the cathodes, in contrast to the lines of cesium atoms, continued to grow even at high discharge current densities (up to ~ 30 A/cm^2) (Fig. 5e), and at the center of the discharge (on slit II) they reached saturation at discharge current densities of the order of 10 A/cm^2. Comparing the intensities of the emission lines of cesium atoms and ions shows that in regimes I–III cesium is predominantly ionized during the discharge pulse.

Molybdenum. Of the plasma spectral lines investigated, the lines of molybdenum atoms were the weakest. In regime I molybdenum is knocked out of the electrodes of the surface-plasma sources predominantly at the start of the gas-discharge pulse (Fig. 6a). In the more cesiated regime Ia the emission and excitation of molybdenum occurred more uniformly over the entire discharge pulse (Fig. 6b). The decrease in the sputtering of molybdenum and saturation of the intensity of the molybdenum radiation within 200 μs after the start of the pulse are due to the increase in the degree of coverage of the cathodes by cesium up to a dynamic-equilibrium value determined by the cesium ion current on the cathode and the sputtering coefficient of cesium. In regimes II–IV, with a lower cesium density on the electrodes, the molybdenum emission line intensities increased substantially (Fig. 6c and 6d), and the kinetics of its line intensities was similar to regime I.

The spatial distribution of the brightnesses of the emission lines of molybdenum atoms over the diagnostic slits I–IV was symmetric (Fig. 6e). The brightnesses of the emission lines of molybdenum atoms near the cathodes increased by a factor of two or three, attesting to the fact that molybdenum atoms are pent up near the cathodes. In regime I, with low discharge current densities, the stationary intensity of the emission of molybdenum atoms increases (Fig. 6f) on account of an increase in the electron density in the discharge. As the discharge current density increases, blocking of the cathode by cesium in the planotron increases and the intensities of the molybdenum emission lines in the planotron (curve *1*) start to decrease with increasing discharge current, while in PIG the brightnesses of the molybdenum emission lines near the cathodes (curve *2*) continue to grow. Rapid saturation of the molybdenum emission line brightnesses occurred at the center of the discharge gap at quite low discharge current densities on account of the low density of molybdenum.

3. Fluctuations of the spectral emission line intensities

Fluctuations of the cesium ion current on the cathode cause local variations of the cesium coverage and emissivity of the cathode. This could be the reason for the low-frequency oscillations of the electron density in the discharge and the associated fluctuations in the intensities of the spectral emission lines (Fig. 7). As the discharge current increased, the frequency and level of low-frequency ($\sim 10^6$ Hz) fluctuations and noise increased. However, the relative level of the fluctuations, determined as the ratio of the rms value of the fluctuations to the average value of the intensity of the spectral emission lines, as a rule, decreased as the discharge current density increased. As the magnetic field of the sources increased, and also as the amount of hydrogen and cesium introduced into the discharge decreased, the level of low-frequency fluctuations increased (Fig. 7e). In regimes II and III the level of the intensity fluctuations was much higher (Fig. 7a and 7b) than in regime I (Fig. 7c and 7d). Regime I in PIG depended strongly on the cesium density in the discharge. A negligible decrease in the amount of cesium introduced into the discharge resulted in the appearance of different instabilities in the discharge and a transition from a noisy to noise-free mode during the discharge pulse. This produced a qualitative change in the kinetics of the emission intensities of all spectral lines of the plasma elements with the exception of the molybdenum lines.

A clear correlation was observed between the oscillations of the intensities of the emission lines and the oscillations of the current (voltage) in the discharge of the source. Oscillations of the discharge current and voltage of the source unequivocally lead to fluctuations of the emittance of the beam of negative hydrogen ions extracted from the dis-

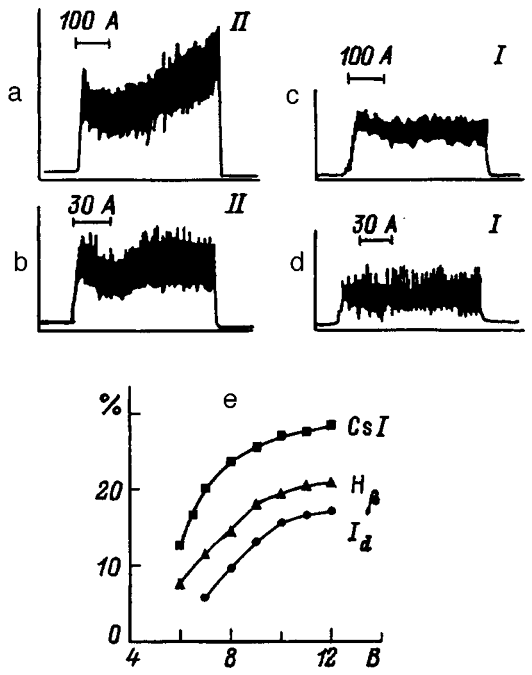

FIG. 7. Parameters of low-frequency fluctuations of the CsI and H_β spectral line intensities and the discharge current: oscillograms of the CsI (852.1 nm) emission line intensity (a,b) in regime II and the H_β line intensity (c,d) in regime I, obtained using a wide-band detection system (scale division along the ordinate 0.2 (a, b) and 1 V (c,d)); e — relative level of low-frequency fluctuations and noise (%) of CsI and H_β spectral lines and the discharge current versus the external magnetic induction.

charge by a high-voltage field (Special Report No. 7205, Siberian Physicotechnical Institute, 1987). For discharge current density of the order of 30 A/cm^2 in the source, increasing the magnetic field from 6 to 9×10^{-2} T increased the frequency of the oscillations of the discharge current from 0.5 to 10 MHz and increased the H^- beam emittance by almost a factor of two (Special Report No. 7504, Siberian Physicotechnical Institute, 1989). The increase in the emittance of the ion beam could also be due in part to an increase in the nonuniformities of the magnetic field of the source and of the space charge of the beam.

4. Temperature of hydrogen atoms in the discharge

The temperature T_{H_0} of hydrogen atoms in the discharge plasma of the source was determined from the Doppler broadening of the H_α line

$$T_{H_0}[\mathrm{eV}]=4\times10^2[\Delta\lambda_D(\mathrm{nm})]^2. \tag{1}$$

The Doppler width $\Delta\lambda_D$ of the spectrum of the H_α line was found from interferograms of the emission spectra. Typical interferograms of the H_α line emission spectra obtained for a planotron (a) and PIG (b) during the time of one discharge pulse are presented in Figs. 8a and 8b. Microphotometry of the interferograms did not show a fine structure of the H_α line. This fine structure can be neglected, if the the hydrogen atoms are excited mainly due to the dissociation of hydrogen molecules by electron impact, when two thirds of the energy is radiated in the transition $3^2D_{5/2}\rightarrow2^2P_{3/2}$ (Ref. 8). In a pulsed discharge an appreciable fraction of the hydrogen atoms is formed as a result of desorption from the surface of the electrodes, which can result in broadening of the H_α line, including that due to fine structure.

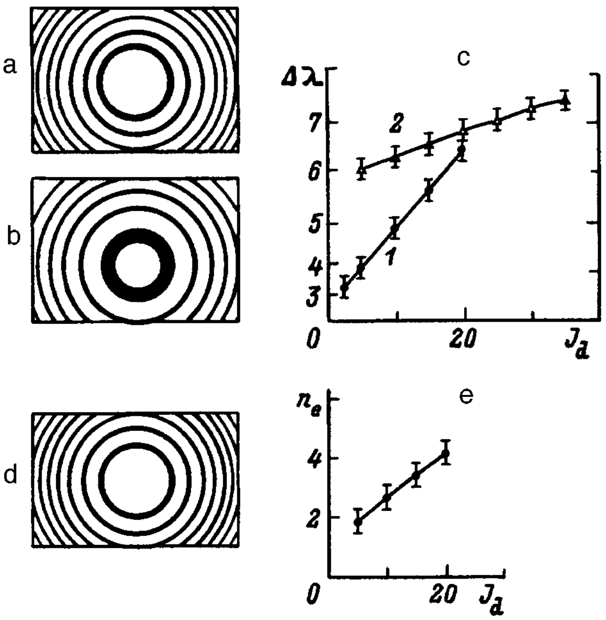

FIG. 8. Interferograms of the H_α line spectra for a planotron (a) and PIG (b); temperature of hydrogen atoms T_{H_0} (eV) and spectrum width $\Delta\lambda$ (nm $\cdot10^{-2}$) of the H_α line radiation versus discharge current density of a planotron (*1*) and PIG (*2*) in regime I (c); interferogram of the H_β line emission spectrum for a planotron in regime I (d); electron density n_e (cm^{-3}, $\times10^{13}$) in a discharge of a planotron versus the current density in regime I (e).

The temperature of the hydrogen atoms in the discharge plasma and the emittances of the H^- beams in the planotron (*1*) and PIG (*2*) increased linearly with the discharge current density (Fig. 8c). Up to source discharge current densities $\leqslant25$ A/cm^2 the temperature of the hydrogen atoms and the emittance of the H^- beam extracted from the source were lower in the planotron than in the PIG (Special Report No. 7504, Siberian Physicotechnical Institute, 1989).

5. Electron temperature and density in the discharge plasma

The electron density in the discharge plasma was estimated according to the Stark broadening of the H_β line. The typical interferogram of the spectrum of this line is shown in Fig. 8d. The electron density in the plasma was a linear function of the discharge current density, and at had the value $n_e=4\times10^{13}$ cm^{-3} at 20 A/cm^2 (Fig. 8e).

The electron temperature in the discharge was found from the ratio of the Balmer line intensities in the approximation of local thermodynamic equilibrium and a Maxwellian electron energy distribution. The temperature calculated by the method of Ref. 9 was 0.4 eV. Anomalously low values of the electron temperature were also obtained from the intensity distribution in the spectrum of the recombination continuum on the $5d$ level (550–590 nm).[10] The ''cesium'' temperature of the electrons was $\sim$0.1 eV and was essentially independent of the cesium density.

The anomalously low values obtained for the electron temperature show that the model of local thermodynamic equilibrium has limited applications under the conditions of a high-current glow discharge with a nonequilibrium processes at the electrode surfaces. In this case, together with the low-energy group of Maxwellized electrons, it is necessary to take account of the role of the group of fast primary electrons.

CONCLUSIONS

Through the use of spectroscopic methods we were able to perform nonperturbative monitoring and to measure reliably all the basic parameters of the discharge plasma of surface-plasma sources of H^- ions. The results obtained make it possible to construct real theoretical models of these sources. It was shown that the planotron source, which has the lowest temperature of the hydrogen atoms in the discharge plasma and the smallest divergence of the H^- ion beam extracted from the source, is the preferred source for obtaining accelerated H^- beams with the lowest possible emittance.

[1] V. V. Antsiferov, V. V. Beskorovaĭnyĭ, and Yu. I. Bel'chenko, Preprint No. 88–116 iR, Institute of Nuclear Physics, Siberian Branch of the Soviet Academy of Sciences, Novosibirsk, 1988, 21 pp.

[2] V. V. Antsiferov, V. V. Beskorovaynyy, and Yu. I. Belchenko, AIP Con. Proc., No. 10, (1990) pp. 427–447.

[3] V. V. Antsiferov and V. V. Beskorovaĭnyĭ, Zh. Tekh. Fiz. **63**(4), 50 (1993) [Sov. Phys. Tech. Phys. **38**, 291 (1993)].

[4] V. V. Antsiferov and V. V. Beskorovaĭnyĭ, Zh. Tekh. Fiz. **63**(5), 41 (1993) [Sov. Phys. Tech. Phys. **38**, 390 (1993)].

[5] V. V. Antsiferov, Preprint No. 93–49 iR, Institute of Nuclear Physics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, 1993, 23 pp.

[6] R. K. Janev, Report PPPL-TM-368, 20 pp.

[7] Yu. I. Bel'chenko, G. I. Dimov, V. G. Dudnikov, and A. S. Kupriyanov, Preprint No. 88–14, Institute of Nuclear Physics, Siberian Branch of the Soviet Academy of Sciences, Novosibirsk, 1988, 36 pp.

[8] G. N. Polyakova, A. I. Ranyuk, and V. F. Erko, Zh. Eksp. Teor. Fiz. **73**, 2131 (1977) [Sov. Phys. JETP **46**, 1117 (1977)].

[9] H. V. Smith and P. Allison, AIP Conf. Proc., 1987, No. 58, pp. 181–192.

[10] F. G. Baksht *et al.*, *Thermionic Converters and Low-Temperature Plasma*, edited by L. K. Hansen, Technical Information Center, U.S. Dept. of Energy, Springfield, Virginia, 1978; Russian original, edited by B. Ya. Moĭzhes and G. E. Pikus, Nauka, Moscow, 1973, 480 pp.

Translated by M. E. Alferieff

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Stress state of a matrix and an inclusion consisting of a twisted multifilament superconductor with a normally-conducting core in the presence of an interaction of the longitudinal current with a transverse magnetic field

E. A. Devyatkin

Institute of Mechanics Problems, Russian Academy of Sciences, 117526 Moscow, Russia

(Submitted September 14, 1997)
Zh. Tekh. Fiz. **68**, 39–42 (October 1998)

A study is made of the stress state arising in an infinite matrix into which is "potted" a long cylindrical inclusion consisting of a twisted multifilament composite superconductor with a core of normal metal as a result of the interaction of the longitudinal component of the transport current with a uniform external magnetic field. It is shown that the approximation of a homogeneous conductor with average elastic properties can be used to calculate the macroscopic stresses in a NbTi–Cu wire embedded in an epoxy matrix. © *1998 American Institute of Physics.* [S1063-7842(98)00610-2]

Substantial stresses and strains can arise in the structural components of superconducting magnet systems, resulting in mechanical dissipative processes, which are undesirable at low temperatures, and/or a change in the superconducting properties of the wires.[1–4] In this connection, it is of interest to study the stress–strain state of superconductors and the materials surrounding them. The stress state arising in an infinite matrix and a long macroscopically uniform cylindrical inclusion consisting of a twisted multifilament composite superconductor (TMCS) carrying a transport current as a result of the interaction of the longitudinal component of the current with a uniform transverse magnetic field was studied in Refs. 5 and 6. The serially produced TMCS wires that are widely used in practice can have a normally conducting core whose elastic properties are different from those of the annular region containing the superconductor. A similar two-dimensional quasistatic problem for such a wire with a critical transport current is studied below. It is assumed that the macroscopic approach can be used to describe the physicomechanical properties of the superconducting region, that there is no macroscopic screening of the external field, and that the superconducting properties of the wire do not depend on the stresses and strains in it.

Consider an inclusion of unit radius "potted" in an infinite nonferromagnetic matrix and consisting of a TMCS in which a longitudinal transport current I_s flows in the region $\delta_s < r < 1$ containing the superconducting filaments (Fig. 1). We shall assume that the matrix, the superconducting region of the inclusion, and also the core of the inclusion $0 \leqslant r < \delta_s$ are linearly elastic, homogeneous, and isotropic, and their Young's moduli and Poisson's ratios are, respectively, $E^{(n)}$ and $\nu^{(n)}$ $(n = m, j, i)$. A Lorentz force $\mathbf{f} = \mathbf{j}_s \times \mathbf{B}_\perp$ acts on a region with a longitudinal current density $\mathbf{j}_s$ in an external uniform transverse field with induction $\mathbf{B}_\perp$. We neglect the interaction of the tangential component of the transport current with the magnetic field and the distortion produced in the distribution of $\mathbf{j}_s$ by the twisting of the wire.

In the region $\delta_s < r < 1$ the equations of equilibrium and compatibility have the form[7]

$$\sigma^{(j)}_{xx,x} + \sigma^{(j)}_{xy,y} = 0, \quad \sigma^{(j)}_{yy,y} + \sigma^{(j)}_{xy,x} + 1 = 0,$$

$$\Delta(\sigma^{(j)}_{xx} + \sigma^{(j)}_{yy}) = 0. \tag{1}$$

Here and below all quantities with the dimension of pressure are normalized to $2B_j^{(s)}B_\perp/[\mu_0(1-\delta_s^2)]$; $B_j^{(s)}$ is the magnetic field generated on the surface of the inclusion by the longitudinal current; μ_0 is the magnetic permeability of free space; and, Δ is the Laplacian operator. The stresses $\sigma^{(i)}_{kl}$ in the core of the wire and the stresses $\sigma^{(m)}_{kl}$ $(k, l = x, y)$ in the matrix surrounding the wire must satisfy similar homogeneous equations. We assume that far from the inclusion the matrix is load-free ($\sigma^{(m)}_{kl} \to 0$ as $t \to \infty$). From the condition that the media are joined together it follows that at the interfaces of the regions there are no jumps of the displacements or in stresses normal σ_n and tangential σ_τ to the boundaries:

$$[\mathbf{u}] = 0, \quad [\sigma_n] = [\sigma_\tau] = 0 \quad \text{for} \quad r = \delta_s, 1. \tag{2}$$

We seek the solution of the problem in the same form as for a homogeneous inclusion[6]

$$\begin{aligned}
&\sigma^{(i)}_{rr} = ar\sin\theta, \quad \sigma^{(m)}_{rr} = [-gr^{-3} + (h+q)r^{-1}]\sin\theta,\\
&\sigma^{(i)}_{\theta\theta} = 3ar\sin\theta, \quad \sigma^{(m)}_{\theta\theta} = (gr^{-3} + hr^{-1})\sin\theta,\\
&\sigma^{(i)}_{r\theta} = -ar\cos\theta, \quad \sigma^{(m)}_{r\theta} = (gr^{-3} - hr^{-1})\cos\theta,\\
&\sigma^{(j)}_{rr} = [(b-1)r - cr^{-3} + (d+e)r^{-1}]\sin\theta,\\
&\sigma^{(j)}_{\theta\theta} = [(3b-1)r + cr^{-3} + dr^{-1}]\sin\theta,\\
&\sigma^{(j)}_{r\theta} = -(br - cr^{-3} + dr^{-1})\cos\theta.
\end{aligned} \tag{3}$$

In the case of a plane strain $\sigma^{(n)}_{zz} = \nu^{(n)}(\sigma^{(n)}_{rr} + \sigma^{(n)}_{\theta\theta})$.

In the limit $r \to \infty$ the stresses in the matrix must be the same as in the case of the well-known solution for a concentrated force acting on a plane (see, for example, Ref. 7).

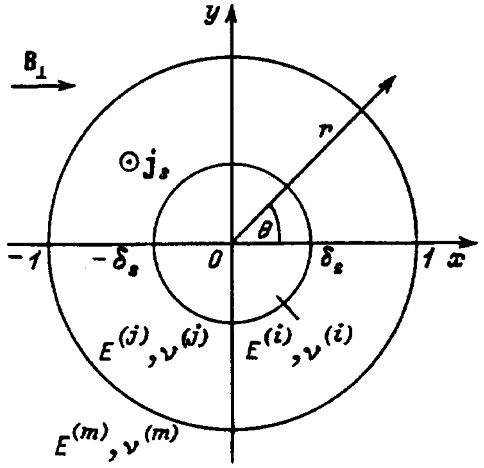

FIG. 1. Inclusion consisting of a twisted multifilament composite superconductor carrying a transport current in a transverse magnetic field.

Since in the case at hand the Lorentz force acts in the direction of the $0y$ axis and equals $I_s B_\perp$ per unit length of the inclusion, we obtain

$$h=\frac{1-\delta_s^2}{4}\,\frac{1-2\nu^{(m)}}{1-\nu^{(m)}},\quad q=\delta_s^2-1. \tag{4}$$

From continuity of the stresses (2) we have

$$a-b+\frac{c}{\delta_s^4}-\frac{d}{\delta_s^2}=0,\quad e=\delta_s^2,\quad b-c+d+g=h. \tag{5}$$

The strains $\varepsilon_{kl}^{(n)}$ $(k,l=r,\theta)$ are related with the displacements $\mathbf{u}^{(n)}$ by the relations

$$\varepsilon_{rr}^{(n)}=u_{r,r}^{(n)},\quad \varepsilon_{\theta\theta}^{(n)}=\frac{u_r^{(n)}+u_{\theta,\theta}^{(n)}}{r},$$

$$\varepsilon_{r\theta}^{(n)}=\frac{u_{r,\theta}^{(n)}-u_\theta^{(n)}}{r}+u_{\theta,r}^{(n)}. \tag{6}$$

Determining the strains corresponding to the stresses (3) from Hooke's law, and integrating Eqs. (6) in each of the regions considered, we find

$$u_r^{(i)}=\frac{1+\nu^{(i)}}{2E^{(i)}}[p+4\nu^{(i)}ar^2]\sin\theta,$$

$$u_\theta^{(i)}=\frac{1+\nu^{(i)}}{2E^{(i)}}[p-(5-4\nu^{(i)})ar^2]\cos\theta,$$

$$u_r^{(j)}=\frac{1+\nu^{(j)}}{2E^{(j)}}\left\{s+[(1-4\nu^{(j)})b-1+2\nu^{(j)}]r^2+cr^{-2}\right.$$
$$\left.+\frac{e}{2}\,\frac{3-4\nu^{(j)}}{1-\nu^{(j)}}\ln r\right\}\sin\theta,$$

$$u_\theta^{(j)}=\frac{1+\nu^{(j)}}{2E^{(j)}}\left\{s-[(5-4\nu^{(j)})b-1+2\nu^{(j)}]r^2-cr^{-2}\right.$$
$$\left.+\frac{e}{2}\,\frac{1+(3-4\nu^{(j)})\ln r}{1-\nu^{(j)}}\right\}\cos\theta,$$

$$u_r^{(m)}=\frac{1+\nu^{(m)}}{2E^{(m)}}\left(gr^{-2}-\frac{1-\delta_s^2}{2}\,\frac{3-4\nu^{(m)}}{1-\nu^{(m)}}\ln r\right)\sin\theta,$$

$$u_\theta^{(m)}=-\frac{1+\nu^{(m)}}{2E^{(m)}}\left[gr^{-2}+\frac{1-\delta_s^2}{2}\,\frac{1+(3-4\nu^{(m)})\ln r}{1-\nu^{(m)}}\right]\cos\theta. \tag{7}$$

Here s and p are integration constants. In addition, it follows from the single-valuedness of the displacements that

$$d=-\frac{e}{4}\,\frac{1-2\nu^{(j)}}{1-\nu^{(j)}}. \tag{8}$$

Since only dimensionless quantities are employed in this paper, to obtain dimensional displacements the right-hand sides of Eqs. (7) must be multiplied by $2B_j^{(s)}B_\perp R/[\mu_0(1-\delta_s^2)]=j_sB_\perp R^2$, where R is the radius of the wire. Using the continuity of the displacements at the interfaces and taking account of Eqs. (4) and (5), we find the unknown coefficients

$$a=\frac{\beta}{2\lambda_i},\quad b=\frac{\beta+3-4\nu^{(j)}}{8(1-\nu^{(j)})},$$

$$g=-\frac{1}{4}\left(\frac{2\beta}{1-\lambda_m}+\frac{1-\delta_s^2}{1-\nu^{(m)}}\right),$$

$$c=\frac{\delta_s^4}{8(1-\nu^{(j)})}\left[1+\beta\left(1-4\,\frac{1-\nu^{(j)}}{\lambda_i}\right)\right], \tag{9}$$

where

$$\lambda_i=1+(3-4\nu^{(i)})\,\frac{1+\nu^{(i)}}{1+\nu^{(j)}}\,\frac{E^{(j)}}{E^{(i)}},\quad \lambda_m=\frac{1+\nu^{(m)}}{1+\nu^{(j)}}\,\frac{E^{(j)}}{E^{(m)}},$$

$$\beta=(1-\delta_s^2)^2\left[1-\delta_s^4-4(1-\nu^{(j)})\left(\frac{1}{1-\lambda_m}-\frac{\delta_s^4}{\lambda_i}\right)\right]^{-1}.$$

Here we present only the constants required for calculating the stresses. For $E^{(i)}=E^{(j)}$ and $\nu^{(i)}=\nu^{(j)}$ the solution (3)–(5) and (7)–(9) is identical to that obtained in Ref. 6 for a homogeneous inclusion. In the limit $\lambda_m\to 1$ we have $a\to 0$. In this case the region $0\leqslant r\leqslant\delta_s$ is stress- and strain-free for any admissible values of its elastic parameters and, as one can see from Eqs. (7), it moves as a rigid body. The solution for $\mathbf{u}^{(m)}$ in the limit $r\to\infty$ is identical to the well-known solution for displacements produced by a concentrated force,[7] but it cannot be used directly to calculate them in the entire plane since in the limit $r\to\infty$ we have $\mathbf{u}^{(m)}\to\infty$. To calculate the displacements in the case, for example, that they are absent at distances $R_m\gg 1$ from the inclusion, a term linear in r must be added to the solution (3) for the stresses in the matrix and the integration constant must be retained in the solution for $\mathbf{u}^{(m)}$, as done for other regions, and the con-

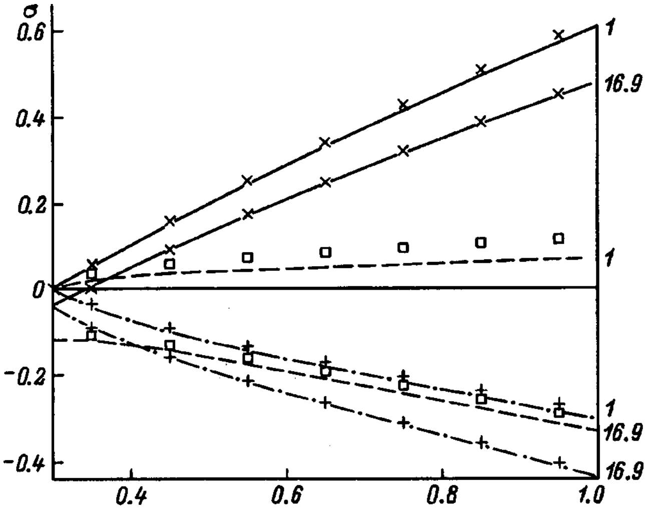

FIG. 2. Stress distribution in the superconducting region of a conductor consisting of a twisted multifilament superconductor.

dition $\mathbf{u}^{(m)}=0$ must hold at $r=R_m$. The presence of integration constants does not change the strains and therefore the stresses also. In the limit $R_m \to \infty$ the solution for the stresses in such a problem is identical to the solution (3) and the coefficients appearing in it equal those found above. Setting $\delta_s \ll 1$ and passing to the limit $E^{(i)} \to 0$ we obtain a solution of the problem for a wire containing a narrow coaxial channel through which coolant can be pumped. The coefficient $\beta \approx [1-4(1-\nu^{(j)})/(1-\lambda_m)]^{-1}$ is essentially identical to the analogous coefficient for a continuous conductor even in the presence of a soft matrix ($E^{(j)} \gg E^{(m)}$, i.e., $\lambda_m \gg 1$), if $\delta_s^4 \ll \lambda_i/|1-\lambda_m|$ and $\delta_s^2 \ll 1$.

As an example, let us consider the stress distribution in a NbTi–Cu conductor which is surrounded by an epoxy matrix and where Young's moduli of the central normally conducting and superconducting regions can differ appreciably. On the basis of the values given in Refs. 1 and 4 for the elastic parameters of the materials at liquid-helium temperatures, we take $E_i=1.32\times10^{11}$ Pa, $E_j=9.7\times10^{10}$ Pa, $E_m=8\times10^{9}$ Pa, $\nu_i=\nu_j=0.33$, $\nu_m=0.36$ ($\lambda_i=2.24$, $\lambda_m=12.4$). If such a conductor is in a copper matrix, we have $\lambda_m=0.74$. For a macroscopically homogeneous conductor ($\lambda_i=2.68$) with Young's modulus close to that of copper surrounded by an epoxy matrix, we obtain $\lambda_m=16.9$. If the elastic properties of all materials are identical, then $\lambda_m=1$. It is clear from the symmetry of the problem that the stresses are maximum at $\theta=\pm\pi/2$ or $\theta=0, \pi$, pure shear obtaining in the latter case. Figure 2 displays the stress distributions in the superconducting region of a homogeneous conductor with $\delta_s=0.3$ for the cases $\lambda_m=1$ and 16.9. The solid, dashed, and dot-dashed curves show, respectively, the radial distributions of the following components of the stresses: $\sigma_{rr}^{(j)}(r,-\pi/2)$, $\sigma_{\theta\theta}^{(j)}(r,-\pi/2)$, and $\sigma_{r\theta}^{(j)}(r,0)$. The symbols in Fig. 2 represent the stresses in an inhomogeneous conductor with $\lambda_m=0.74$ and 12.4 and correspond to the curves closest to them. One can see that when the Young's moduli of the superconducting and normally conducting regions of the wire surrounded by an epoxy matrix are appreciably different, the stress distribution in it does not differ much from the case of a homogeneous conductor. As the stiffness of the matrix increases, the difference in the $\sigma_{\theta\theta}^{(j)}$ component of the stresses for these cases can become substantial.

In summary, the approximation of a homogeneous conductor with average elastic properties can be used to calculate the macroscopic stresses in a NbTi–Cu wire with a near-critical transport current placed in a soft matrix. The error due to this approximation decreases as the volume content of the superconductor in the wire decreases and when the softer copper is used as its matrix (in Ref. 1, the value 1.24×10^{11} Pa is given for its Young's modulus at 4.2 K).

Taking for estimates $j_s=1.5\times10^{9}$ $\mathrm{A\cdot m^{-2}}$, $R=7.5\times10^{-4}$ m, and $\delta_s=0.3$ we obtain that the maximum shear stresses arising in the matrix near the surface of the wire reach the maximum shear strength of the epoxy resin, equal to $1-6\times10^{6}$ Pa at 4.2 K[1] in 2.1–12.5 T fields. The maximum tensile stresses in the wire are much lower than the yield stresses of most metals at liquid-helium temperatures. For example, in Nb_3Sn wires the critical current changes little when the stresses in them change by about 10^{7} Pa (see Refs. 2 and 3 and the literature cited therein). However, such stresses produced in a wire by Lorentz forces in the presence of an 8–10 T external magnetic field and external transverse compressive loads $>5\times10^{7}$ Pa can produce appreciable changes in the transport current.[3]

The solution obtained makes it possible to calculate the stresses produced by Lorentz forces in the components of structures containing circular conductors consisting of TMCS with a weak elastic interaction between them. When the distance between the surfaces of the conductors equals only several radii of the conductors, such an interaction can be taken into account as an action produced in the matrix by concentrated forces.[6] This is because the terms in the solution for the stresses in the matrix decay rapidly, as $\sim r^{-3}$, which makes it possible to treat them as concentrated sources even at comparatively small distances away from the conductors. When all materials have close elastic properties, the indicated interaction can be taken into account simply by superposing the corresponding solutions for concentrated forces on the solution obtained here. In this case the maximum errors arising in the calculation of the stresses in a wire as a result of replacing the other conductor by a concentrated force are about 1 and 8% for the stress components $\sigma_{rr}^{(j)}$ and $\sigma_{\theta\theta}^{(j)}$, $\sigma_{r\theta}^{(j)}$, respectively, even when their surfaces are separated by a distance equal to three radii and $\nu_i=\nu_j=\nu_m\approx0.3$ and $\delta_s^2\ll1$.

[1] E. S. Bobrov, J. E. S. Williams, and Y. Iwasa, Cryogenics **25**, 307 (1985).
[2] G. Paszter, A. Anghel, B. Jakob, and R. Wesche, IEEE Trans. Magn. Mag- **30**, 1938 (1994).
[3] J. W. Ekin, J. Appl. Phys. **62**, 4829 (1987).
[4] W. Goldacker, C. Reiger, and W. Maurer, IEEE Trans. Magn. Mag- **27**, 946 (1991).
[5] E. A. Devyatkin, Pis'ma Zh. Tekh. Fiz. **22**(10), 74 (1996) [Sov. Tech. Phys. Lett. **22**(5), 420 (1996)].
[6] E. A. Deviatkin, Cryogenics **37**, 129 (1997).
[7] A. E. H. Love, *A Treastise on the Mathematical Theory of Elasticity*, Cambridge University Press (1927), 4th edition [ONTI, Leningrad (1935), 675 pp.].

Translated by M. E. Alferieff

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Dynamic and fractal properties of SP-28 steel under high-speed loading conditions

B. K. Barakhtin, Yu. I. Meshcheryakov, and G. G. Savenkov

Institute of Mechanical Engineering, Russian Academy of Sciences, 199178 St. Petersburg, Russia

(Submitted January 27, 1997)
Zh. Tekh. Fiz. **68**, 43–49 (October 1998)

Using an interferometric method to record the velocity of the free surface of a target subjected to two-dimensional shock loading, it is shown experimentally that the decrease in the compression pulse amplitude is due to the nonstationary nature of mesoscale processes — the amplitude decrease is progressively larger for higher rates of change of the variance of the mesoparticle velocity. It is shown theoretically that the loading rate influences the spallation strength of a material in a planar collision only if the variance of the particle velocity is nonzero. A fractal analysis of the spallation surfaces of steel samples is performed by quantitative fractography methods. An expression relating the fractal dimension of the spallation fracture surface and the variance of the mesoparticle velocity is derived. For typical values of the load pulse parameters for which back-side spallation occurs the fractal dimension agrees satisfactorily with the fractal dimensions for triadic Koch islands. *© 1998 American Institute of Physics.* [S1063-7842(98)00710-7]

INTRODUCTION

In view of the widespread use of SP-28 steel in critical structures it is of practical interest to study its properties under high-speed loading.[1,2] We report below the results of dynamic tests and microstructural investigations of SP-28 steel in a highly tempered state with the following mechanical characteristics: proportional limit $\sigma_{02}=540$ MPa, ultimate strength $\sigma_b=745$ MPa, relative elongation $\delta=25\%$, and shock viscosity 9.25 MJ/m^2.

Our objective in the present investigations was to establish the relation between the spallation strength of a material and the width of the mesoparticle (0.1–10 μm) velocity distribution[3] on the one hand and the fractal dimension of the spallation fracture surface on the other.

EXPERIMENTAL PROCEDURE

The testing procedure includes shock loading of planar targets under uniaxial deformation conditions in the impactor velocity range 100–500 m/s. The spallation strength was calculated in the acoustic approximation from the formula

$$\sigma_{\mathrm{sp}}=0.5\rho_0 C_1(U_{\max}-U_{\min}), \tag{1}$$

where $\rho_0=7.8$ g/cm^3 is the density of the sample material, $C_1=5.95\times 10^5$ cm/s is the longitudinal sound velocity, and $U_{\max}$ and $U_{\min}$ are, respectively, the maximum and minimum velocities of the free surface of the sample during spallation, which were read off of the time profile. The temporal profiles of the velocity of the free surface of the targets were recorded with a high-speed two-channel interferometer.[4–6] In each act of shock loading, besides the time profile $\overline{U}(t)$ of the mean velocity, the width $\langle\Delta U(t)\rangle^2$ of the mesoparticle velocity distribution (or the square root of the variance of the mesoparticle velocity) and the kurtosis of the distribution function were also recorded.[1] Laser interferometry makes it possible to measure independently the variance of the particle velocity and the kurtosis of the distribution function. The variance is determined on the basis of measurements of the degree of contrast of the interference signal. This method is applicable only to stationary wave fronts, since the method assumes the particle velocity distribution function to be an equilibrium function and symmetric (Maxwellian) (Fig. 1). In the case of nonstationary wave fronts the distribution function is asymmetric and is characterized not only by the mean value and the variance of the particle velocity but also by the kurtosis $\Delta\Delta u(t)$.

The measurement of the variance has been discussed many times in previous works,[4–6] so that here we dwell only on the method for determining the kurtosis of the distribution function. As will be shown below, the strength behavior of a material under dynamic loading conditions is related with the kurtosis, just as with the variance.[7–9]

The determination of the kurtosis of the particle velocity distribution function is based on measurement of the number of missing beats of the interference signal as compared with the number of beats corresponding to a stationary wave front. In the ideal case of a stationary plastic front the number of beats of the interference signal in the time profile of the velocity of the free surface from the onset of the wave front (including the elastic precursor) up to the point of the transition to the plateau of the compression pulse (point A in Fig. 2) should equal the ratio of the impactor velocity to the interferometer constant V_0/U_{int} (provided that the acoustic impedances of the impactor and target are identical and therefore $V_0=U_{fs}$).

In the literature there is disagreement concerning the reasons for the decrease in the velocity on the plateau of the compression pulse. In Ref. 10 it is concluded on the basis of an analysis of the interaction of the elastic precursor, reflected from the free surface of the target, with the incident plastic front that the amplitude of the plastic wave decreases approximately by the amount of the elastic precursor. On the

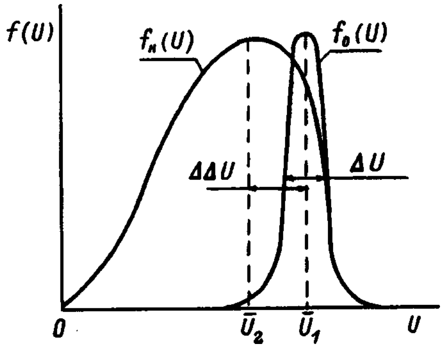

FIG. 1. Equilibrium f_0 and nonequilibrium f_n particle velocity distribution functions ($\overline{U}_1$, $\overline{U}_2$ — mean velocity for the equilibrium and nonequilibrium distributions).

other hand, in Ref. 5 it is shown on the basis of a series of experiments on shock loading of two brands of steel that for stationary wave fronts the amplitude of the plastic front does indeed decrease by the amount of the elastic precursor, while for nonstationary waves no interaction was observed between the plastic front and the elastic precursor. Since we are interested in nonstationary fronts, it is of interest to analyze this phenomenon from the standpoint of the nonequilibrium kinetics of mesosize particles.

The propagation of an elastoplastic front in a structurally nonuniform medium is a random process, which in the general case can be characterized by a nonequilibrium particle velocity distribution function $f(r,v,t)$. For a system of particles with a long-range interaction, which mesoparticles are by definition,[3] the behavior of the distribution function is described by the Fokker–Planck equation[11,12]

$$\frac{\partial f}{\partial t}+v\frac{\partial f}{\partial r}+\frac{F}{m}\frac{\partial f}{\partial t}=-\frac{\partial}{\partial v}(D_1 f)+\frac{1}{2}\frac{\partial^2}{\partial v^2}(D_2 f), \tag{2}$$

where F is the external force, m is the particle mass, $D_2=d\langle\Delta u\Delta u\rangle/dt$ is the diffusion coefficient in velocity space and characterizes the rate of change of the variance, and $D_1=d\langle\Delta u\rangle/dt$ is the so-called dynamic friction coefficient, which characterizes the rate of change of the standard deviation of the particle velocity from the average (flow) particle velocity in the wave; D_1 has the dimensions of acceleration and, since it is multiplied by the mass of a mesoparticle, equals the stopping force of the mesoparticles, which is directed opposite to the external force F which initiates the motion of the plastic front.

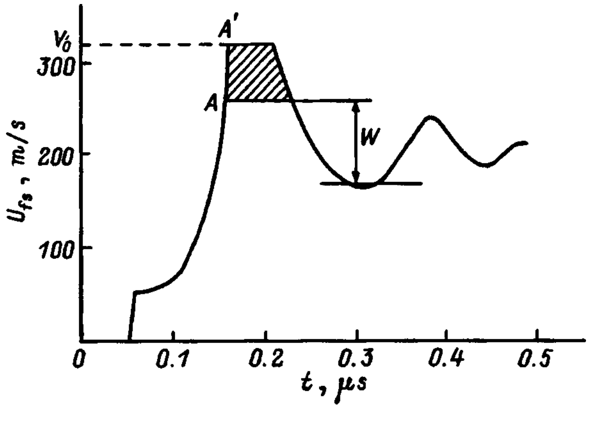

FIG. 2. Temporal profile of the velocity pulse on the free surface of a target.

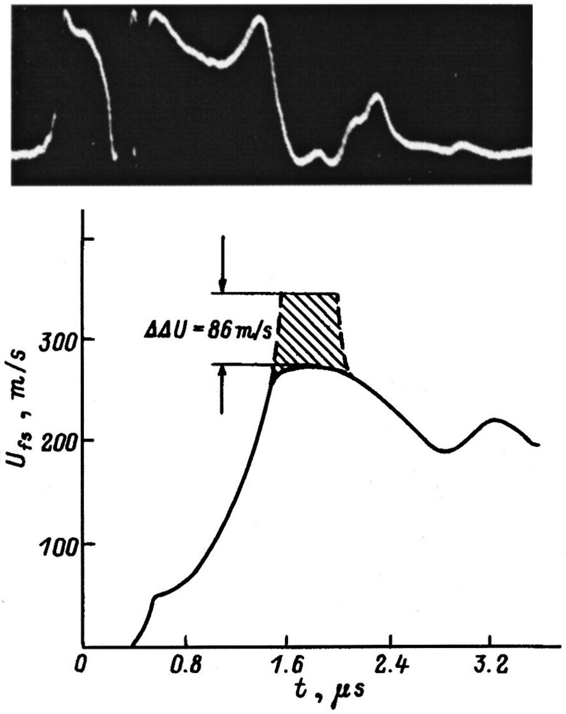

FIG. 3. Interferogram and temporal profile of the velocity of the free surface for a Cr–Ni–Mo steel target.

This stopping force decreases the amplitude of the compression pulse as a consequence of the stochastic nature of the motion of elementary carriers of plastic deformation, in this case mesoparticles — the so-called ''fluctuation'' damping of waves in a nonuniform medium. The total decrease in the amplitude of the plastic wave as a result of fluctuation damping is determined as

$$\Delta\Delta u(t)=\int_0^{\tau} D_1(t)\,dt. \tag{3}$$

In Refs. 11 and 12 it is shown that the diffusion coefficients D_1 and D_2 are related as

$$D_1=\frac{1}{2}\frac{dD_2}{dv}. \tag{4}$$

Specifically, it follows from expression (4) that the kurtosis of the particle velocity distribution function and the related decrease of the amplitude of the compression pulse are determined by the rate of change of the variance in velocity space. In other words, kurtosis appears only when the variance starts to change.

Figure 3 shows an actual interferogram, an elastoplastic pulse of the average velocity of the free surface, and the time variation of the particle velocity variance for the case of shock loading of SP-28 steel with a velocity of 350 m/s. The compression pulse is characterized by the decrease of the amplitude by the amount $\Delta\Delta u=86$ m/s. Graphical differentiation of the variance according to Eq. (4) and subsequent

TABLE I.

Impactor velocity, m/s	Target thickness, mm	Projectile thickness, mm	Width of distribution, m/s	Kurtosis of distribution, m/s	Spallation velocity, m/s
487	8.81	3.07	28.6	137	150
380	11.7	2.87	50	86.4	86.8
376	8.46	3.08	13.6	26	91.03
321	10.04	2.04	0	64	81.5
318	10.01	3.1	14.2	48	77.5
311	10.0	3.07	16	73	78
308	10.09	3.09	0	49.5	75
273	11.9	2.93	0	123	75
269	9.99	3.08	0	59	67
213	11.87	2.92	0	39	78
196.3	9.96	3.09	0	31	75
196	8.97	3.09	0	24	78
180.5	10.01	3.09	0	62.5	68
	9.56	3.09	0		73
155.5	9.27	3.1	25	60.5	68
152.5	10.42	3.09	0	36.5	79
150	9.53	3.09	0	63.4	75
142.2	10.58	3.1	0	52	88
130.9	9.96	3.07	0	46	64
104.7	9.95	3.1	0	55	47
97.5	9.97	3.1	0	6.6	80
80.84	10.05	3.04	0	5.84	

integration over the duration of the front (expression (3)) give an kurtosis of 83 m/s, which shows that almost the entire decrease in the amplitude of the compression pulse is due to fluctuation damping.

It follows from the foregoing analysis that to determine the kurtosis of the particle velocity distribution function in nonstationary plastic fronts it is sufficient to have either the temporal profile of the variance or the impactor velocity, measured independently of the temporal profile of the velocity of the free surface of the target. These methods were used in the tests of SP-28 steel.

ANALYSIS OF THE EXPERIMENTAL DATA

The complete set of test data is presented in Table I. In SP-28 steel, up to impactor velocity of 318 m/s, the mesoscale variance of the particle velocity $\langle \Delta u \rangle_{\mathrm{ms}}^2$ was found to be zero and only a kurtosis was registered. In the dependence of the kurtosis $\Delta\Delta u$ on the impactor velocity, presented in Fig. 4, one is drawn to the fact that at low deformation velocities the kurtosis does not depend on the loading velocity, but at 350 m/s it starts to increase linearly with the impactor velocity. If this linear dependence is continued downwards to the abscissa, it can be shown that the kurtosis starts to increase precisely at the moment when the mesoparticle velocity variance changes. In the experiment, the appearance of a kurtosis is delayed up to impactor velocity 350 m/s because of the fact that the velocity ''shortfall'' on the plateau of the compression pulse (hatched upper part of the compression pulse in Fig. 2) is determined not only by the value of the kurtosis but also by other factors decreasing the amplitude of the compression pulse. In the present experiments an appreciable mesoparticle velocity variance appears at impactor velocity 318 m/s, and at this level a kurtosis of the particle velocity distribution function appears. However, initially, its value remains less than the background value of the ''shortfall'' of the mean mesoparticle velocity on the plateau of the compression pulse. As the kurtosis increases its effect on the velocity ''shortfall'' becomes dominant. This corresponds to impactor velocity 350 m/s.

Analysis of the results on the spallation strength, which are presented in Table I and in Fig. 4, also shows that up to impactor velocity $V_0 = 318$ m/s the loading velocity has no effect on this characteristic, though such a relation does exist to a greater or lesser degree.[13–16] Evidently, the independence of σ_{sp} from V_0 (or, what is the same thing, from the rate of strain) is due to the fact that up to $V_0 = 318$ m/s there is no particle velocity variance $\langle \Delta u \rangle_{\mathrm{ms}}^2$.

Phenomenologically, the interrelation of the spallation stress and the mesoparticle velocity variance can be ex-

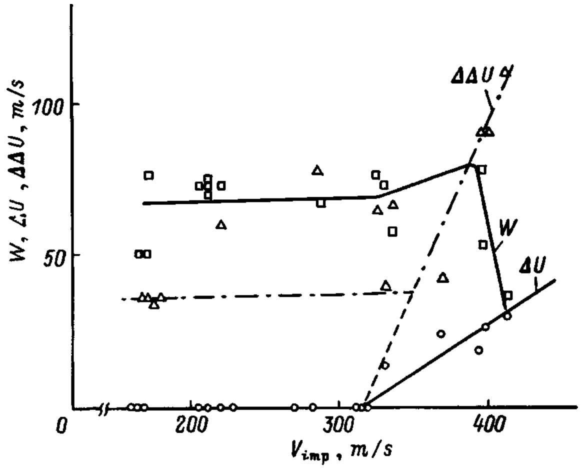

FIG. 4. Width ΔU of the particle velocity distribution, kurtosis of the distribution function, and spallation velocity on the free surface of the target versus the impactor velocity.

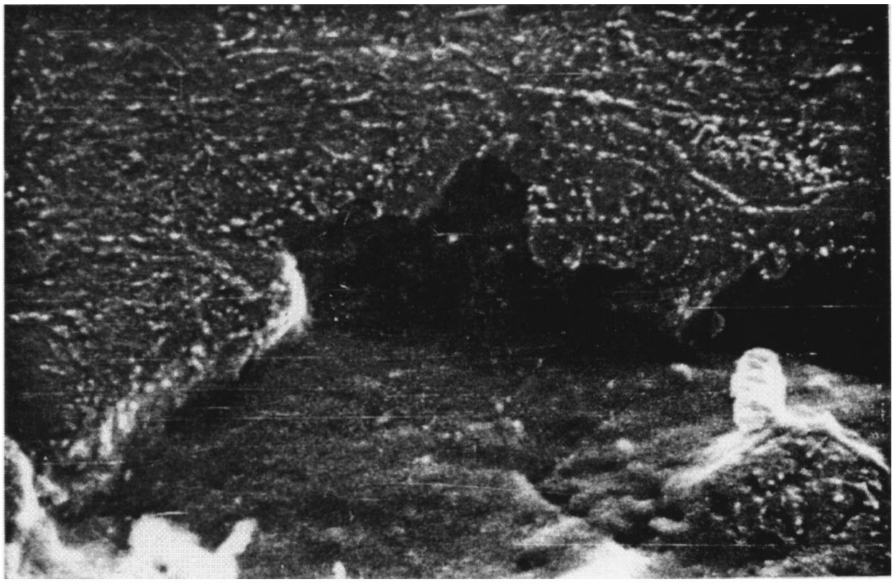

FIG. 5. Fractogram of a section of the spallation surface of the Koch-curve type.

plained as follows. As is well known, the resistance σ_{sp} to spallation fracture under uniaxial strain ε_r in the plane waves of the load can be expressed as[15]

$$\sigma_{\mathrm{sp}}=k\varepsilon_r+\frac{4}{3}(\tau_0+\mu\dot{\gamma}), \tag{5}$$

where k is the bulk modulus, τ_0 is the static resistance to shear, μ is the dynamic viscosity coefficient, and $\dot{\gamma}$ is the rate of plastic shear strain.

Keeping in mind the analogy between the turbulent motion of a liquid[17] and nonuniform high-speed deformation,[18] the viscosity of a medium can be written in the form

$$\mu=\beta\rho_0\Delta u_{\mathrm{ms}}\Delta h. \tag{6}$$

Here $\beta\approx1$ is the correlation coefficient. The plastic shear strain equals

$$\gamma=\frac{V_0}{2C\Delta t}=\frac{V_0}{\Delta h}, \tag{7}$$

C_p is the velocity of longitudinal plastic waves, Δt is the time during shear occurs, and Δh is the width of the shear zone.

Substituting expressions (6) and (7) into Eq. (5) we obtain

$$\sigma_{\mathrm{sp}}=k\varepsilon_r+\frac{4}{3}(\tau_0+\beta\rho_0V_0\Delta u). \tag{8}$$

One can see that until $\Delta u_{\mathrm{ms}}=0$ the value of V_0 has no effect on the spallation strength. When a finite variance appears, the spallation velocity also increases. This is confirmed by the behavior of $W_{\mathrm{sp}}(V_0)$ in Fig. 4.

However, the spallation strength stops increasing at impactor velocity 380 m/s, i.e., when the kurtosis of the mesoparticle velocity distribution function starts to predominate over the influence of the variance. The spallation velocity decreases because the mean mesoparticle velocity $\overline{u}$ decreases as a result of the asymmetry of the distribution function. Since the spallation velocity W_{sp} is read from the particle velocity actually existing on the plateau, an increase of the kurtosis of the distribution function and, in consequence, a decrease in the amplitude of the compression pulse result in a decrease of the spallation velocity at the meso level. This situation is qualitatively illustrated in Figs. 1 and 2.

STRUCTURAL INVESTIGATIONS

Metallographic analysis of the initial samples confirmed that their internal structure is identical to both quenched and highly tempered (to sorbite) SP-28 steel with grain sizes in the range 20–100 μm. A detailed examination of the tested targets using optical and scanning electron microscopes revealed a large number of discontinuities of different sizes which form in the material under impact loading. In the range of magnifications from 500 to $5000^{\times}$ these discontinuities had a regularly varying surface relief of ruptures and cracks. Thus, at impactor velocities $V_0=130-150$ m/s small ruptures and sinuous cracks connecting them exhibited a shape similar to a regular Koch surface[19] with the smallest (visible) figure-generation threshold 0.8–1.0 μm (Fig. 5). Such a geometry of discontinuities is observed up to impactor velocities $V_0=269$ m/s. After this velocity is reached, the discontinuities show up, and as V_0 increases further, sawtooth-shaped cracks predominate (Fig. 6).

The structural investigations performed made it possible to draw the following conclusions: A substantial number of discontinuities with an extended surface relief are present in the spallation fracture zone in a wide range of spatial scales, and the shape of the profile of their surface relief is similar to the elements of Koch curves. These conclusions indicate that fractal properties could manifest in the mechanisms of nucleation and growth of cracks. To check this conjecture, a fractal analysis of crack surfaces was performed. Since the planes of the microsections were oriented in the direction of propagation of the shock wave, the method of vertical sections could be used to find the fractal dimension of the observed discontinuities.[20] To this end, the length L of the con-

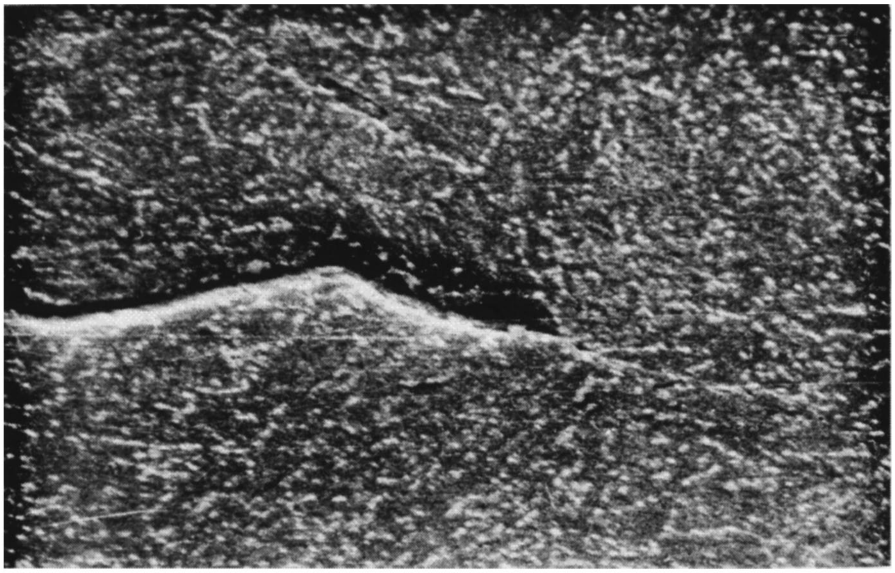

FIG. 6. Fractogram of a section of the spallation surface with saw-tooth-shaped component.

tour of the defects was measured on the photographs obtained of the structure at magnifications from 50 to $5000^{\times}$ with different unit scales n, a plot of $\log L - \log n$ was constructed, and the fractal dimension $D = \log L / \log n$ was determined from the slope.

For the same number of images in each of three groups of photographs analyzed with the chosen scale units 1–0.1, 0.1–0.01, and 0.01–0.001 mm, the fractal dimensions found are $D = 0.7$, 1.2, and 1.26 for $\log n = 1$, 2, and 3, with correlation coefficients 0.692, 0.938, and 0.972, respectively. These dimensions correspond well to the way that the fractal dimension of Koch islands with finite generations depends on the number of generations.

The results obtained show that a statistical ''joining'' of the analyzed arrays with one another is required. In practice, this means that the number of analyzed observations must be increased almost tenfold and the observation scale decreased tenfold, which is difficult to do. A fractal dimension $D = 1.1$ was obtained, with a correlation coefficient of 0.81, from the total $\log L$–$\log n$ data set, with the scale varying over three orders of magnitude. This agrees quite satisfactorily with the fractal dimension for triadic Koch islands ($D = 1.2618$ for $\log n = \log 3$; Ref. 20). Thus the result obtained attests to the fractal properties of the cracks that are formed in SP-28 steel under high-speed loading conditions.

Since the fractal dimension characterizes a physical system as a whole, it is used in the micromechanics of fracture as a measure of the degree of order (or, conversely, chaos) of the internal structure of the deformed solid when analyzing the dynamics of cracks from the standpoint of synergetics.[21] From this standpoint, it would be interesting to use fractal geometry to obtain relations between the velocity distribution function of components of the material and the topological parameters of the medium and the fractal dimension. To this end, we shall integrate the relation proposed in Ref. 14 between the rate of growth $\dot{N}$ of the density of microcracks and the spallation stresses,

$$\dot{N} = K_1 K_2 \sigma_p^{n+1}, \tag{9}$$

where K_1, K_2, and n are constants in the equation.

Integrating Eq. (9) gives

$$N = K_1 K_3 \sigma_p^{n+1} \Delta t. \tag{10}$$

Here Δt is the passage time of the reflected wave. Assuming the passage times of the forward and backward waves to be the same (this is true to a high degree of accuracy for the steel studied), we shall determine Δt in the form of the relation obtained in Ref. 19:

$$\Delta t = l_{2c} / \Delta\Delta u, \tag{11}$$

where l_{2c} is the critical length of a shear crack.

We shall determine the spallation stress σ_p from a relation which is also justified in Ref. 9:

$$\sigma_p = K_{1a} / (\pi l_{1c})^{0.5}, \tag{12}$$

where K_a is the critical coefficient of the stress intensity with respect to the stopping of a crack, and l_{1c} is a rupture crack of critical size.

Substituting expressions (12) and (11) into Eq. (10), we obtain

$$M = \frac{K_1 K_2 K_{1a}^{n+1}}{(\pi l_{1c})^{\frac{n+1}{2}}} \frac{l_{2c}}{\Delta\Delta u}. \tag{13}$$

On the other hand, according to Ref. 22 [*sic*], after some modification the quantity N can be written as

$$N = \lambda l_{1c}^{-2D}, \tag{14}$$

where λ is a constant.

Equating expressions (13) and (14) we obtain

$$\frac{\Theta}{\Delta\Delta u} = \frac{l_{1c}^{m-D}}{l_{2c}}, \tag{15}$$

where $\Theta = K_1 K_2 k_{1a}^{(n+1)} / \pi$ and $m = (n+1)/2$.

Taking the logarithm of expression (13), we present the final result in the form

$$D=m+\frac{\log(\Delta\Delta u)-\log(l_{2c})-\log\Theta}{\log(l_{1c})}. \tag{16}$$

Substituting typical values of the parameters appearing in relation (16), $m=1.5$, $l_{1c}=10^{-4}$ m, $l_{2c}=5\times10^{-5}$ m, $\Delta\Delta u=60$ m/s, yields the fractal dimension $D=0.943$, which is quite close to the value $D=1.1$ found experimentally on the basis of quantitative fractography.

In summary, it can be concluded that the relation obtained determines a correspondence between the kurtosis of the velocity distribution function of the mesoscale components of a medium, the mechanical and topological properties of the medium, and the fractal dimension of the discontinuities formed as a result of spallation fracture.

This work was supported by the Russian Fund for Fundamental Research (Project No. 96-02-16807-a).

---

1) In probability theory the so-called Pearson asymmetry measure $s=(\overline{u}-u_m)/\Delta u$, where u_m is a dynamical variable (in the present case the particle velocity) corresponding to the maximum of the distribution, $\overline{u}$ is the mean particle velocity (mathematical expectation), which in the case of an asymmetric distribution is different from the distribution maximum, and Δu is the distribution width, is often used to characterize the degree of asymmetry of the distribution function. In the present work it is more convenient to use the absolute value of the degree of asymmetry $u-u_m=\Delta\Delta u$, which in what follows we shall call the kurtosis.

[1] I. P. Polukhin, G. Ya. Gun, and A. M. Galkin, *Plastic Resistance of Metals and Alloys* [in Russian], Metallurgiya, Moscow, 1983, 352 pp.

[2] G. R. Épishteĭn, *Structure of Metals Deformed by Explosion* [in Russian], Metallurgiya, Moscow, 1988, 280 pp.

[3] V. I. Vladimirov, in *Collective Effects in Ensembles of Defects* [in Russian], Nauka, Leningrad, 1987, pp. 43–57.

[4] A. K. Divakov, L. S. Kokhanchik, Yu. I. Meshcheryakov, and M. M. Myshlyaev, Zh. Prikl. Mekh. Tekh. Fiz., No. 3, 133 (1987).

[5] Yu. I. Mescheryakov and A. K. Divakov, DYMAT J. **1**, 271 (1994).

[6] Yu. I. Mescheryakov, in *Shock Compression of Condensed Matter-1995*, Seattle, 1995. Pt. 1, pp. 623–625.

[7] B. K. Barakhtin, Yu. I. Meshcheryakov, and G. G. Savenkov, Zh. Tekh. Fiz. **61**(6), 8 (1991) [Sov. Phys. Tech. Phys. **36**, 589 (1991)].

[8] Yu. I. Meshcheryakov and G. G. Savenkov, Problemy Prochnosti, No. 12, 83 (1990).

[9] Yu. I. Meshcheryakov and G. G. Savenkov, Zh. Prikl. Mekh. Tekh. Fiz., No. 3, 138 (1993).

[10] D. E. Grady, in *Proceedings of the International Conference on ''Metallurgical Applications of Shock-Waves and High-Strain-Rate Phenomena,'' Explomet-85,* pp. 587–597.

[11] J. Hubburd, Proc. Roy. Soc. A **260**, 114 (1961).

[12] Yu. I. Mescheryakov and E. I. Prokuratova, Int. J. Solids Struct. **32**, 1711 (1995).

[13] M. N. Stepnov, *Statistical Methods for Analyzing the Results of Mechanical Tests* [in Russian], Metallurgiya, Moscow, 1985, 232 pp.

[14] G. I. Kanel', S. V. Pazorenov, and V. E. Fortov, Zh. Prikl. Mekh. Tekh. Fiz., No. 5, 60 (1984).

[15] G. V. Stepanov, *Elastoplastic Deformation and Fracture of Metals Under Pulsed Loading Conditions* [in Russian], Naukova Dumka, Kiev, 1991, 288 pp.

[16] I. A. Volkov, Zh. Prikl. Mekh. Tekh. Fiz., No. 2, 19 (1993).

[17] Yu. V. Grinyaev and V. E. Panin, in *Experimental Investigation and Theoretical Description of Disclinations* [in Russian], Nauka, Leningrad, 1988, pp. 66–72.

[18] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Pergamon Press, New York, 1987; Nauka, Moscow, 1988).

[19] E. E. Undertwood, J. Appl. Mech. , No. 10, 10 (1990).

[20] C. W. Luny and S. S. Zhny, Preprint No. 76, Intern. Atomic Agency and United Nation Educat. Science and Cultural Organ. (1989).

[21] V. S. Ivanova, *Synergetics. Strength and Fracture of Metallic Materials* [in Russian], Nauka, Moscow, 1992, 160 pp.

Translated by M. E. Alferieff

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Possibilities for using a pulsed magnetic field to influence structural states in oxide glass

V. I. Alekseenko

A. A. Galkin Donetsk Physicotechnical Institute, Ukrainian National Academy of Sciences, 340114 Donetsk, Ukraine

(Submitted January 5, 1997; resubmitted June 5, 1997)
Zh. Tekh. Fiz. **68**, 50–54 (October 1998)

The structural–energy spectrum of the states of a bismuth-containing oxide glass, the sensitivity of these states to the action of a pulsed magnetic field, and the thermodynamic and kinetic stability of the structure excited by a pulsed magnetic field are studied by the method of measuring the internal friction. It is established that a pulsed magnetic field influences the structural states and that this leads to irreversible changes in the structure and the crystallization parameters of the glass. It is found that the efficacy of acting on a material with a pulsed magnetic field also depends on the parameters of the field and the structural–energy state of the material and that the action itself is of a thermally-activated relaxational character. The optimal conditions for the action of a pulsed magnetic field on glass are determined. © *1998 American Institute of Physics.* [S1063-7842(98)00810-1]

INTRODUCTION

Since 1973, when the method of pulsed magnetic field treatment (PMFT) of condensed systems was proposed, extensive experimental data has been accumulated proving that a considerable change in the real structure and macroscopic characteristics of materials can be brought about by exposure to a sequence of pulses of a relatively weak magnetic field with intensity $H < 10^6$ A/m. Data have been obtained on the decay of impurity phases in crystalline matrices,[1,2] the precipitation of finely dispersed phases of hardeners,[3] the possibility of coherent ordering of an impurity, right up to the appearance of spatial dissipative structures,[4] the change in the relative arrangement and shape of clusters of electrically active centers in semiconductors,[5,6] and other data. They are correlated with data on the change in hardness and durability of high-speed steels and change in microstresses.[7] A change in the viscosity of petroleum after PMFT has also been observed.[8] These are all residual effects and they have been observed in materials subjected to PMFT at room temperature, where the magnetic field energy $\mu_B H$ is three orders of magnitude lower than the thermal energy (for $H = 10^6$ A/m).

The intensity of irreversible relaxation processes, which result in the formation of structure and properties in the material after PMFT, is determined to a large degree by the thermodynamic conditions of the residual effect, while the role of the regimes and the character of PMFT remains, to some extent, in the background. However, a number of results attest unequivocally to the existence of changes occurring in the properties of materials during treatment in a magnetic field. Thus, the viscosity of chalcogenide glassy semiconductors has been observed to change as a result of the action of a magnetic field,[9] an ac field having a stronger effect than a dc field, and the change depends on the temperature. In Ref. 10 it was observed that the electrical conductivity of polyacetylene films decreases when an external magnetic field is switched on. It is interesting that the effect of a magnetic field was observed in thermodynamically nonequilibrium systems. These results suggested that for each specific material, besides optimal thermodynamic conditions for the residual effect, there also exist regimes for optimal absorption of a magnetic field, as a result of which the maximum changes in particular properties can be achieved.

The main goals of the present investigations are to conduct a scientific search for the optimal regimes for absorption of a pulsed magnetic field in bismuth-containing glass and to study the mechanism and the influence on the crystallization of the glass of such absorption.

RESULTS OF PRELIMINARY INVESTIGATIONS

Superconducting phases are obtained by crystallization of glass, which by its very nature is a strongly nonequilibrium material with respect to both the stabilized amorphous and crystalline state. This enabled us to employ PMFT as a potentially possible method for influencing the crystallization process and the properties of superconducting phases. Since relatively little energy is expended, the intent was to improve the efficiency of the process of crystallization of the glasses in superconducting phases and to improve the physicochemical properties of the glasses. The results of preliminary investigations in the indicated glass by means of differential-thermal analysis, x-ray crystallographic analysis, and measurements of the mechanical properties served as the basis for this. Thus, treatment of $Bi_{1.6}Pb_{0.4}Sr_2CaCu_2O_x$ glass with magnetic field pulses with $H = 10^6$ A/m and pulse repetition frequency $f_i = 1$ Hz for 5 min at room temperature decreased the temperature of crystallization in the 2201 superconducting phase and the activation energy of crystallization (Fig. 1) and increased the crystallization rate (Fig. 2) compared with these parameters in the untreated material.[11] In this connection, as a result of PMFT of glass, a relative

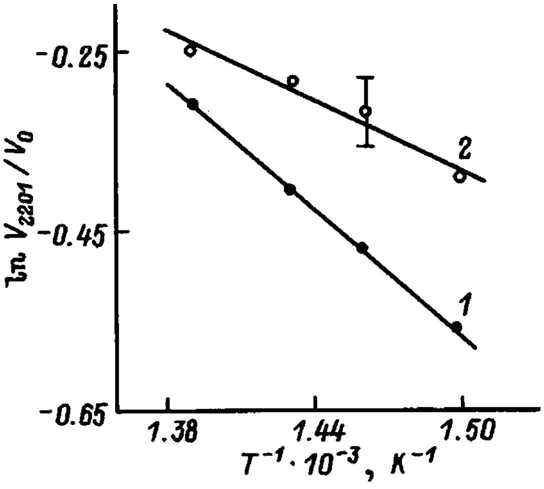

FIG. 1. Temperature dependence of the relative quantity of the 2201 phase in the process of crystallization of glass in 20 min: *1* — Initial glass, *2* — after treatment with a pulsed magnetic field in regime 5 (Table IV).

increase in the quantity of the 2201 superconducting phase and a decrease in the level of internal microstresses in it were observed. The final material was stronger. Similar results were also observed for the superconducting phases 2212 and 2223.

It became clear that PMFT of the glass chosen gives a positive result and can serve as one way of improving the technology for obtaining high-T_c superconducting materials, at least for those materials which are obtained in the glass variant. The remaining problem consisted of optimizing the PMFT regimes for the glass investigated.

RESULTS OF BASIC INVESTIGATIONS AND DISCUSSION

To solve this problem it is necessary to study the structural–energy states in the quenched glass, specifically, to determine the energy spectrum of the structural states, their sensitivity to external action (mechanical, magnetic, or electromagnetic since pulsed magnetic fields are used), the relaxational characteristics of the processes leading to a transition of the structural elements into excited states as a result of an external action (the relaxation time τ and the rate factor τ_0 — a parameter determining the character of the relaxational process), thermodynamic reversibility or irreversibility, and the kinetic stability of the structure of the experimental material in excited states. For this, the internal friction (IF) was studied by the compound vibrator method at a frequency of longitudinal oscillations 91 kHz with a constant amplitude of the cyclic strain $\varepsilon=6\times10^{-8}$ in the temperature interval 80–500 K. The measurement error was equal to 10%.

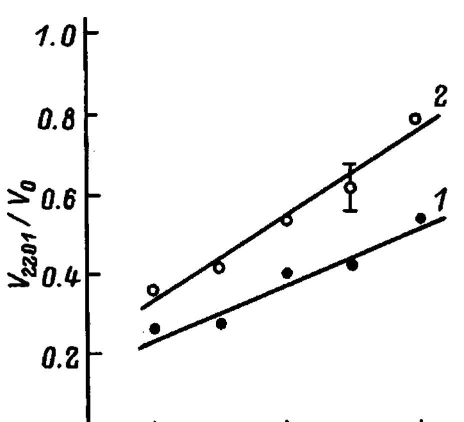

FIG. 2. Kinetic dependence of the relative amount of the phase 2201 in the process of crystallization of the glass at 773 K (*1,2* — the same as Fig. 1).

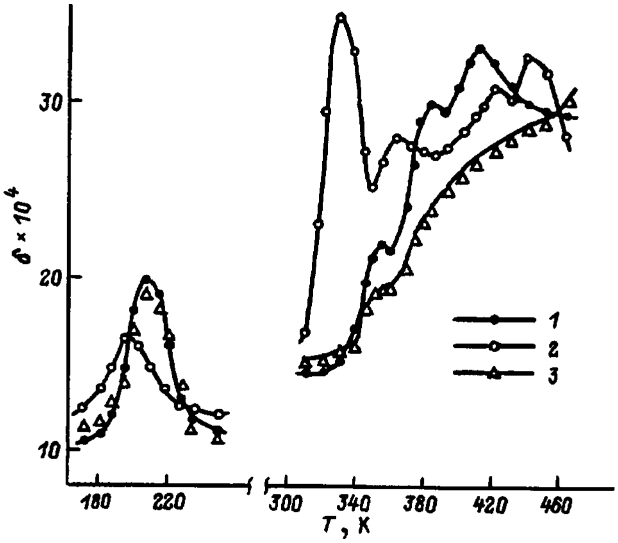

FIG. 3. Temperature dependence of the damping decrement in glass: *1,2* — the same as in Fig. 1; 3 — measurements of internal friction in the experimental temperature range after several temperature runs.

Figure 3 shows the temperature dependence of the damping decrement δ, which to within a constant equals the IF (Q^{-1}). As one can see, peaks of the IF are observed at the temperatures $T_{m1}=205\,\text{K}$, $T_{m2}=353$ K, $T_{m3}=378\,\text{K}$, and $T_{m4}=413\,\text{K}$. These peaks are due to relaxational processes leading to a transition of the structural elements into excited states in a field of cyclic stresses. From the standpoint of the efficiency of the absorption of mechanical energy by the material, the IF peaks characterize the conditions for maximum utilization of the external mechanical energy in the observed relaxational processes with their own thermodynamic (activation energies) and kinetic (relaxation times) parameters. The activation energies of the relaxation processes were determined using the Wert–Marx expression[12]

$$U=RT_m\ln\frac{RT_m}{h\omega}+T_m\Delta S, \tag{1}$$

where ΔS is the activation entropy, ω is the angular frequency of the measurements, and $\Delta S=4-20$ J/mole·K;[12] the energies were estimated for $\Delta S=10$ J/mole·K.

The relaxation times were determined from an expression which relates the temperature positions of the maxima of the IF peaks and the activational characteristics of the relaxational processes[12]

$$\omega\tau_0\exp(U/kT_m)=1, \tag{2}$$

TABLE I.

Peak No.	T_m, K	τ, s	U, eV
1	205	$1.7\cdot 10^{-6}$	0.30
2	353	$4.7\cdot 10^{-6}$	0.53
3	378	$3.2\cdot 10^{-6}$	0.57
4	413	4.0×10^{-6}	0.63

where $\tau_0 \exp(U/kT_m) = \tau$.

The limitations of expression (1) should be noted. This expression is valid only when the relaxational process satisfies certain prerequisites, since the derivation of the expression employed an assumption that the frequency factor is the same for all relaxational processes and is 10^{-13} s in order of magnitude. This corresponds to the Debye frequency of the oscillations of a single atom $\nu = 1/\tau_0$. Since we know of no other mechanisms of IF in oxide glasses (for example, for the dislocation mechanism $\tau_0 = 10^{-9} - 10^{-10}$ s) other than the mechanism of thermally activated displacement of single atoms, we shall employ expression (1). The results of a numerical analysis are presented in Table I.

Therefore the choice of method for measuring the IF is successful in the sense that it reveals the possibilities for the material to absorb external energy and makes it possible to determine the character and mechanisms of energy dissipation processes. But, most importantly, a characteristic of the method is the selective character of the absorption of external mechanical energy, also observed in the case of ac and pulsed magnetic fields applied to a material. For this reason, the information obtained about the spectrum of structural–energy states of the glass under study makes it is possible to find the specific conditions of the PMFT of the glass such that absorption of the field occurs and, as a result, if the states excited by the magnetic field are kinetically unstable, changes occur in the structure and properties of the glass.

Next, it is necessary to solve the question of the kinetic stability of the observed states (IF peaks) in the case that they are excited by a mechanical field and pulsed magnetic fields. It was found that the observed IF peaks are converted during the action of pulsed magnetic fields on the material (Fig. 3, curve *2*) and during mechanical action (Fig. 3, curve *3*) into a system of new peaks with different characteristics. This experimental fact shows that periodic external actions drive the material into excited states which are inherently kinetically unstable, which is what leads to the development of structural relaxation (SR) and irreversible changes in the structure. As follows from Fig. 3, the results of structural relaxation are different for magnetic and mechanical actions. The reason could be that in the case of PMFT of the experimental glass not all observed states are excited in the glass, but rather only the states which are inherently magnetically active.

We shall take the last step in the solution of the problem by making a number of assumptions, hoping in advance that the experimental check will elucidate the situation. If all observed structural–energy states of the glass are assumed to be magnetically active, then the absorption of energy from the magnetic field will be maximum at magnetic field frequency 91 kHz, just as in the case of a cyclic mechanical field with such a frequency at $T_{m1} = 205$ K, $T_{m2} = 353$ K, $T_{m3} = 378$ K, and $T_{m4} = 413$ K. However, we are dealing with a pulsed magnetic field in which, depending on the pulse duration τ_i and the pulse repetition frequency f_i, certain harmonics with different amplitudes H_n and frequencies $n\Omega_i$ will be present, and their energy will be absorbed efficiently at definite temperatures. The Fourier transform was used to determine the harmonic components present in the carrier of the working magnetic field pulses. The Fourier series for periodic magnetic field pulses is

TABLE II.

PMTF conditions H_i, A/m; f_i, Hz; τ_i, μs	H_0, A/m	Working harmonics range	Working amplitude range, A/m	Working frequency range, Hz
10^5; 1; 500	50	$1-1000$	$100-60$	$6.3-6300$
10^5; 10; 500	500	$1-1000$ (except 200*n*)	$1000-60$	$63-63000$
10^4; 1; 500		H_n of all harmonics below H_g		
10^4; 10; 500	50	$1-100$	$100-60$	$6.3-6300$

$$H(t) = H_0 + \sum_{n=1}^{\infty} H_n \sin(n\Omega_i t + \Psi_n), \tag{3}$$

where $H(t)$ is the intensity of the pulsed magnetic field with repetition frequency $T_i = 1/f_i = 2\pi/\Omega_i$, $H_0 = H_i K$ is the static component of the magnetic field, H_i is the amplitude of the intensity of the magnetic field pulse, $K = \tau_i/T_i$ is the filling factor, and H_n and Ψ_n are the intensity amplitude and phase of the *n*th harmonic.

For the case of a symmetric arrangement of the pulse relative to the ordinate and initial phases $\Psi_n = 90°$ the series will consist of H_0 and cosinusoidal components

$$H(t) = H_0 + \sum_{n=1}^{\infty} H_n \cos(n\Omega_i t), \tag{4}$$

where the amplitudes of the harmonic components are given by

$$H_n = \frac{2H_i}{n\pi} \sin(n\pi K). \tag{5}$$

We shall use Eq. (4) to analyze some of our PMFT regimes with respect to the main parameters of the pulses H_i and f_i with $\tau_i = 500$ μs to determine the working ranges of the amplitudes and frequencies. The range of working harmonics which were included in the choice of optimal PMFT regimes was determined by their amplitudes and was conditionally limited by the intensity of the geomagnetic field with $H_g = 46$ A/m. The working frequency range was also determined according to this criterion. The results are presented in Table II.

Next, we present estimates of the temperature regimes of PMFT of bismuth-containing glass in the working frequency range for the spectrum of structural states determined by the method of measuring the IF. Since the relaxational processes

TABLE III.

U, eV	Working frequency range, Hz	H_i, A/m	Working temperature range, K	f_i, Hz	τ_i, μs
0.30	6.3–6300	10^5	120–160	1	500
	63–63000	10^5	130–180	10	500
0.53	6.3–6300	10^5	210–280	1	500
	63–63000	10^5	225–310	10	500
0.57	6.3–6300	10^5	225–300	1	500
	63–63000	10^5	245–335	10	500
0.63	6.3–6300	10^5	250–330	1	500
	63–63000	10^5	270–365	10	500

TABLE IV.

Regime No.	H_i, A/m	f_i, Hz	PMFT T, K	τ_i, μs	PMFT t, min	Relative increase in 2201, %
1	10^5	1	150	500	5	10
2	10^5	1	250	500	5	60
3	10^5	1	250	500	2	50
4	10^5	1	250	500	15	50
5	10^5	1	300	500	5	90
6	10^5	1	320	500	5	40
7	10^5	1	320	500	15	35
8	10^5	10	150	500	5	10
9	10^5	10	250	500	5	50
10	10^5	10	300	500	5	120
11	10^5	10	320	500	5	95

leading to the excitation of the structure in a cyclic elastic field are thermally activated, we shall perform the calculations using expression (2), making the assumption that the character of the relaxation is the same in pulsed magnetic fields. Such a point of view could be consistent because of the fact that the energy of the magnetic field employed, just as that of the mechanical field (for $\varepsilon = 6\times10^{-8}$), is much less than the thermal energy and especially the activation energy of relaxational processes studied in the present work. For this reason, neither a magnetic field nor a mechanical field influences the density of thermal fluctuations responsible for any particular excitation process. The results of determining the temperature regimes of PMFT of our glass for $H_i = 10^5$ A/m are presented in Table III.

The results presented in Table III show the temperature intervals where maximum absorption of energy from a pulsed magnetic field with fixed parameters occurs in a bismuth-containing glass.

To check the reliability of the computed results, it was necessary to choose a parameter of the material that could characterize qualitatively and quantitatively the result of PMFT of the glass investigated. As shown above, PMFT of glass changes the parameters of the crystallization of the glass in superconducting phases. In this connection, after PMFT a relative increase was observed in the amount of superconducting phases as compared with glass not treated with a pulsed magnetic field. The increase in the 2201 phase on crystallization by PMFT of the glass was used as a control parameter for the experimental check of the computational results. The results of such a check are presented in Table IV.

CONCLUSIONS

As one can see from Table IV, the magnitude of the effect presented there depends strongly on the PMFT regime. Thus, PMFT in regimes 1 and 8 gives the weakest effect. This is due to the fact that only one structural–energy state with relaxation energy 0.3 eV is excited. The amplitude of the IF peak for the relaxational process with an activation energy of 0.3 eV is the lowest of all peaks investigated (Fig. 3, curve *1*) and therefore this state makes the minimal contribution to the absorption of a pulsed magnetic field. In the case of glass treated in a pulsed magnetic field in the regimes 2, 3, 4, and 9, states with energies 0.53 (to the greatest degree) and 0.57 eV participate in absorption of the magnetic field. Their contribution is higher than the preceding one. Comparing the regimes 2–4 according to the PMFT time it can be concluded that the treatment does not greatly influence the magnitude of the effect. The greatest increase in the 2201 phase as a result of PMFT of the glass is observed for regimes 5, 10, and 11, where states with energies 0.53, 0.57, and 0.63 eV, respectively, operate. It should be noted that a correlation between crystallization and magnetic field absorption may not be observed because not all PMFT-induced changes in the glass structure can influence the crystallization of the glass. This is serious question requiring a separate analysis.

The investigations revealed that the mechanism leading to the absorption of pulsed magnetic fields or in general the mechanism whereby the pulsed magnetic field influences the glass is of a temperature-dependent relaxational character. For this reason, the action of magnetic fields on a material must be judged not from the standpoint of the ‘‘thermal paradox’’ ($kT \gg \mu_B H$) but rather from the standpoint of thermally activated relaxation. The efficcacy of PMFT in the sense described in the present paper is determined not only by the magnetic field parameters but also by the structural–energy state of the material, which, as observations showed, changes both during and after PMFT. This is why prolonged PMFT does not give an additional result. For this reason, each PMFT treatment must be preceded by a study of the structural–energy state of the material, which makes it possible to determine the optimal PMFT regime. Since the process of absorption of a pulsed magnetic field is of a relaxational character with a characteristic maximum of absorption, a static magnetic field will be ineffective for diamagnetic materials. The negligible effects of treating diamagnetic systems (including also the system investigated) with a static magnetic field could be due to the presence of small quantities of paramagnetic centers and other possible mechanisms of the effects of magnetic fields.

An important point is that the results presented above determined the specific approach to studying the effect of cyclic magnetic fields on any material with different degree of ordering. Once the conditions for excitation of thermally-activated relaxational processes are known, the conditions for absorption of pulsed magnetic fields can be determined with a definite degree of accuracy. This method makes it possible to find the correct approach for studying the influence of variable or periodic magnetic fields on any thermally-activated relaxation processes, for example, brittle–plastic transition in crystalline structures, strain aging, nucleation and motion of dislocations, phase formation in amorphous systems, and many other processes. Here a special place should be given to pulsed magnetic fields, which, in view of their intrinsic wide spectrum of harmonics, are capable of selective action on different processes.

[1] S. N. Postnikov, *Electric Phenomena Accompanying Friction and Cutting* [in Russian], Volgo-Vyatskoe Book Publishers, Gor'kiĭ (1975), 280 pp.

[2] G. I. Distler, V. M. Kanevskiĭ, V. V. Moskvin *et al.* Dokl. Akad. Nauk SSSR **268**, 591 (1983) [Sov. Phys. Dokl. **28**, 43 (1983)].

[3] S. N. Postnikov, *Electrophysical and Electrochemical Phenomena in Friction, Cutting, and Lubrication*, New York (1978) 281 pp.

[4] S. N. Postnikov, V. P. Sidorov, A. V. Ilyakhinskiĭ *et al.*, in *Applied Problems in Strength and Plasticity* [in Russian], Gor'kiĭ (1980), pp. 138–143.

[5] A. N. Buzynin, V. N. Buzynin, N. A. Butylkina *et al.*, in *Abstracts the Eighth All-Union Conference on Methods for Obtaining and Analyzing Ultrapure Materials* [in Russian], Gor'kiĭ (1988), pp. 174–176.

[6] S. N. Postnikov, *Treatment with Pulsed Magnetic Fields* [in Russian], Sofiya (1989), 133 pp.

[7] S. N. Postnikov and V. P. Sidorov, in *Applied Problems of Strength and Plasticity* [in Russian], Gor'kiĭ (1981), pp. 150–154.

[8] S. N. Postnikov, V. G. Krasnov, and Yu. D. Sedov, Elektron. Obrab. Materialov, No. 2, 45 (1986).

[9] S. A. Dembovskiĭ, E. A. Chechetkina, and S. A. Kozyukhin, JETP Lett. **41**, 88 (1985).

[10] E. L. Frankevich, D. I. Sokolik, D. I. Kadyrov, and V. M. Kobryanskiĭ, JETP Lett. **36**, 486 (1982).

[11] Yu. D. Tret'yakov and P. E. Kazin, Neorg. Mater. **29**, 2246 (1993).

[12] S. N. Golovin, A. Pushkar, and D. M. Levin, *Elastic and Damping Properties of Metal Construction Materials* [in Russian], Metallurgiya, Moscow (1987), 190 pp.

[13] A. Feltz, *Amorphous Inorganic Materials and Glasses*, VCH Publ., [Weinheim–New York (1993) Russian translation from German, Mir, Moscow (1987)], 556 pp.

[14] H. S. Chen, *Amorphous Metallic Alloys* [in Russian], Mir, Moscow (1986), pp. 164–183.

[15] V. I. Alekseenko, G. K. Volkova, T. E. Konstantinova *et al.*, Fiz. Tverd. Tela (St. Petersburg) **36**, 1597 (1994) [Phys. Solid State **36**, 873 (1994)].

Translated by M. E. Alferieff

Fabrication and investigation of $SnO_2-As_2(Se_{0.9}Te_{0.1})_3$ and $SnO_2-(As_{0.67}Sb_{0.33})_2Se_3$ heterojunctions

I. P. Arzhanukhina, K. P. Kornev, and Yu. V. Seleznev

Kaliningrad State University, 236041 Kaliningrad, Russia

(Submitted November 26, 1997)
Zh. Tekh. Fiz. **68**, 55–57 (October 1998)

A method for fabricating and the results of an investigation of $SnO_2-As_2(Se_{0.9}Te_{0.1})_3$ and $SnO_2-(As_{0.67}Sb_{0.33})_2Se_3$ heterojunctions are described. The spectral and current–voltage characteristics of the heterojunctions obtained are presented. © *1998 American Institute of Physics.* [S1063-7842(98)00910-6]

INTRODUCTION

For semiconductor electronics, the search for new structures that can be used to develop a variety of semiconductor devices, specifically, photodetectors operating in different parts of the spectrum, is now very important. Examples of such structures are heterostructures based on amorphous materials, including glassy chalcogenide semiconductors (GCSs). Semiconductor heterojunctions are the focus of active recent investigations. The technology of heterostructures, apparently, will play an important role in the growing microelectronics market.[1] As materials, glassy chalcogenide semiconductors are interesting for use in electronics, since their properties can be purposefully altered by continuously varying the combination of components in a compound.

Our objective in the present work was to fabricate and investigate the characteristics of $SnO_2-As_2(Se_{0.9}Te_{0.1})_3$ and $SnO_2-(As_{0.67}Sb_{0.33})_2Se_3$ heterostructures.

SAMPLE PREPARATION AND MEASUREMENT PROCEDURE

The samples whose structures are displayed in Fig. 1 were prepared for the investigations. A tin layer was deposited on a 13×13 mm glass substrate, using masks, in vacuum at room temperature. Next, the deposited structures were heated in an oxygen atmosphere up to a temperature of approximately 700 K. This yielded two electrodes consisting of a transparent (in the visible part of the spectrum) *n*-type conducting tin dioxide film with a band gap of 4.0 eV for direct transitions and 2.8 eV for indirect transitions.[2] The ionization energy of the donor centers, which is associated with the presence of oxygen vacancies, was equal to 0.07 eV.[2,3] Next, layers of a glassy chalcogenide semiconductor and aluminum, which partially overlapped one another, were deposited successively, using different masks, on the electrodes at room temperature (Fig. 1).

The sample was illuminated on the glass substrate side through the transparent tin dioxide electrode. The aluminum layer served as the second electrode. A voltage was applied between the Al and SnO_2 electrodes. The polarity of the voltage was determined by the sign of the voltage on SnO_2.

The measurements were performed on a specially developed apparatus that made it possible to measure the photocurrent in the sample in the region of the spectrum with wavelength from $\lambda=0.4$ to $1.2\ \mu$m, with a spectral slit width from $\delta\lambda=0.002\ \mu$m for the short-wavelength part of the spectrum up to $\Delta\lambda=0.008\ \mu$m for the long-wavelength part of the spectrum. The minimum measured current was equal to 10^{-12} A. The voltage applied to the sample could be varied from $+100$ to -100 V. The IVCs obtained in the dark and the IVCs obtained by illuminating the sample with light in the region of maximum sensitivity (light IVCs) were measured for both types of samples. The spectral characteristics of photosensitivity (SCP) were measured with both polarities of the voltage applied to the sample. To obtain the SCP, the dependence of the photocurrent of the sample on the wavelength of the incident light was measured in the range of wavelengths from 400 to 900 nm with a step of 25 nm. The spectral sensitivity of the sample was determined from the measured dependence according to the formula

$$S(\lambda)=J_{\mathrm{ph}}(\lambda)/I(\lambda),$$

where $J_{\mathrm{ph}}(\lambda)$ is the photocurrent in the sample in amperes and $I(\lambda)$ is the power of the lamp in watts in the wavelength interval $\Delta\lambda$.

The photocurrent was determined according to the voltage drop across a known resistance.

RESULTS AND DISCUSSION

The current–voltage characteristics for samples based on both compositions are nonlinear and asymmetric. For both the dark and light IVCs measured in the region of maximum sensitivity the current obtained with positive polarity of the tin dioxide collecting electrode is much higher than the current obtained with negative polarity. A potential barrier for holes, which are the majority carriers in GCSs,[4] is formed at the metal–semiconductor interface (Fig. 2). When the collecting electrode is positive, the barrier decreases and a larger number of holes can flow into the aluminum from the GCS. For opposite bias the barrier for holes emitted from aluminum decreases as the voltage increases; the current increases but to a lesser extent, since a barrier for holes exists at the GCS–SnO_2 interface. In this case the current through the structure will be determined by the recombination rate of holes and electrons near the GCS–SnO_2 interface. Therefore,

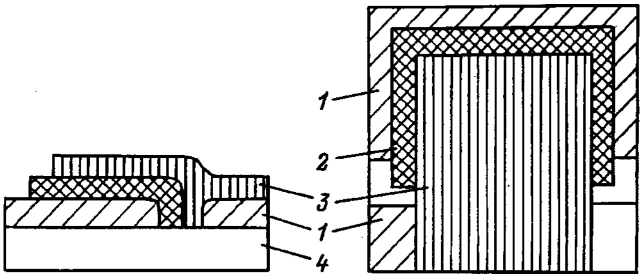

FIG. 1. Structure of the sample: *1* — SnO_2, *2* — GCS, *3* — Al, *4* — substrate.

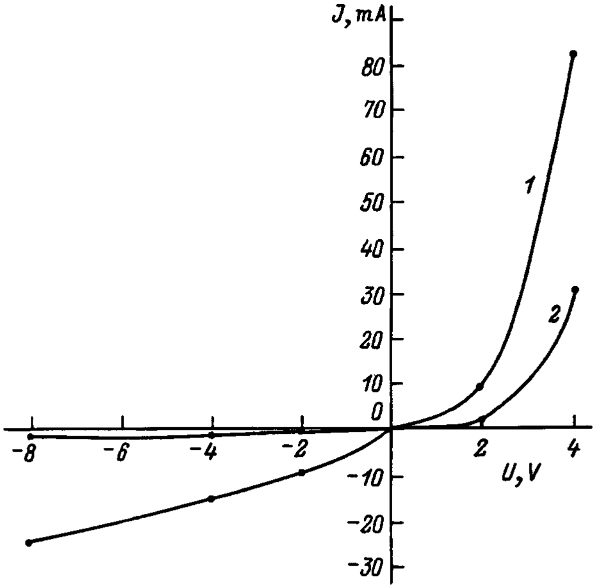

FIG. 3. Current–voltage characteristics of a $(As_{0.67}Sb_{0.33})_2Se_3$ heterostructure: *1* — in light, *2* — in dark.

as the voltage increases, the current with a positive collecting electrode should grow more rapidly than for a negative collecting electrode. This is observed in Fig. 3, which displays the IVCs for the $(As_{0.67}Sb_{0.33})_2Se_3$ heterostructures. The IVCs of the samples based on $As_2(Se_{0.9}Te_{0.1})_3$ have the same form.

The spectral characteristics of photosensitivity $S(\lambda)$ for $As_2(Se_{0.9}Te_{0.1})_3$ heterostructures with a negative correcting electrode and GCS film thickness $d=7$ μm are presented in Fig. 4. As one can see, the region of photosensitivity of the structure extends from $\lambda=0.47$ to 0.71 μm at the $0.1S_{\max}$ level with a maximum at $\lambda=0.58$ μm. For film thickness $d=1$ μm the sensitivity region ranges from $\lambda=0.4$ to 0.7 μm with a maximum at $\lambda=0.53$ μm, while for film thickness $d=10$ μm it extends from $\lambda=0.53$ to 0.76 μm with a maximum at $\lambda=0.65$ μm.

In the case of a collecting electrode with positive polarity the region of spectral sensitivity lies in shorter wavelength part of the spectrum relative to the curves obtained for negative polarity, but in this case the long-wavelength edge of the curves is the same for both polarities.

For the $(As_{0.67}Sb_{0.33})_2Se_3$ heterostructures with negative polarity the region of the spectral sensitivity at the $0.1S_{\max}$ level ranges from $\lambda=0.515$ to 0.730 μm with a maximum at $\lambda=0.65$ μm ($d=1$ μm) and from $\lambda=0.575$ μm to 0.725 μm with a maximum at $\lambda=0.625$ μm ($d=10$ μm).

In the case of positive polarity, only the long wavelength part of the SCP is observed in the range of wavelengths investigated. The shift of the maximum of the SCP equals 0.5 eV. As noted above, a potential barrier for holes exists at the Al–GCS interface. According to photoemission measurements,[5] the magnitude of this barrier equals approximately 0.6 eV for the present compositions. The existence of a shift of the maximum of the spectral sensitivity to higher photon energies with a positive voltage on SnO_2 is an additional confirmation of the existence of a potential barrier. The higher the energy of the absorbed photons, the shorter their penetration depth is, the closer the absorption of most photons to the surface barrier is, and the higher the average energy of the photogenerated holes is. For this reason, more holes capable of overcoming the potential barrier appear as the photon energy increases. The result is that the maximum of the SCP with a positive voltage shifts into a shorter wavelength region of the spectrum relative to the maximum with

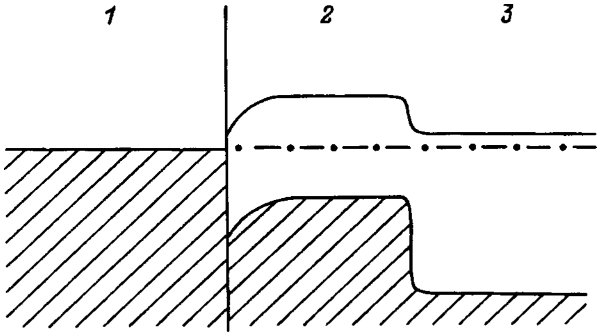

FIG. 2. Band diagram of the heterojunction: *1* — Al, *2* — GCS, *3* — SnO_2.

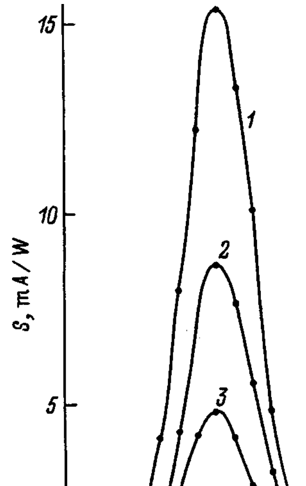

FIG. 4. Spectral characteristics of a $As_2(Se_{0.9}Te_{0.1})_3$ heterostructure: *1* — 8*B*, *2* — 4*B*, *3* — 2*B*.

a negative voltage. In the case of a negative voltage on SnO_2, the photogenerated holes do not have to overcome the potential barrier indicated above, so that in this case the position of the photosensitivity maximum is determined by the band gap in GCS.

CONCLUSIONS

1. Samples based on $SnO_2-As_2(Se_{0.9}Te_{0.1})_3$ and $SnO_2-(As_{0.67}Sb_{0.33})_2Se_3$ heterojunctions were fabricated and investigated.

2. The spectral and current–voltage characteristics were measured.

3. The current–voltage characteristics are nonlinear and asymmetric.

4. The region of spectral sensitivity was determined, and the maximum sensitivity was found for both types of heterojunctions.

5. It was found that for positive polarity the maximum of SCP is shifted to higher photon energies.

[1] V. Venkataraman, Curr. Sci. **64**, 855 (1994).

[2] I. Yu. Dzyurkevich and K. P. Kornev, in *Methods and Apparatus for Magnetic Measurements and Monitoring* [in Russian], Omsk Polytechnical Institute Press, Omsk (1977).

[3] H. S. Soni, S. D. Sathage, and A. P. B. Sinka, Indian J. Pure Appl. Phys. **21**, 197 (1983).

[4] Sh. Sh. Sarsembinov, O. Yu. Prikhod'ko, M. Zh. Mal'tekbasov *et al.*, Fiz. Tekh. Poluprovodn. **25**, 564 (1991) [Sov. Phys. Semicond. **25**, 341 (1991)].

[5] A. S. Kochemirovskiĭ and K. P. Kornev, in *Glassy Semiconductors* [in Russian], Physicotechnical Institute Press, Leningrad (1985), 80–81 pp.

Translated by M. E. Alferieff

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Self-propagating high-temperature synthesis and solid-phase reactions in bilayer thin films

V. G. Myagkov, V. S. Zhigalov, L. E. Bykova, and V. K. Mal'tsev

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

(Submitted May 13, 1997)
Zh. Tekh. Fiz. **68**, 58–62 (October 1998)

Self-propagating high-temperature synthesis (SHS) in Al/Ni, Al/Fe, and Al/Co bilayer thin films is investigated. It is established that SHS is achieved in thin films at initiation temperatures 300–350° lower than in powders. The mechanism of SHS in thin films is similar to the process of explosive crystallization. It is shown that at the initial stage solid-phase reactions arising on the contact surface of condensate films can be self-propagating high-temperature synthesis. SHS could find application in different technologies for obtaining film components for microelectronics. © *1998 American Institute of Physics.* [S1063-7842(98)01010-1]

Self-propagating high-temperature synthesis (SHS) is widely used to obtain many different compounds. Ordinarily, the reagents participating in SHS are in a powdered form or one reagent is in a gaseous state.[1,2] Infrequently, SHS has been studied in bimetallic systems[3] and foils.[4] In both cases, the reagents were several microns in size. SHS in thin-film samples (thickness up to 200 nm) has not been investigated at all.

The present work is devoted to the experimental study of the characteristic features of SHS in thin films and their differences from SHS in powders.

The following pairs were chosen as the initial materials for synthesis: Ni–Al, Fe–Al, and Co–Al. A layer of one of the ferromagnetic metals $M=$ Ni, Fe, and Co was deposited successively by thermal evaporation and an Al layer was deposited on top. Each layer was 30–100 nm thick. Films with layers of equal thickness were used for the investigations; this corresponds to ~40% at.Al in the sample. Mica or glass, 0.1–0.2 mm thick with linear dimensions of 5–10 mm, was used as the substrate. The two-layer system obtained was placed on a tungsten heater and heated at a rate of the order of 20° per second (thermal explosion). The vacuum during deposition and heating was 1×10^{-4} Pa. At the temperature T_{1i} a nucleus of a phase of the reaction products appeared and propagated with the velocity $V\sim0.5\times10^{-2}$ m/s until it covered the entire surface of the film. The morphology of the surface of the new phase changed, and the reflection from the film surface became dull and differed sharply from the mirror surface of the initial sample, so that the motion of the new phase was easily observed visually (Fig. 1a). The initiation temperature T_{1i} in the experiments was not constant; it depended on the rate of heating and the ratio of the thicknesses of each film and varied in the range 250–400 °C for the systems Co–Al and Fe–Al and 200–300 °C for Ni–Al. If t is the characteristic reaction time, then the velocity of the front is $V\sim\sqrt{\chi/t}$. In this time the diffusion will extend to the thickness of the film $d\sim\sqrt{Dt}$. Taking average experimental values $V=1\times10^{-2}$ m/s, $d=1\times10^{-7}$ m, and thermal diffusivity $\chi=10^{-5}-10^{-6}$ m^2/s (Ref. 5), we estimate the diffusion coefficient as $D=10^{-12}-10^{-13}$ m^2/s. Such values of D are characteristic for the Al diffusion in Fe, Co, or Ni at temperature 1200–1400 K.[6] This temperature agrees with the results of direct measurements, presented below, of the temperature of an Al/Fe film. Hence it follows that at these temperatures Al on the front is in a liquid state, while the lower Fe, Co, or Ni layers are in a solid phase. The reaction front is convex (Fig. 1a), because of the fact that the heat losses at the edge of the film are greater than at the center. They decrease the front temperature T_f and the front velocity from center to edge. On the other hand, the front temperature is higher than the melting temperature of aluminum T_m (Al). The arrow in Fig. 1a shows the liquid zone of aluminum, bordering the reaction front and possessing a high reflectance, so that it is distinguished from the initial and reacted parts of the film. If, after initiation, the substrate temperature is made to be less than T_{1i}, then quenching of the reaction occurs.

The self-maintaining character of the propagation of the nucleus is determined by the fact that the reaction zone lies at the interface of the initial film and the reaction products. Intense heat release on the front appreciably raises the temperature there. As a result of the Arrhenius temperature dependence of the diffusion coefficient the combustion process proceeds exclusively on the front. The proposed mechanism of SHS in thin films is similar to the process of autowave oxidation of metals[7,8] and the process of explosive crystallization.[9,10] The main characteristics of the processes are similar in the following cases: the existence of the reaction initiation temperature T_{1i}, self-maintained propagation of the front of a new phase, high temperature of the front, identical temperature dependences of the front propagation velocity, and possibility of the existence of a liquid zone on the front. The process of explosive crystallization has been well studied theoretically, so that it can be applied to the analysis of phenomena arising during SHS in thin films and autowave oxidation of metals. Figure 1b shows schematically the temperature profile on the reaction front, which, assuming that no phase transformations occur, is

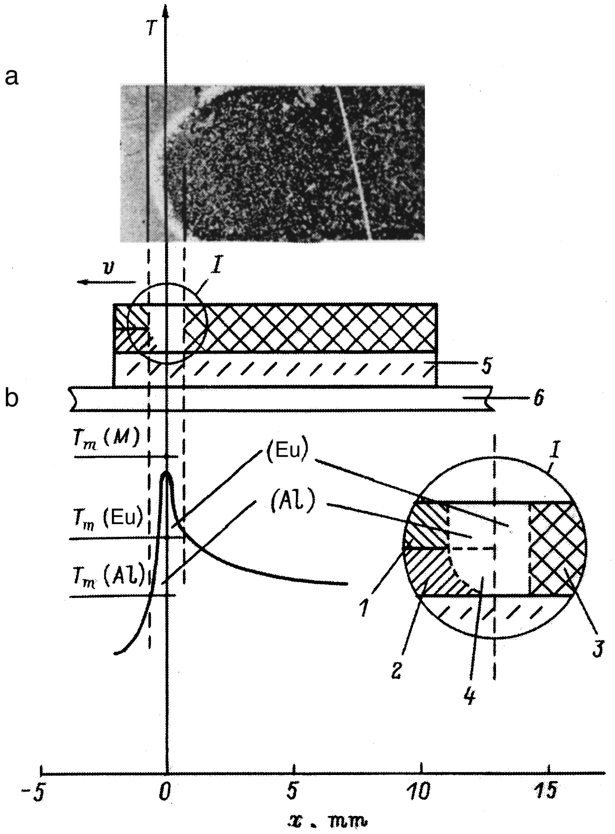

FIG. 1. Photograph, schematic illustration of SHS with a liquid aluminum zone, which is shown by the arrow, on the front (a), and temperature profile, perpendicular to the reaction front, with a liquid aluminum zone at the front (b): *1* — Al film; *2* — metal film $M=$ Fe, Co, Ni; *3* — reaction products; *4* — liquid zone on the SHS front; *5* — substrate; *6* — heater at temperatures $T>T_{2i}$.

exponential.[11] The front temperature T_f lies in the range $T_m(M)>T_f>T(\mathrm{Al})$, so that a liquid phase of aluminum exists at the front. Aluminum diffuses into the bottom metal layer ($M=$ Ni, Fe, Co), which is in the solid phase. If the lowest eutectic temperature $T_m(\mathrm{Eu})$ of the reaction products is less than the front temperature $T_f>T_m(\mathrm{Eu})$, then the liquid zone should include, besides liquid aluminum (Al), liquid reaction products (Eu). The width of the liquid zone depends on the temperature profile of the reaction front (Fig. 1b).

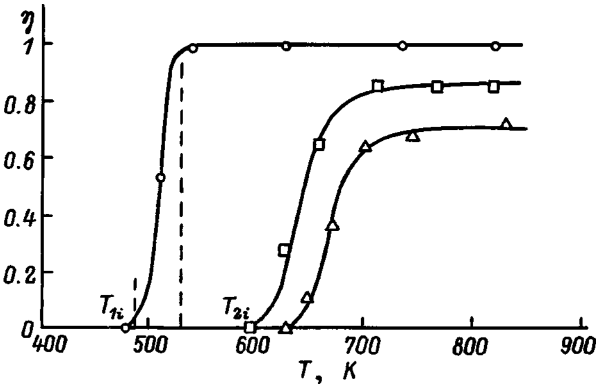

FIG. 2. Degree of transformation of bilayer systems as a function of substrate temperature: ○ — Al/Ni, □ — Al/Fe, △ — Al/Co; layer thicknesses: Al ~60 nm, M~50 nm.

The magnetic moment $M(T_l)$ of the sample was measured in the experiments. It is proportional to the volume of the ferromagnetic part of the film, depending on the substrate temperature. The degree of transformation $\eta(T_l)=(M(0)-M(T_l))/M(0)$ was determined assuming all intermetallic phases of Al with Ni, Co, and Fe to be nonmagnetic. Here $M(0)$ is the magnetic moment of the initial sample at room temperature and $M(T_l)$ is the magnetic moment after the substrate is heated to temperature T_l and held for 10 s — the time required for the combustion wave to propagate through the film. Figure 2 shows the dependence $\eta(T_l)$ for the systems Ni–Al, Co–Al, and Fe–Al. From the dependence $M(T_l)$ follows the existence of the synthesis onset temperature T_{1i} and temperature T_{2i} at which the degree of transformation has a maximum value. In Fig. 2 the temperatures T_{1i} and T_{2i} are marked only for Al/Ni films. Thus, at temperatures $T>T_{2i}$ nickel reacts completely with aluminum, and there is not enough time for 0–30% of the cobalt and 0–20% of the iron to react. Analysis of the surface morphology suggests that in the temperature range $T_{2i}>T>T_{1i}$ synthesis does not proceed to the entire depth, but covers a thickness *d* at the interface of the films. The thickness *d* increases rapidly with increasing substrate temperature, and at temperature T_{2i} synthesis proceeds to the entire depth. The large fraction of the surface where the reagents are in contact with one another decreases appreciably the initiation temperature T_{1i}. For the system Ni–Al this temperature is 300–350° lower than the corresponding temperature in powders.[11,12] In many technologies for obtaining thin-film coatings, the films cool down at substrate temperatures higher than the initiation temperature T_{1i}. This suggests that if the substrate temperature T_l is higher than the initiation temperature T_{1i} when the multilayer films are deposited, SHS can proceed and change the expected phase composition and structure of the samples. To confirm this conjecture, an Al layer (~50 nm thick) at different substrate temperatures was deposited on Co, Ni, Fe films (~50 nm thick) deposited on mica substrates. The degree η of transformation was determined as a function of the substrate temperature T_l by the method presented above. Figure 3 shows the dependence $\eta(T_l)$, which shows that the SHS process is initiated during deposition of the upper layer

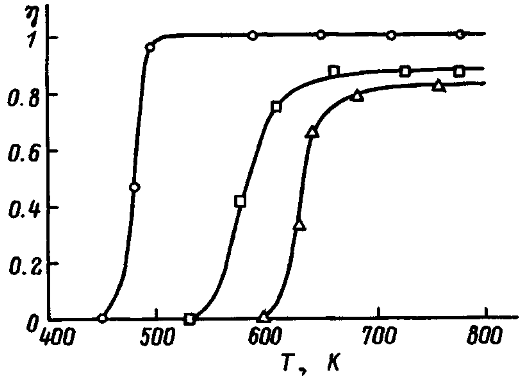

FIG. 3. Degree of transformation versus substrate temperature after deposition of an Al layer on a film $M=$ Ni, Fe, Co: ○ — Al/Ni, □ — Al/Fe, △ — Al/Co.

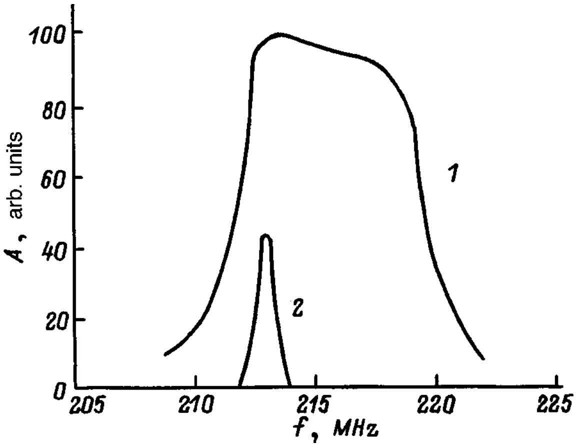

FIG. 4. Frequency spectra of nuclear spin echo of a bilayer film system Al/Co: *1* — initial sample, *2* — after passage of a SHS wave through the sample.

of aluminum. In addition, the initiation temperatures of the systems Ni–Al, Fe–Al, and Co–Al are close to the corresponding temperatures during heating of bilayer films of the same systems (Fig. 2).

Al/Co bilayer film systems were also investigated by the nuclear magnetic resonance method. Figure 4 shows the frequency spectra of the nuclear spin echo (NSE) of the initial films and of the same samples after annealing, where SHS was realized. The typical Al/Co spectrum of the initial samples is characteristic for polycrystalline Co and is formed by two of its allotropic modifications: the low-temperature hexagonal close-packed (hcp) α-Co phase and the high-temperature face-centered cubic (fcc) β-Co phase (the $\alpha \leftrightarrows \beta$ transition temperature $\sim$700 K) with central frequency 213 MHz. After passage of the SHS wave the line shape of the spectrum changes appreciably, and only the signal from the fcc phase remains. The reaction products experience high rates of cooling, which results in, specifically, stabilization of the high-temperature fcc phase of Co. Thus, taking the rise in temperature in the combustion wave $\Delta T \sim 1000$ K and the characteristic reaction time $t=\chi/v^2 \sim 0.1-0.01$ s we estimate the cooling rate as $\sim 10^4-10^5$ K/s. Such cooling rates are sufficient not only for fixing the high-temperature phases but also for obtaining an amorphous phase in the alloys. The absence of satellite lines in the low-frequency region of the spectrum shows that after the passage of a SHS wave a solid solution of aluminum in cobalt is not formed, but rather intermetallic compounds are formed. The area under the curve of the frequency spectrum determines the amount of cobalt present in the sample. The degrees of transformation, determined from the ratio of the areas of the spectra before and after the reaction and from magnetic measurements, agree well with one another.

The phase composition of the samples after the passage of a SHS wave was investigated by the methods of x-ray crystallographic analysis. The samples investigated had layers of equal thicknesses, each layer being $\sim$100 nm thick. The diffraction patterns the samples of the system Al/Ni where SHS occurred as a result of annealing at temperature $T > T_{2i}$ are completely identical to the diffraction patterns of samples where SHS occurred with an aluminum film deposited on a nickel film at temperature $T > T_{2i}$. The interpretation of the diffraction patterns shows the presence of a main phase Ni_2Al_3 with a small quantity of a NiAl phase and show residual aluminum and nickel to be absent, in agreement with the data presented in Figs. 2 and 3. For Al/Fe films where SHS occurred as a result of annealing the diffraction patterns show reflections from residual α-Fe as well as the high-temperature phase γ-Fe and the intermetallic compounds $FeAl_3$ and Fe_2Al_5. The high-temperature phase γ-Fe was stabilized just as β-Co as a result of the high rates of cooling which arise on the combustion front. Similar diffraction patterns of Al/Co samples, just as their magnetic measurements and NMR spectra, confirmed the presence of residual β-Co. In addition, other phases, which can be identified as Al_5Co_2 and AlCo, are present in the sample.

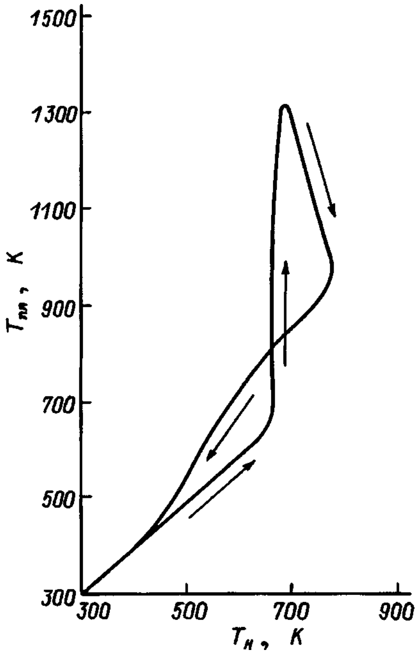

FIG. 5. Temperature of a bilayer film sample Al/Fe versus the heater temperature. The arrows show the path of the temperature as the sample is heated up to 770 K and then cooled.

The film temperature the during passage of a SHS wave was measured in the system Al/Fe. To this end, a Pd layer (50 nm thick) was deposited on the substrate and Fe and Al layers (each layer 50 nm thick) were deposited successively on the Pd layer. The Pd and Fe layers were used as thermocouples. The sample was placed on a heater whose temperature was determined with a standard chromel–copel thermocouple and heated at rate 10 K/s. The heater was switched off at temperature 770 K, which is higher than the initiation temperature T_{1i}. Figure 5 shows the dependence of the sample temperature measured with the film thermocouple on the heater temperature. Before the temperature T_{1i} was reached, both thermocouples showed the same temperature. This signified that the heater temperature was identical to the film temperature. However, above $T_{1i}=660$ K the film temperature increased rapidly to a value of the order of 1330 K as a result of an exothermic reaction. This value agrees well with the estimate of the front temperature made earlier. After the passage of the SHS wave, the temperature indicated by the film thermocouple dropped. However, the reverse path of the film temperature was different from the forward path and was higher than the heater temperature, which indicated continuation of the exothermic reaction. The reaction stops only at temperatures 300–350 K, and the film and heater temperatures become the same.

To determine the total reaction time the electrical resistance of the samples during the reaction at temperature T_{1i} was measured. The electrical resistance increases continuously as the temperature increases from room temperature to T_{1i}. This increase is typical for the temperature dependence of the resistance of metals. When the temperature reaches the initiation value T_{1i} the electrical resistance of the sample increases rapidly — for 20 min for Al/Co systems and for 10 min for Al/Ni, Al/Fe film systems. Experiments to determine the sample temperatures and electrical resistance show that after the passage of a SHS wave through the film (for the present samples not more 10 s), the further heat release in the film is due to a post-combustion process. Post-combustion processes have been observed in previous works on SHS.[2] Hence it follows that the reaction process consists of two stages. During the first stage a fast autowave combustion process occurs. It results in appreciable mixing of the reagent layers and formation of reaction products. The second stage is slower. At this stage the initial products that have not completely reacted in the first stage undergo post-combustion. This stage can be accompanied by recrystallization and formation of new phases. During the second stage the combustion process is not a wave process, but rather the reaction products are produced by a process of nucleation and growth, whose kinetic law is described by the Kolmogorov–Avrami–Johnson–Mehl equation.[13]

The low initiation temperatures of SHS in bilayer film systems suggest that many solid-phase reactions in the initial stage at the interface of film condensates are self-propagating high-temperature synthesis. Thus, in Ref. 14 solid-phase reactions were investigated in Ni/Al multilayer films. Similarly to the present work, solid-phase Ni/Al reactions in multilayers materialize both as a result of annealing and during combined deposition of nickel and aluminum films. The fact that in our work and in Ref. 14 the initiation temperatures of the reactions $T_{1i}\sim 500$ K are the same suggests that SHS in multilayer films was also observed in the work mentioned. For this reason, it should be expected that other solid-phase reactions arising at the contact surface of film condensates can initially be self-propagating high-temperature synthesis. This is the reason for the search for solid-phase reactions that could be realized in bilayer film samples. They should occur in systems where SHS has been obtained.[1] Conversely, for films on whose boundary solid-phase reactions arise (see, for example, Ref. 15) SHS should be expected to occur in the corresponding powders.

The following conclusions can be drawn on the basis of the initial experiments investigating SHS in thin films: 1) SHS materializes in thin films at initiation temperatures 300–350° lower than in powders; 2) the mechanism of SHS in thin films is similar to explosive crystallization; 3) SHS in thin films can also be obtained on the surface of powders if a second reagent in a liquid phase is present; 4) the possibility of SHS initiation should be taken into account in existing technologies for obtaining thin-film coatings; 5) different solid-phase reactions observed at the interface of bilayer and multilayer films and arising at low heat-treatment temperatures can initially be self-propagating high-temperature synthesis; 6) the high rates of cooling after the passage of a SHS wave in bilayer film samples can stabilize high-temperature and metastable phases; and, 7) SHS could find application in different technologies for obtaining film components for microelectronics.

This work was supported by the Russian Fund for Fundamental Research, Grant No. 96-32327a/410.

[1] A. G. Merzhanov, in *Physical Chemistry* [in Russian], edited by Kolotyrkin, Khimiya, Moscow (1983), pp. 6–45.

[2] A. G. Merzhanov and I. P. Borovinskaya, Dokl. Akad. Nauk SSSR **204**, 366 (1972) [Dokl. Chem. **201**, 429 (1972)].

[3] S. G. Vadchenko, A. M. Bulaev, Yu. A. Gal'chenko, and A. G. Merzhanov, Fiz. Goren. i Vzryva, No. 6, 46 (1987).

[4] U. Anselmi-Tambuini and Z. A. Munir, J. Appl. Phys. **66**, 5039 (1989).

[5] I. S. Grigor'ev and E. Z. Meĭlikhov [Eds.], *Physical Constants* [in Russian], Énergoatomizdat, Moscow (1991), 1232 pp.

[6] V. N. Larikov and V. I. Isaĭchev, *Structure and Properties of Metals and Alloys. Diffusion in Metals and Alloys* [in Russian], Naukova Dumka, Kiev (1987), 511 pp.

[7] V. G. Miagkov, L. I. Kveglis, G. I. Frolov, and V. S. Zhigalov, J. Mater. Sci. Lett. **13**, 1284 (1994).

[8] V. G. Myagkov and N. V. Baksheev, Pis'ma Zh. Tekh. Fiz. **18**(6), 14 (1992) [Sov. Tech. Phys. Lett. **18**(3), 174 (1992)].

[9] V. A. Shklovskiĭ and V. M. Kuz'menko, Usp. Fiz. Nauk **157**, 311 (1989) [Sov. Phys. Usp. **32**, 163 (1989)].

[10] G. H. Gilmer and H. J. Leamy, in *Laser and Electron-Beam Processing of Materials*, Academic Press, New York (1980), pp. 227–232.

[11] K. A. Philpoh, Z. A. Munir, and J. B. Holt, J. Mater. Sci. **22**, 159 (1987).

[12] C. Michaelsen, G. Lucadamo, and K. Barmak, J. Appl. Phys. **80**, 6689 (1996).

[13] R. W. Cahn and P. Haasen [Eds.], *Physical Metallurgy*, (North-Holland, New York, 1983, Vol. 2; Mir, Moscow, 1968), 490 pp.

[14] E. O. Coldan, C. Cabral, Kotecki Jr. *et al.*, J. Appl. Phys. **77**, 614 (1995).

[15] L. Persson, M. El. Bouanani, and M. Hult *et al.*, J. Appl. Phys. **80**, 3347 (1996).

Translated by M. E. Alferieff

Fast nonequilibrium induction detectors based on thin superconducting films

I. G. Gogidze, P. B. Kuminov, A. V. Sergeev, A. I. Elant'ev, E. M. Men'shchikov, and E. M. Gershenzon

Moscow Pedagogical State University, 119435 Moscow, Russia

(Submitted September 8, 1997)
Zh. Tekh. Fiz. **68**, 63–69 (October 1998)

A new type of fast detector is proposed. The operation of the detector is based on the change induced in the kinetic inductance of NbN and $YBa_2Cu_3O_{7-\delta}$ superconducting films by nonequilibrium quasiparticles produced by electromagnetic radiation. The speed of a NbN detector is essentially temperature-independent and is less than 1 ps. A model based on the Omen–Scalapino scheme describes well the experimental dependence of the voltage–power sensitivity of a NbN detector on the modulation frequency of the radiation. A low equilibrium quasiparticle density and a high quantum yield give detecting power $D^*=10^{12}\,W^{-1}\cdot cm\cdot Hz^{1/2}$ at temperature $T=4.2$ K and $D^*=10^{16}\ W^{-1}\cdot cm\cdot Hz^{1/2}$ at temperature $T=1.6$ K. The time constant of the low-temperature YBaCuO induction detector is determined only by the electron–phonon interaction time τ^d_{e-ph} in the nodal regions. © *1998 American Institute of Physics.* [S1063-7842(98)01110-6]

INTRODUCTION

The investigation and development of fast devices whose operation is based on nonequilibrium processes occurring in both conventional and high-T_c superconductors (HTSCs) is one of the important directions of modern superconducting electronics. Picosecond resistance detectors consisting of low-temperature superconducting thin films with high sensitivity operate similarly to bolometers in the range from millimeter wavelengths to visible light.[1,2] Similar HTSC detectors can be another order of magnitude faster, but they are less sensitive.[3,4]

Resistance detectors employing as the working component thin films of conventional superconductors with a short electron mean free path are described well by a nonequilibrium two-temperature heating model.[5] This model also explains the basic forms of the photoresponse of HTSCs, specifically, the two-component decay of the photoresponse in HTSCs has the same nature as in conventional superconductors with a high superconducting transition temperature.[6,7]

In superconducting thin films the hierarchy of characteristic relaxation times is determined by the distribution functions of the nonequilibrium quasiparticles and phonons under the action of radiation, i.e., as quasiparticles and phonons respond to a short laser pulse.[8] The response mechanism is determined by the ratio of the times of relaxation processes determined by phonons. At low temperatures the time τ_{es} in which phonons escape from a thin film into the substrate is shorter than the phonon–electron scattering time τ_{ph-e}, which is a necessary condition for the appearance of a purely nonequilibrium response.[9] The ratio of τ_{es} and τ_{ph-e} changes as the working temperature increases in the range from liquid-helium to liquid-nitrogen temperatures. Moreover, the form of the photoresponse depends strongly on the ratio of the electron–phonon interaction time τ_{e-ph} and the phonon–electron scattering time τ_{ph-e}. At liquid-helium temperatures τ_{e-ph} is greater than τ_{ph-e}. In this case, if $\tau_{es}\ll\tau_{ph-e}$, then a purely nonequilibrium response with a characteristic decay time τ_{e-ph} materializes, while for $\tau_{es}\gg\tau_{ph-e}$ the response is exclusively bolometric with a corresponding decay time τ_{es}. For thin films of conventional superconductors with a high critical temperature, the ratio of τ_{e-ph} and τ_{ph-e} changes at $T=5$–10 K, and at higher temperatures the response becomes multicomponent.

The development of nonequilibrium HTSC resistance devices has now encountered fundamental difficulties. The obstacle is the high thermal resistance at the film–substrate boundary for all substrates suitable for growing high-quality films. As a result, the heat removal from the film is inadequate even for ultrathin films, which makes it impossible to eliminate the bolometric component of the detector response.

The solution of the problem must be sought not in the resistive state of the film but rather in the superconducting state, i.e., not in the change in the resistance of the sensitive element but rather in the change of the kinetic inductance. In this case, not only is heat not released during the flow of the bias current, but, and most importantly, the appearance of a nonequilibrium response at low temperatures in superconductors with *s* pairing decreases the detector time constant τ_D, while in superconductors with nodal regions, i.e., with *d* pairing, it changes the hierarchy of relaxation times and makes it possible to realize a phonon thermostat, i.e., to eliminate the bolometric effect.

Of the works concerning kinetic inductance in conventional superconductors, we call attention to Refs. 10–14. In these works, the first attempts were made to employ the inductance for applied purposes: radiometers, SQUID-based low-noise amplifiers, and so on. The authors noted that, in their opinion, near the critical temperature T_c the thermal coupling of nonequilibrium particles and the lattice is very good and the film is heated as a whole, i.e., the inductive response is of a purely bolometric character. Appreciably

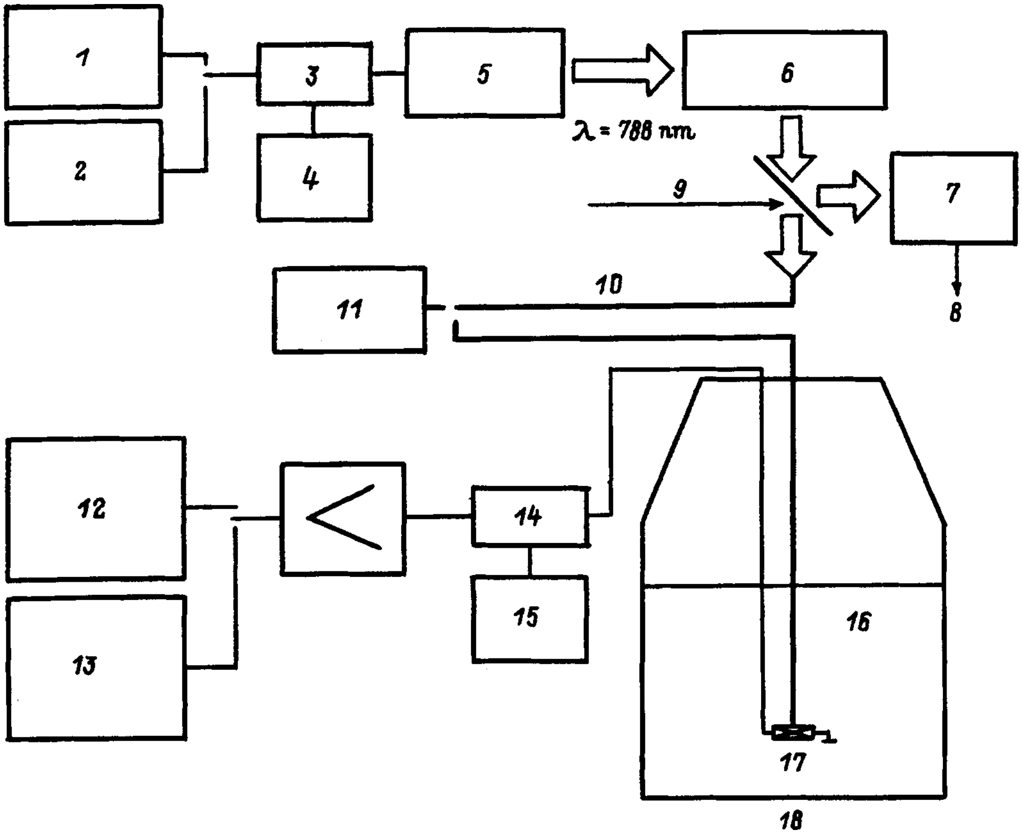

FIG. 1. Block-diagram of the experimental apparatus: *1* — G4-151 generator (1–500 MHz); *2* — Ya2R-74 generator (0.01–2 GHz); *3,14* — adapter; *4,15* — power supply; *5* — RLD-78 M20 laser diode; *6 — focusing system; 7* — LFDG-70 avalanche photodiode; *8* — to spectrum analyzer; *9* — beam splitter; *10* — 50 μm in diameter fiber-optic cable; *11* — radiation power meter; *12* — SKA-59 spectrum analyzer (0.01–110 MHz); *13* — SKA-60 spectrum analyzer (0.01–12 GHz); *16* — liquid helium; *17* — sample; *18* — transport Dewar vessel.

below T_c the situation changes: Electromagnetic radiation increases the temperature of only the quasiparticles without affecting the lattice, and the response can be explained only by the nonequilibrium superconductivity.

Most research groups studying the inductive response of HTSC films now believe that only a nonequilibrium response exists in the superconducting state. To explain the mechanism of such response a theory developed a long time ago[15] for the deep superconducting state ($T \ll \Delta$) of films of conventional superconductors, which in principle is not applicable to HTSC materials,[16] is invoked.

Investigations of relaxation processes in thin superconducting films are of great general-scientific value and extremely important for designing not only induction detectors but also other components of superconductor electronics which operate under nonequilibrium conditions, such as devices based on tunneling injection, superconducting transistors, and others.

Our objective in the present work is to study the inductive photoresponse of NbN and $YBa_2Cu_3O_{7-\delta}$ thin films in the superconducting state $T \ll \Delta$ and to determine the basic characteristics of nonequilibrium induction detectors.

EXPERIMENTAL SAMPLES, PROCEDURE, AND TECHNIQUE

Thin films of $YBa_2Cu_3O_{7-\delta}$ and niobium nitride NbN were chosen as the objects of investigation. YBaCuO films are the best-known HTSC materials: The fabrication technology is well developed, and a relatively large amount of knowledge about the main physical properties has been accumulated in the course of investigations over many years. Our samples on $LaAlO_3$ substrates consisted of 50 nm thick films with critical temperature 88 K in a structure in the form of a single bridge or three parallel strips with a total area of 10×10 μm. NbN films with a high critical temperature and a thickness of several tens of angstroms were needed for the experimental investigations. Our technology group solved this problem. The result were less than 10 nm thick NbN films on a sapphire substrate. The transition temperature T_c of the films was controllable with reproducibility better than 0.5 K in the temperature range 4.5–14.7 K.[17] The NbN samples had the structure of a meander (strip width $w = 1.5$ μm) filling an area of 54×72 μm.

A block diagram of the experimental apparatus for measuring the frequency dependence of the signal obtained by applying to the superconducting sample amplitude-modulated near-IR electromagnetic radiation is presented in Fig. 1. A RLD-78M20 laser diode with wavelength $\lambda = 788$ nm, connected through a high-frequency adapter to the power supply, served as the source of radiation. The laser had maximum radiation power $P = 10$ mW with bias current $I = 65$ mA. The laser radiation, modeled by a generator of sinusoidal signals, was fed into a 50 μm in diameter G1-50/125-AXAI fiber-optic cable by means of a short-focal length objective lens. Another such cable brought radiation up to the experimental sample. The cables were connected through an optical reducer, which made it possible to measure, using a RIFOCS-555B power meter, the power of the radiation brought up to the sample. A beam-splitter after the focusing system diverted part of the radiation to a LFDG-70 avalanche photodiode with a measuring frequency band up to 6 GHz, controlling the amplitude–frequency characteristic of the laser diode. The experimental sample was placed at the end of a holder placed in a liquid-helium Dewar vessel. The signal from the sample was fed through flexible coplanar and stiff coaxial lines into the detecting apparatus, which consisted of the adapter of the power supply for the sample, an amplifier, and SK4-59 and SK4-60 spectrum analyzers or a EGG-5202 phase-sensitive device. The latter was used when the problem of measuring very low power signals, $P_s \sim -150$ dBm, in the band of detected frequencies $f = 1$–50 MHz, arose.

Measurements of the relaxation times of the response of thin superconducting films by the modulation method are more accurate than measurements performed by pulsed methods. Signal detection occurs in a narrow band of detection frequencies, which depends on the stability of the generator of the modulation of the laser power and the minimum input band of the spectrum analyzer (or other device). Signals with power $\sim 50\times 10^{-18}$ W have been detected in the synchronous detection regime, while the pulsed method has made it possible to measure signals with peak power only down to $\sim 50\times 10^{-9}$ W.[18]

The registration times of the impedance of the experimental samples were determined from the amplitude–frequency characteristics. Since the relaxation times of the response, in general, vary according to an exponential law, the frequency characteristic of the signal voltage per unit intensity is similar to the transmission characteristic of a low-frequency filter

$$\Delta U(f) = U_0 \frac{2\pi f \tau}{\sqrt{1+(2\pi f \tau)^2}}, \tag{1}$$

where $\Delta U(f)$ is the signal level at frequency f, U_0 is the signal level at zero frequency, and τ is the characteristic time, also called the time constant of the detector.

EXPERIMENTAL RESULTS AND DISCUSSION

a) *Nonequilibrium NbN induction detector.* In this section we propose a superconducting nonequilibrium detector whose operation is based on the radiation-induced change in the kinetic inductance of the superconducting condensate. We note that the inertial bolometer proposed previously[13,19] employs the temperature dependence of the kinetic inductance near T_c. We shall investigate a fast detector whose operation is based on the nonequilibrium response of thin films at a temperature much less than T_c. The time constant τ_D of such a detector is determined by the development time of an electron avalanche, which is much shorter than τ_{e-ph}. Moreover, the detecting power of this detector is much greater than that of electronic resistance bolometers with comparable values of τ_D.

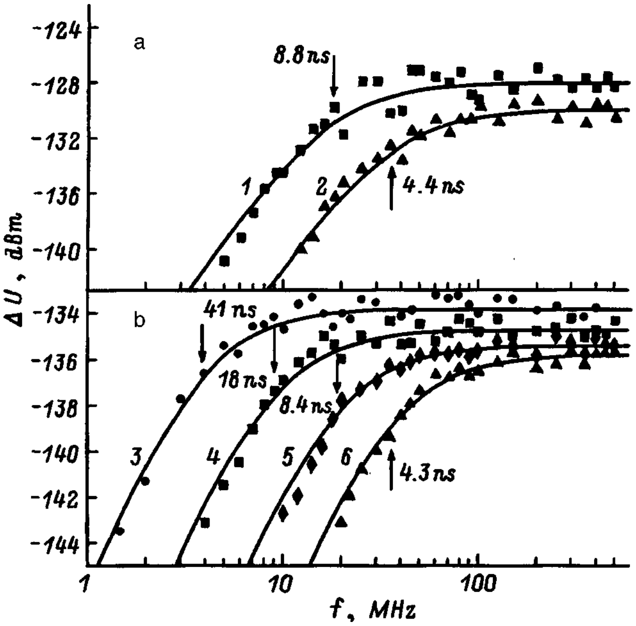

FIG. 2. Inductive signal ΔU versus modulation frequency for 5 (a) and 10 nm (b) thick NbN films: T/T_c: *1* — 0.34, *2* — 0.39, *3* — 0.29, *4* — 0.34, *5* — 0.4, *6* — 0.45.

Figure 2 shows the signal voltage versus the modulation frequency f for two samples at different temperatures. The signal amplitude is approximated well by the dependence 1. The values of τ are marked in Fig. 2 by arrows. For a constant reduced temperature T/T_c the values of τ are proportional to the film thickness. For each sample, at low temperatures τ is an exponential function of temperature. The signal magnitude on the plateau of the frequency curve grows weakly as $T^{-1/2}$ with decreasing temperature.

The experimental results presented can be explained on the basis of the following model, which is based on the Omen–Scalapino scheme.[20] When light with photon energy $h\nu$ greater than the gap (Δ) is absorbed, Cooper pairs are broken and quasiparticles are formed. This changes the kinetic inductance L of the film and therefore results in the appearance of a voltage equal to IdL/dt. The rate of change of the inductance is determined by the modulation frequency of the radiation and relaxation processes in the film. At first, the photoexcited quasiparticles lose energy, emitting phonons with energy greater than 2Δ, which in turn destroy Cooper pairs and produce new quasiparticles. For quasiparticle energy less than 3Δ, further quasiparticle multiplication becomes impossible. Next, the nonequilibrium quasiparticles cool down, emitting low-energy phonons, and accumulate in a level with energy Δ. At this stage, which lasts no longer than 1 ps, the development of the electron avalanche is completed and the further, relatively slow, evolution results in a decrease of the nonequilibrium quasiparticle density. When two quasiparticles recombine, a phonon with energy 2Δ is emitted, and this phonon can either leave the film in a time $\tau_{es} = 4d/\alpha u_s$ (u_s is the sound speed in the substrate and α is the transmittance of the film–substrate boundary) or destroy a Cooper pair in a time τ_B. As a result, the effective quasiparticle lifetime is

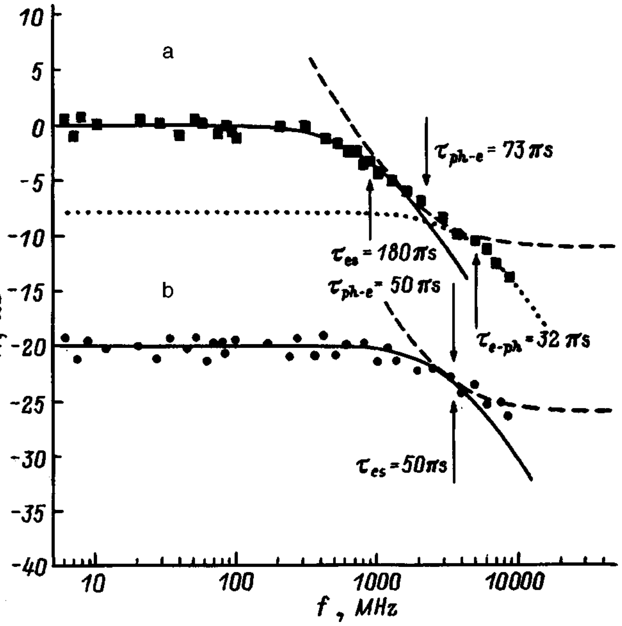

FIG. 3. Amplitude–frequency characteristics of the photoresponse of NbN thin films in the resistive state: a — 20 nm thick sample, $T_c = 6.5$ K, $\Delta T = 0.4$ K, $T = 6.1$ K; b — 5 nm thick sample, $T_c = 12.4$ K, $T = 12.3$ K.

$$\tau_l = \tau_R\left(1 + \frac{\tau_{es}}{\tau_B}\right). \tag{2}$$

Following Ref. 16, the voltage–power sensitivity can be represented in the form

$$S = \frac{LI\exp(-\Delta/k_BT)}{n_{\text{eq}}V\Delta}\,\frac{2\pi f\tau_l}{\sqrt{1+(2\pi f\tau_l)^2}} \times \frac{1}{\sqrt{1+(2\pi f\tau^*)^2}}, \tag{3}$$

where τ^* is the development time of an electron avalanche.

We shall now estimate for the thinnest film the numerical values of the characteristic times appearing in Eqs. (2) and (3). The recombination time at temperatures much less than T_c can be determined in terms of the electron–phonon relaxation time at the critical temperature $\tau_{e-ph}(T_c)$:

$$\frac{\tau_R(T)}{\tau_{e-ph}(T_c)} = 0.1\left(\frac{T_c}{T}\right)^{1/2}\exp\left(\frac{\Delta}{k_BT}\right). \tag{4}$$

To estimate the recombination time we measured the escape time of nonequilibrium phonons from a 5 nm thick NbN thin film and obtained $\tau_{es} = 50$ ps (Fig. 3b). Since the electron–phonon interaction time is essentially independent of film thickness, we employed a 20 nm thick film with $T_c = 6.5$ K to determine it. Decreasing the working temperature increased the electron–phonon interaction time, making it possible to measure it directly in the frequency band of our apparatus (Fig. 3a). Taking account of the temperature dependence $\tau_{e-ph} \sim T^{-1.6}$ (Ref. 21) for a 5 nm thick film with $T_c = 12.4$ K we have $\tau_{e-ph}(T_c) = 10$ ps, which gives $\tau_R = 50$ ps with temperature $T = 4.2$ K. The time τ_B does not depend on temperature and can be estimated from the relation

$$\tau_B = \frac{u_s}{u_F}\,\frac{h}{2\pi\Delta}, \tag{5}$$

which gives $\tau_B = 2$ ps. Therefore $\tau_{es}/\tau_B \approx 20$, which indicates strong reabsorption of nonequilibrium high-energy phonons by quasiparticles. The quasiparticle lifetime calculated from Eq. (2) using the value of τ_R (Eq. (4)) agrees well with the experimental values of τ presented in Fig. 2. We note that strong reabsorption of phonons has a positive effect on the characteristics of the detector, since in this case the maximum quasiparticle multiplication factor $K = h\nu/2\Delta$ is reached as a result of an electron avalanche.

We shall now discuss the parameters of a nonequilibrium induction detector. To realize the detector we propose using the plateau of the frequency dependence in Fig. 2. Then the detector possesses an output frequency band Δf ranging from the reciprocal of the quasiparticle lifetime to the reciprocal of the development time of an electron avalanche. The avalanche development time has been estimated in optical experiments according to the modulation of the transmittance.[22] They showed that $\tau^* < 0.5$ ps.

The voltage–power sensitivity on the plateau is

$$S = \frac{LI\exp(-\Delta/T)}{n_{\text{eq}}V\Delta} \propto T^{-1/2}, \tag{6}$$

where n_{eq} is the equilibrium quasiparticle density, which decreases exponentially with decreasing temperature.

The main source of the detector noise is generation–recombination noise. The corresponding detecting power is

$$D^* = \sqrt{\frac{\tau_l}{4\Delta^2 n_{\text{eq}} d}}. \tag{7}$$

Taking the electron density $\sim 10^{22}\,\text{cm}^{-3}$ we obtain the equilibrium quasiparticle density at 4.2 K $n_{\text{eq}} \approx 10^{17}\,\text{cm}^{-3}$ and correspondingly D^* equals approximately $10^{12}\ \text{W}^{-1}\cdot\text{cm}\cdot\text{Hz}^{1/2}$.

b) *Nonequilibrium* $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ *induction detector*. The difficulty of studying the electronic kinetics in HTSC materials is due primarily to the structure of the order parameter. The existence of nodal regions, where the superconducting gap is small, makes the electronic scattering processes qualitatively different from the corresponding processes in conventional superconductors. As a result, the characteristic times in HTSC materials are much shorter, which makes nonequilibrium phenomena in them more complicated to detect, but more promising for applications in fast (picosecond) electronics.

In high-temperature superconductors at low temperatures the quasiparticles are concentrated in nodal regions, so that the recombination and relaxation times differ only by the coherence factors and are of the same order of magnitude. Both times increase only as a power law with decreasing temperature. The electronic relaxation time in the resistive state near the superconducting transition in YBaCuO films was determined in Ref. 23 and is $\simeq 2$ ps. Therefore the expected characteristic electronic times are shorter than 100 ps even at liquid-helium temperatures. Experimental investigations of the inductive response near a superconducting transition show that the kinetics of the electrons is essentially identical to the processes occurring in the resistive state.[18] Two-time relaxation is observed in this temperature range. At these temperatures the picosecond processes are due to electron–phonon interaction, while the nanosecond processes are due to the escape of phonons through the HTSC-film–substrate boundary (bolometric effect). The bolometric effect can be eliminated by lowering the temperature and using thinner films.

The results of recent experiments on the microwave response, NMR, and photoemission unequivocally demonstrate the existence of low-energy excitations in the nodal regions.[24] Investigations of the phase coherence of corner SQUIDs confirm the idea of d pairing,[25] and d symmetry also occurs from a model where pairing arises as a result of the exchange of antiferromagnetic spin fluctuations.[26] On account of the presence of nodal lines, the low-temperature behavior of a superconductor is very sensitive to impurities. Without electron–impurity scattering the quasiparticle density of states near the Fermi surface is a linear function of energy. Even low impurity concentrations lead to a constant quasiparticle density of states, which depends on the electron–impurity potential.[27]

Since there are no features in the density of states, the quasiparticle recombination time does not growth exponentially at low temperatures. For this reason, the description of the photoresponse by the Omen–Scalapino model[20] is not justified. As already noted above, experiments near the transition show a very strong electron–electron interaction. Assuming that the electron–electron interaction prevails over the electron–phonon interaction in the superconducting state also, the kinetic scheme with a nonequilibrium electron temperature can be used.

The amplitude–frequency characteristics of the inductive signal for two different working temperatures are shown in Fig. 4. As one can see, at temperature $T=4.2$ K the amplitude of the experimental signal grows linearly with the modulation frequency of the incident laser power in the entire band of measuring frequencies (Fig. 4a). As working temperature increases ($T=25$ K), a plateau appears in the amplitude–frequency characteristic of the signal after the modulation frequency of the radiation power $f=30.8$ MHz with characteristic time $\tau \simeq 5.2$ ns (Fig. 4b).

Low-temperature investigations of the experimental samples, for which $\tau_{es} \simeq 52 \pm 0.2$ ns,[28] show the following: at working temperature $T=4.2$ K, up to the modulation frequencies of the electromagnetic radiation $f=50$ MHz (instrumental time $\tau=3.18$ ns), the amplitude–frequency characteristic of the sample has a rising form, i.e., the signal amplitude is directly proportional to the modulation frequency of the incident laser power. If the response had exhibited two-component relaxation, then even with $f=30.8$ MHz we would have observed a plateau in the AFC in our band of measuring frequencies ($\Delta f=50$ MHz).

Our experimental results confirm that a purely nonequilibrium response in HTSCs materializes at low temperature, i.e., there is no bolometric effect at working temperatures far below the superconducting transition, and they give a basis for the possible development of fast nonequilibrium induction detector on the basis of YBaCuO thin films.

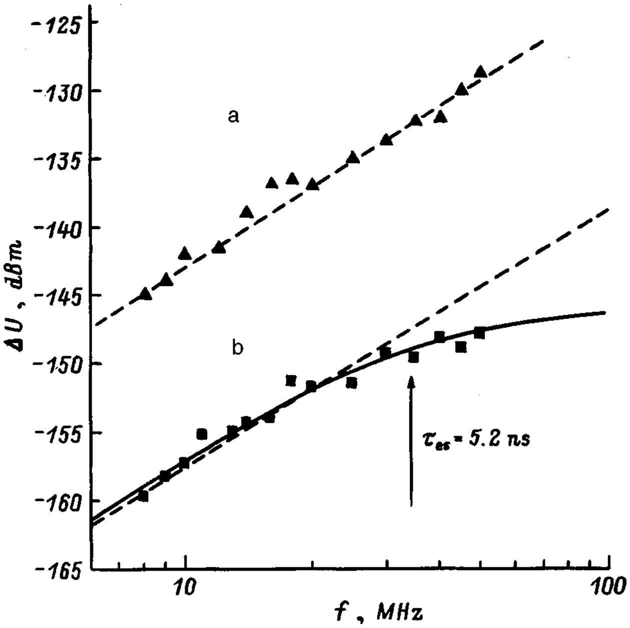

FIG. 4. Amplitude–frequency characteristics of the inductive signal of a $YBa_2Cu_2O_{7-\delta}$ detector at 4.2 (a) and 25 K (b).

We shall now discuss the main parameters of the YBaCuO induction detector, such as the time constant τ_D and the detecting power D^*. The kinetic scheme with a nonequilibrium electron temperature in a HTSC film far from the superconducting transition presupposes that the time constant of the YBaCuO detector equals the electron–phonon interaction relaxation time τ_{e-ph}^d in the nodal regions. At temperature $T=4.2$ K $\tau_{e-ph}^d \simeq 15$ ps.[23] Then at temperatures near $T=10$ K the time constant of a YBaCuO nonequilibrium induction detector is $\tau_D \simeq 10$ ps.

Following Ref. 16, the detecting power of a HTSC nonequilibrium induction detector is

$$D^* = \sqrt{\frac{\tau_{e-ph}^*}{4T^2 c_q d}}, \tag{8}$$

where c_q is the quasiparticle density in the nodal regions, which can be estimated from the formula

$$\frac{c_q(T)}{c_e(T_c)} \approx 1.54 \left(\frac{T}{T_c} \right)^2, \quad T \ll \Delta. \tag{9}$$

At temperature $T=10$ K one has $c_q \simeq 4 \times 10^{-5}$ J $\cdot \mathrm{cm}^{-3} \cdot \mathrm{K}^{-1}$, and, accordingly, the detecting power of a YBaCuO induction detector is $D^* = 10^9$ $\mathrm{W}^{-1} \cdot \mathrm{cm} \cdot \mathrm{Hz}^{1/2}$.

CONCLUSIONS

NbN and YBaCuCo induction detectors are compared with other superconductor detectors in Fig. 5. For a NbN nonequilibrium induction detector the time constant equals the avalanche development time, which is essentially temperature-independent. The detecting power increases with decreasing temperature (Eq. (7)) and $D^* = 10^{12}$ W^{-1} $\cdot \mathrm{cm} \cdot \mathrm{Hz}^{1/2}$ at $T=4.2$ K and $D^* = 10^{16}$ $\mathrm{W}^{-1} \cdot \mathrm{cm} \cdot \mathrm{Hz}^{1/2}$ at

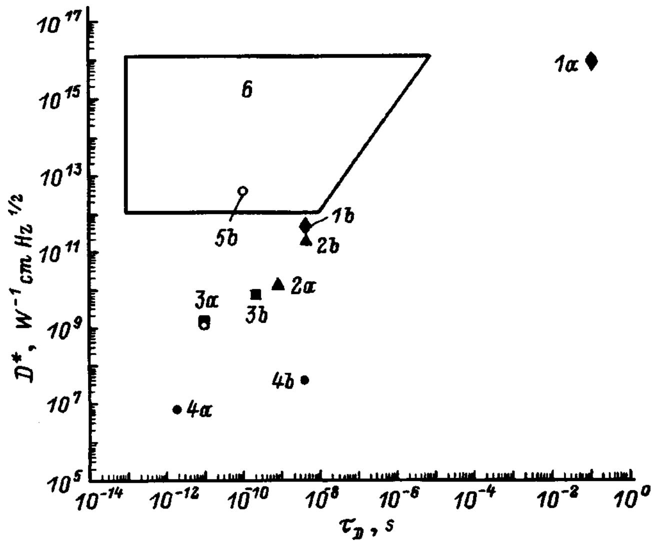

FIG. 5. Response time τ_D and parameter D^* of superconductor detectors: *1a* — Al bolometer (1 K),[29] *1b* — nonequilibrium Al resistance detector (1.7 K),[30] *2a* — nonequilibrium Nb resistance detector (1.7 K),[1] *2b* — nonequilibrium Nb resistance detector (4.2 K),[1] *3a* — nonequilibrium NbN resistance detector (10 K),[2] *3b* — NbN bolometer (10 K),[31] *4a* — nonequilibrium YBaCuO resistance detector (90 K),[4] *4b* — YBaCuO bolometer (90 K),[31] *5a* — nonequilibrium YBaCuO induction detector (10 K), 5b — nonequilibrium YBaCuO induction detector (1 K), *6* — nonequilibrium NbN induction detector (1.6—4.2 K).

$T=1.6$ K. In Fig. 5 the left-hand boundary of the circumscribed region corresponds to this value of D^*. At the same time, at a fixed temperature the output frequency band is bounded below by τ_l^{-1}, which depends on temperature and is presented in Fig. 5 by the right-hand boundary of the circumscribed region. The top boundary corresponds to a working temperature of 1 K and the bottom boundary corresponds to 4.2 K.

The time constant of a $YBa_2Cu_3O_{7-\delta}$ low-temperature nonequilibrium induction detector is determined only by the electron–phonon interaction relaxation time τ^d_{e-ph} in the nodal regions. The detector has the following limiting characteristics: As the working temperature decreases from 10 to 1 K, the time constant τ_D changes from 10 to 100 ps and the detecting power D^* correspondingly improves from 10^9 to 4×10^{12} $W^{-1}\cdot cm\cdot Hz^{1/2}$ (Fig. 5).

To compare with nonequilibrium induction detectors we chose a sensitive Al membrane bolometer with a record-breaking value of D^* as well as fast Nb and NbN nonequilibrium resistance detectors operating near the transition. As one can see, nonequilibrium induction detectors possess a high detecting power because of a low quasiparticle density and record-breaking speed. The drawback of such detectors is a low voltage–power sensitivity, which places quite stringent restrictions on the amplifier in the detection circuit.

This work is supported by the State Program ''Topical Problems in Condensed Matter Physics'' (topic ''Superconductivity'' as part of Project No. 96128) and the Russian Fund for Fundamental Research (Project No. 1-068/4).

[1] E. M. Gershenzon, M. E. Gershenzon, G. N. Gol'tsman *et al.*, Zh. Tekh. Fiz. **59**(2), 111 (1989) [Sov. Phys. Tech. Phys. **34**, 195 (1989)].

[2] B. M. Voronov, E. M. Gershenzon, G. N. Gol'tsman *et al.*, SFKHT **5**, 955 (1992).

[3] E. M. Gershenzon, I. G. Gogidze, G. N. Gol'tzman *et al.*, Pis'ma Zh. Tekh. Fiz. **17**(22), 6 (1991) [Sov. Tech. Phys. Lett. **17**(11), 786 (1991)].

[4] M. Danerud, D. Winkler, M. Lindgren *et al.*, J. Appl. Phys. **76**, 1902 (1994).

[5] S. L. Anisimov, B. L. Kapelevich, and T. L. Perel'man, Zh. Éksp. Teor. Fiz. **66**, 776 (1974) [Sov. Phys. JETP **39**, 375 (1974)].

[6] G. L. Eesley, J. Heremans, M. S. Meyer *et al.*, Phys. Rev. Lett. **65**, 3445 (1990).

[7] A. D. Semenov, R. S. Nebosis, Yu. P. Gousev *et al.*, Phys. Rev. B **52**, 581 (1995).

[8] A. Frenkel, Phys. Rev. B **48**, 9717 (1993).

[9] E. M. Gershenzon, M. E. Gershenzon, G. N. Gol'tsman *et al.*, Zh. Éksp. Teor. Fiz. **86**, 758 (1984) [Sov. Phys. JETP **59**, 442 (1984)].

[10] D. G. McDonald, Appl. Phys. Lett. **50**, 775 (1987).

[11] J. E. Sauvageau and D. G. McDonald, IEEE Trans. Magn. Magn. MAG-25, 1331 (1989).

[12] J. E. Sauvageau, D. G. McDonald, and E. N. Grossman, ISN, 372 (1990).

[13] E. N. Grossman, D. G. McDonald, and J. E. Sauvageau, IEEE Trans. Magn. MAG-27, 2677 (1991).

[14] J. E. Sauvageau, D. G. McDonald, and E. N. Grossman, IEEE Trans. Magn. MAG-27 , 2757 (1991).

[15] S. B. Kaplan, C. C. Chi, D. N. Langenberg *et al.*, Phys. Rev. B **14**, 4854. (1974).

[16] A. V. Sergeev and M. Yu. Reizer, Int. J. Mod. Phys. B **10**, 635 (1996).

[17] B. M. Voronov, E. M. Gershenzon, L. A. Seĭdman *et al.*, Sverkhprovodimost' (KIAE) **7**, 1097 (1994).

[18] G. N. Gol'tsman, P. B. Kouminov, I. G. Goghidze, and E. M. Gershenzon, IEEE Trans. Appl. Supercond. **AS-5**, 2591 (1995).

[19] N. Bluzer, J. Appl. Phys. **78**, 7340 (1995).

[20] C. S. Omen and D. J. Scalapino, Phys. Rev. Lett. **28**, 1559 (1972).

[21] G. N. Gol'tsman, A. D. Semenov, V. P. Gousev *et al.*, Semicond. Sci. Technol. **4**, 453 (1991).

[22] A. S. Kazeroonion, T. K. Cheng, and S. D. Brorson, Solid State Commun. **78**, 95 (1991).

[23] É. E. Aksaev, E. M. Gershenzon, G. N. Gol'tsman *et al.*, Sverkhprovodimost' (KIAE) **3**, 1928 (1990) [Superconductivity **3**, S328 (1990)].

[24] K. Kitazawa, Physica C **235–240**, p. XXIII.

[25] D. A. Wollman, D. J. Van Harlingen, W. C. Lee *et al.*, Phys. Rev. Lett. **71**, 2134 (1993).

[26] D. Pines, Physica B **199–200**, 300 (1994).

[27] Sun Ye and K. Maki, Phys. Rev. B **51**, 6059 (1995).

[28] A. V. Sergeev, A. D. Semenov, P. B. Kouminov *et al.*, Phys. Rev. B **49**, 9091 (1994).

[29] J. Clarke, G. I. Hoffer, P. L. Richards, and N. H. Yeh, J. Appl. Phys. **48**, 4865 (1977).

[30] E. M. Gershenzon, G. N. Gol'tsman, B. S. Karasik *et al.*, Sverkhprovodimost' (KIAE) **5**, 1129 (1992) [Superconductivity **5**, 1126 (1992)].

[31] P. L. Richards, J. Appl. Phys. **76**, 1 (1994).

Translated by M. E. Alferieff

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Mathematical simulation of the technology and properties of gradient spherical lenses

E. A. Fadeev, V. I. Kosyakov, and A. Sh. Tukhvatulin

St. Petersburg State Technical University, 195251 St. Petersburg, Russia

(Submitted May 19, 1997)
Zh. Tekh. Fiz. **68**, 70–73 (October 1998)

The possibilities of improving the optical characteristics of spherical lenses with an axial refractive index gradient are analyzed. The refractive index profile obtained as a result of the Fickian-diffusion exchange of molecules through a flat boundary of the medium is considered. It is shown for the specific polymer system polydiallyl isophthalate plus polymethyl methacrylate (PMMA) that the spherical aberrations can be reduced by almost an order of magnitude in comparison to the aberrations of a homogeneous lens of the same geometric shape made from PMMA. © *1998 American Institute of Physics.* [S1063-7842(98)01210-0]

Gradient lenses are distinguished from lenses made from homogeneous optical materials primarily in that their optical properties can be controlled by altering the refractive index profile, which provides an additional degree of freedom for designing optical systems. In this paper we propose a technology for fabricating spherical lenses with a refractive index gradient, which have better optical properties than do similar homogeneous lenses, and we investigate the properties of such lenses by mathematical modeling. We consider the case in which one of the surfaces (the entrance surface) of the lens is planar and the exit surface has the shape of a sphere of radius R. The main idea behind this study is to utilize the properties of gradient and homogeneous lenses in a single system.

A refractive index gradient can be created, for example, in polymer materials by a method based on the diffusive exchange of monomers with different molecular refractions in a prepolymer matrix followed by curing.[1] If diffusion takes place through a flat surface into a semi-infinite medium, and then a spherical lens is cut from the sample with a refractive index profile obtained (Fig. 1), the refractive index will vary on the spherical boundary of the segment. This circumstance can be utilized to try to reduce the spherical aberrations inherent in lenses made from a homogeneous material. The refractive index profile which an aberration-free spherical lens should have must be ascertained, mathematical simulation of diffusion in the material of the preform must be performed, the method for cutting a lens from the preform must be selected, and the optical characteristics of the systems obtained must be calculated and compared with the optical characteristics of a homogeneous lens.

In order to answer the question of what refractive index profile an ‘‘ideal’’ (aberration-free) lens should have, we must find the refractive index profile in a spherical lens for which all rays parallel to the optical axis that enter the lens at different heights would be focused onto a single point. Such an ‘‘ideal’’ refractive index profile can be determined in the following manner: 1) the entry of a ray on the flat side of the lens and its propagation in the lens are considered; 2) the conditions for refraction of the ray on the spherical surface of the lens and its intersection with the optical axis are formulated; 3) the required profile is calculated as a function of depth and consequently on the exit (spherical) surface of the lens on the basis of the condition that the rays are focused onto a single point on the axis.

Let the ray L, which has entered the lens at the height y, propagate in it to the boundary between the media without being refracted, and then, after it has been refracted, let it reach the focusing point on the axis (Fig. 2). From the law of refraction we obtain

$$n(x)=\frac{n\sin\beta}{\sin\alpha}. \tag{1}$$

Here n is the refractive index of the medium surrounding the lens (air in a particular case), and $n(x)$ is the refractive index of the lens material at the point where the ray passes into the external medium.

If we take into account that the ray entry height $y=\sqrt{R^2-x^2}$, then

$$\sin\alpha=\frac{y}{R}=\frac{\sqrt{R^2-x^2}}{R},\quad \tan\gamma=\frac{F-x}{y}=\frac{F-x}{\sqrt{R^2-x^2}}.$$

Here F is the focal distance of the lens, x is the distance from the flat surface of the lens, and R is the radius of the lens. We express the angle γ in terms of the angles α and β:

$$\gamma=\frac{\pi}{2}-\beta+\alpha.$$

Then, after some elementary trigonometric transformations, we can obtain

$$\tan\gamma=\frac{\cos\alpha\sqrt{1-\sin^2\beta}+\sin\alpha\sin\beta}{\sin\beta\cos\alpha-\sqrt{1-\sin^2\beta}\sin\alpha}. \tag{2}$$

We isolate $\sin\beta$ from (2):

$$\sin\beta=\frac{\tan\gamma\sin\alpha+\cos\alpha}{\sqrt{\tan^2\gamma+1}}.$$

We substitute the expressions for $\sin\alpha$ and $\sin\beta$ and then the expressions for $\tan\gamma$ into (1):

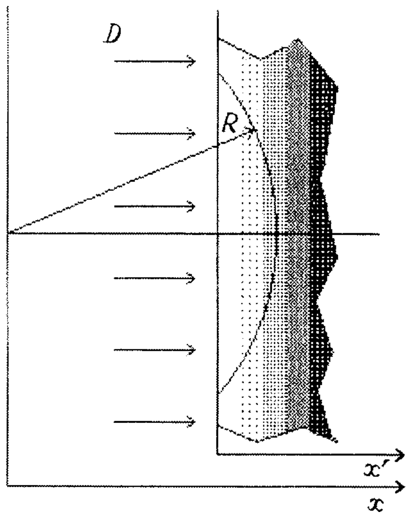

FIG. 1. Schematic representation of the fabrication of an inhomogeneous spherical lens (D is the diffusion flux).

$$n(x)=\frac{n\sin\beta}{\sin\alpha}=\frac{n(\tan\gamma\sin\alpha+\cos\alpha)}{\sin\alpha\sqrt{\tan^2\gamma+1}}$$

$$=\frac{n\left(\dfrac{F-x}{\sqrt{R^2-x^2}}-\dfrac{\sqrt{R^2-x^2}}{R}+\dfrac{x}{R}\right)}{\dfrac{\sqrt{R^2-x^2}}{R}\sqrt{\left(\dfrac{F-x}{\sqrt{R^2-x^2}}\right)^2+1}}.$$

As a result, after some simplification, we obtain the following refractive index profile:

$$n(x)=\frac{F\cdot n}{\sqrt{F^2-2Fx+R^2}}$$

(a similar expression was obtained by a somewhat different method in Ref. 2).

After normalizing all the linear dimensions to the radius of curvature R, we can express x and F in dimensionless units. Then

$$n\left(\frac{x}{R}\right)=\frac{\dfrac{F}{R}n}{\sqrt{\left(\dfrac{F}{R}\right)^2-2\left(\dfrac{F}{R}\right)\left(\dfrac{x}{R}\right)+1}}. \tag{3}$$

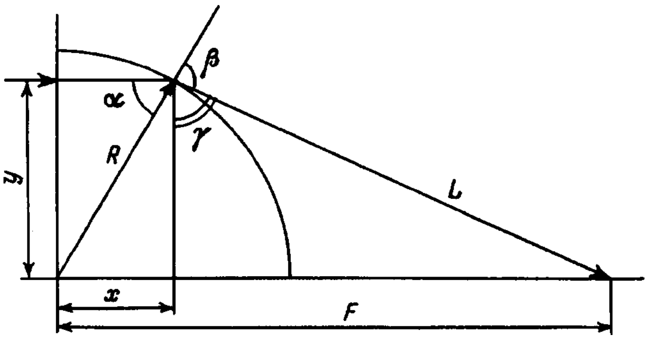

FIG. 2. Calculation of the "ideal" refractive index profile.

This expression is also the condition for the focusing of rays entering the lens along normals to its flat surface. It can be seen that the refractive index should increase as x/R increases. In this case the refraction angles for rays entering the lens far from the optical axis will decrease with resultant compensation of the negative spherical aberration inherent in a homogeneous lens.

In order to ascertain the refractive index profile that can be obtained in practice, let us consider the fairly frequently encountered case of diffusive behavior in accordance with Fick's law. We shall dwell on the case of one-dimensional one-component diffusion in a semi-infinite sample having a flat boundary with constant initial and boundary conditions and a constant diffusion coefficient. Under these conditions the solution of Fick's equation can be found in analytical form:[3]

$$f\left(\frac{x'}{R}\right)=\frac{C\left(\dfrac{x'}{R}\right)}{C_0}=1-\operatorname{erf}\left(\frac{\dfrac{x'}{R}}{2\sqrt{\dfrac{Dt}{R^2}}}\right),$$

where D is the diffusion coefficient, t is the time after the beginning of diffusion, x' is the distance from the surface of the semi-infinite sample, C_0 is the concentration of the diffusant on the boundary of the semi-infinite sample, C is the concentration of the diffusant in the layer at the distance x' from the surface of the semi-infinite sample, and t is the volume fraction of the diffusant at a definite point in the matrix medium.

The refractive index profile was obtained from the concentration profile using the Gladstone–Dale refraction formula

$$n\left(\frac{x'}{R}\right)=n_1-(n_1-n_2)f\left(\frac{x'}{R}\right)$$

$$=n_1\left(1-f\left(\frac{x'}{R}\right)\right)+n_2 f\left(\frac{x'}{R}\right), \tag{4}$$

where $n(x')$ is the refractive index of the lens at a distance x' from the surface of the semi-infinite sample, n_1 is the refractive index of the matrix material (when $f=0$), and n_2 is the refractive index of the diffusant (when $f=1$).

A calculation using the equations presented yields the refractive index profile for the interdiffusion of specified substances for any times. Thus we have the ideal refractive index profile required to satisfy Eq. (3) (Fig. 3) and a family of Fickian refractive index profiles (4) obtained as a result of diffusion for different moments in time (Fig. 4). To create an aberration-free lens, on any suitable Fickian curve (Fig. 3) we must find the portion of the refractive index profile that coincides most closely with the portion of the ideal curve corresponding to the assigned range of variation of the refractive index. However, the curves in Figs. 3 and 4 have positive and negative curvatures, respectively, over the entire range of variation of the distance from the flat surface of the lens to the boundary between the media at a given height. Therefore, to juxtapose them, we should apparently select segments of the curves which can be regarded as straight

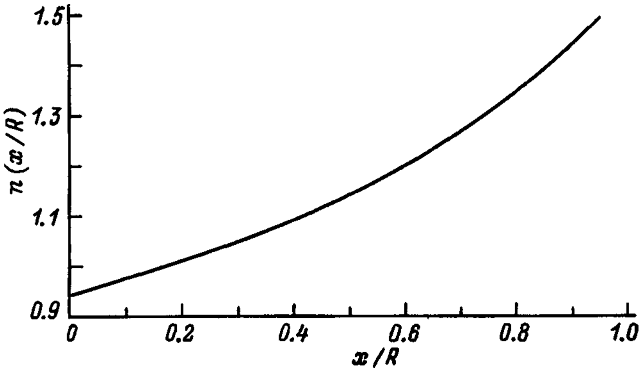

FIG. 3. ''Ideal'' refractive index profile (values $n<1$ mean that a ray will be reflected in that region).

lines with the same slope. Comparing these dependences, we should also bear in mind that in Eq. (3) for the ''ideal'' refractive index profile the coordinate x is measured from the center of the spherical lens, while in Eq. (4) for diffusive refractive index profiles the coordinate x' is measured from the flat boundary separating the diffusion components at the initial moment in time.

As an example, we construct the refractive index profiles for gradient media known from the literature.[1] As the material with a low refractive index we take methyl methacrylate (MMA), whose refractive index $n_1=1.494$ (after polymerization), and as the material with a high refractive index we take diallyl isophthalate (DAIP) with $n_2=1.570$ (after polymerization). The segments of the ''ideal focusing'' curve at $0.95R \leqslant x \leqslant R$ and of the diffusion curve at $0 \leqslant x' \leqslant 0.05R$ turned out to be closest under the conditions given. The slopes of the segments indicated practically coincide, as is illustrated in Fig. 5, in which these two curves have been juxtaposed.

Thus, after cutting a spherical lens with radius R from the semi-infinite sample so that the center of the sphere would be located on the left at a distance equal to $0.95R$ from the planar boundary [in this case the segment of the ''Fickian'' curve described above ($0 \leqslant x' \leqslant 0.05R$) corresponds to the interval $0.95R \leqslant x \leqslant R$ (Fig. 1)], we obtain a spherical lens with optical characteristics that are apparently better than those of a spherical lens with the same geometric dimensions but without a refractive index gradient. For an exact comparison of the aberration characteristics of gradient and homogeneous lenses, we carried out a modeling of the ray paths in the inhomogeneous medium.

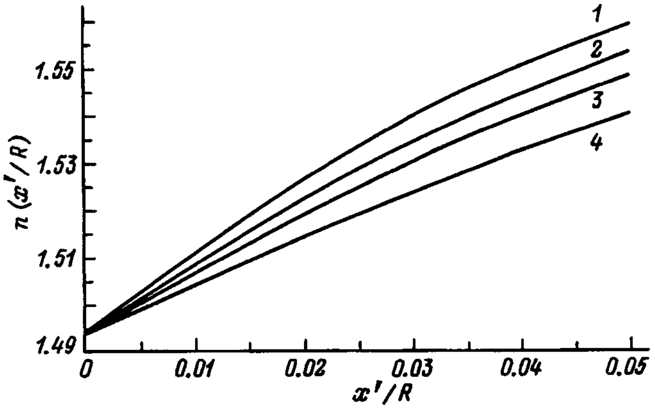

FIG. 4. Family of refractive index profiles obtained as a result of solving Fick's equation. $\frac{R}{2\sqrt{Dt}}$: *1* — 12, *2* — 17, *3* — 18, *4* — 20.

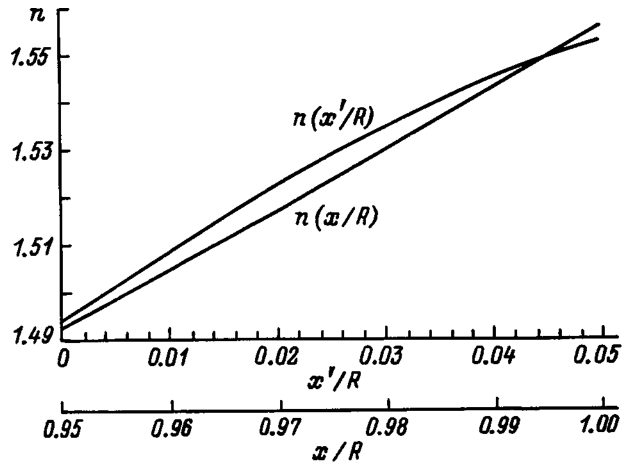

FIG. 5. Segments of curves (*3*) and (*4*) that are close to one another.

The trajectory of a ray in a medium with an arbitrary refractive index profile is usually found by a numerical calculation. The most widely employed technique is the algorithm described in Ref. 4, in which the ray equation

$$\frac{d}{dS}\left(n(\mathbf{r})\frac{d\mathbf{r}}{dS}\right)=\nabla n(\mathbf{r}); \qquad \mathbf{r}=(x,y,z)$$

is transformed by a change of variables

$$t=\int \frac{dS}{n(\mathbf{r})}, \qquad dt=\frac{dS}{n(\mathbf{r})}$$

into a form that is convenient for solving by the Runge–Kutta procedure:

$$\frac{d^2\mathbf{r}}{dt^2}=n(\mathbf{r})\nabla n(\mathbf{r}). \tag{5}$$

Here dS is an element of the ray trajectory; $\mathbf{r}(x,y,z)$ is the radius vector of the element dS, and $n(\mathbf{r})$ is the refractive index at the point (x,y,z). The optical ray vector, whose components are optical direction cosines between the tangent to the trajectory and the x, y, and z axes, has the form

$$T=\frac{d\mathbf{r}}{dt}=n(\mathbf{r})\frac{d\mathbf{r}}{dS}=n(\mathbf{r})\frac{dx}{dS}\mathbf{i}+n(\mathbf{r})\frac{dy}{dS}\mathbf{j}+n(\mathbf{r})\frac{dz}{dS}\mathbf{k}$$
$$=p\mathbf{i}+q\mathbf{j}+l\mathbf{k}.$$

To solve Eq. (5) by the Runge–Kutta method, we introduce column matrices of the components of the trajectory P, the ray vector T (the trajectory slope), and the refractive index gradient M

$$P=\begin{pmatrix} x \\ y \\ z \end{pmatrix}; \quad T=\begin{pmatrix} T_x \\ T_y \\ T_z \end{pmatrix}=n(\mathbf{r})\begin{pmatrix} \dfrac{dx}{dS} \\ \dfrac{dy}{dS} \\ \dfrac{dz}{dS} \end{pmatrix};$$

$$M=n(\mathbf{r})\begin{pmatrix} \dfrac{\partial n(\mathbf{r})}{\partial x} \\ \dfrac{\partial n(\mathbf{r})}{\partial y} \\ \dfrac{\partial n(\mathbf{r})}{\partial z} \end{pmatrix}=\frac{1}{2}\begin{pmatrix} \dfrac{\partial^2 n(\mathbf{r})}{\partial x} \\ \dfrac{\partial^2 n(\mathbf{r})}{\partial y} \\ \dfrac{\partial^2 n(\mathbf{r})}{\partial z} \end{pmatrix}.$$

Then Eq. (5) can be written in the matrix form

$$\frac{d^2P}{dt^2}=M(P).$$

The values of the matrices (P_1,T_1); $(P_2,T_2);\ldots;(P_m,T_m)$ are successively calculated according to the Runge–Kutta scheme with a t step Δt, beginning at the first point at which the initial values (P_0 and T_0) are assigned:

$$P_{m+1}=P_m+\Delta t\left(T_m+\frac{1}{6}(A+2B)\right),$$

$$T_{m+1}=T_m+\frac{1}{6}(A+4B+C).$$

Here the matrices A, B, and C are defined in the following manner:

$$A=\Delta t M(P_m),$$

$$B=\Delta t M\left(P_m+\frac{\Delta t}{2}T_m+\frac{1}{8}\Delta t A\right),$$

$$C=\Delta t M\left(P_m+\Delta t T_m+\frac{1}{2}\Delta t B\right).$$

These formulas enable us to calculate the coordinates of the point of intersection of the optical axis with a ray entering the lens at any height. The optical quality of a lens is generally characterized by the magnitude of the longitudinal spherical aberration, which is equal to the difference between the coordinates of the points of intersection of the optical axis with an arbitrary and a paraxial ray: $\delta(y)=F(y)-F(0)$.

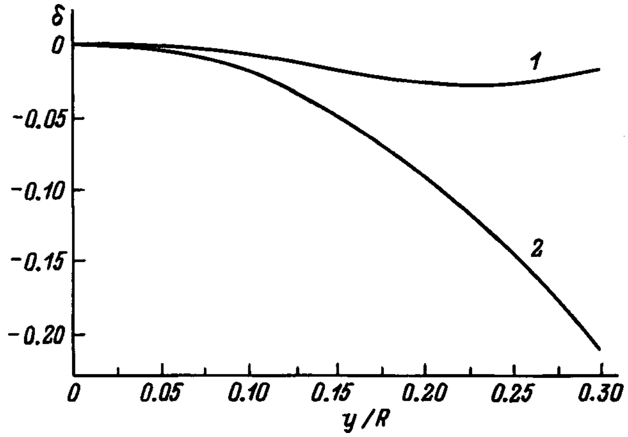

FIG. 6. Aberration characteristics of a homogeneous lens (*2*) and a lens with a refractive index gradient (*1*).

When the aberration characteristics of a gradient and an ordinary lens are compared, it is reasonable to employ a lens fabricated from the material of one of the components of the system as the homogeneous lens. We chose the component with $n=1.494$ for this purpose. The calculations showed that rays entering such a lens with a homogeneous refractive index at different heights y intersect the optical axis at points separated from one another by a maximum distance of $0.212R$, while rays entering a lens with a refractive index gradient intersect at points separated from one another by a maximum distance of $0.028R$ (Fig. 6). It is seen that the maximum longitudinal spherical aberration is almost an order of magnitude smaller in the gradient lens than in the homogeneous lens.

Thus, the requirements placed on the axial distribution of the refractive index in a gradient spherical lens have been formulated as a result of this work. A method for fabricating lenses with a diffusive (Fickian) refractive index profile has been proposed with consideration of these requirements. The optical properties of a gradient lens have been analyzed by mathematical simulation, and it has been shown that the spherical aberration is reduced approximately eightfold in comparison to the analogous lens without a refractive index gradient.

[1] N. B. Galimov, V. I. Kosyakov, A. Sh. Tukhvatulin *et al.*, Opt. Spektrosk. **50**, 545 (1981) [Opt. Spectrosc. **50**, 295 (1981)].

[2] V. I. Tarkhanov, Opt. Mekh. Promst. **55**(2), 23 (1988) [Sov. J. Opt. Technol. **55**, 91 (1988)].

[3] R. M. Barrer, *Diffusion in and through Solids* [Macmillan, New York (1941); IL, Moscow (1948), 504 pp.].

[4] A. Sharma, D. Viziakumar, and A. K. Ghatak, Appl. Opt. **21**, 984 (1982).

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

High-power single-frequency tunable solid-state lasers

V. V. Antsiferov

Nuclear Physics Institute, Siberian Branch of the Russian Academy of Sciences, 630090 Novosibirsk, Russia

(Submitted May 21, 1997)
Zh. Tekh. Fiz. **68**, 74–79 (October 1998)

Experimental results are presented on the achievement of single-frequency tunable lasing in ruby, Nd-glass, and Nd:YAG lasers with electrooptic Q switching of the cavity by the injection of an external signal. An optimization of the parameters is carried out for lasers on neodymium ions in yttrium aluminum garnet, lanthanum beryllate, chromium-doped gadolinium scandium gallium garnet, and lanthanum hexaaluminate with passive Q switching of the cavity by means of lithium fluoride shutters containing F_2^- color centers. High-power single-frequency generation of giant pulses is achieved, with the output wavelength tunable over the half-width of the gain lines of the active media. © *1998 American Institute of Physics.* [S1063-7842(98)01310-5]

INTRODUCTION

The rapid development of the research on ultrahigh-resolution spectroscopy, the selective ionization and excitation of atoms and molecules, the diagnostics of plasmas and of atom and ion beams, etc. using high-power solid-state lasers is imposing increasingly stringent requirements on the parameters of the laser radiation: the power, the spectral width, the width of the tuning range, the temporal and spatial distribution of the intensity, the stability of the spectrum, and the time when the giant pulse appears.

Longitudinal mode selection is accomplished in lasers by creating losses in the laser cavity which depend on the output wavelength and making the radiation losses as small as possible for the modes selected. The efficiency of longitudinal mode selection is considerably higher in lasers with passive Q switching of the cavity than in lasers with active Q switching. This is because the time for the linear development of a giant pulse is of the order of 1 μs in lasers with passive Q switching, while it is an order of magnitude shorter in lasers with active Q switching. During linear development, various modes grow from the level of spontaneous noise independently of one another, and the ratio between the mode amplitudes at the moment of nonlinear development of the giant pulse determines the modes which participate in the lasing. The difference between loss coefficients needed to suppress the weaker modes is inversely proportional to the number of passes k of the photons through the cavity when the maximum laser power is achieved:

$$\frac{P_1}{P_2}=\left[\frac{1-\gamma_1}{1-\gamma_2}\right]^k, \tag{1}$$

where γ is the radiation loss in one pass for each mode.

In the case of passive Q switching one has $k\sim10^3$, and the required difference between the losses for two modes, under which their intensities will differ by an order of magnitude at the end of the linear development, is $(\gamma_2-\gamma_1)=10^{-3}$. Single-frequency lasing is achieved in a laser with passive Q switching under the conditions of weak discrimination of the longitudinal modes. In lasers with active Q switching the number of passes through the cavity is reduced to several tens. This does not permit sufficient discrimination of the longitudinal modes by conventional selection methods so as to obtain lasing on a single longitudinal mode.

METHODS FOR OBTAINING SINGLE-FREQUENCY LASING

1. Dispersive prisms, diffraction gratings, and interference-polarization filters are employed as dispersive elements in lasers having active media with broad gain bands owing to their large dispersion regions, although they do not have very high frequency selectivities for isolating individual modes. Tuning of the output wavelength of a giant pulse with a spectral width of 0.1 nm over a 5.6-nm range can be achieved in a Nd-glass laser with passive Q switching and a prism dispersive cavity.[1] When a compound dispersive cavity consisting of a holographic diffraction grating and an uncoated resonant reflector is employed, the single-frequency generation of a giant pulse with tuning of the output wavelength over a 2.2-nm range is realized.[2] Interference-polarization filters have been employed for selecting longitudinal modes and tuning of the output wavelength under conditions of passive Q switching in ruby,[3] Nd:YAG,[4] and Nd-glass[5] lasers. Tuning of the output wavelength was accomplished by smoothly varying the total thickness of the crystalline wedges, and its range in the Nd-glass laser amounted to 13.4 nm with a spectral width of 4.5 nm.[5]

2. Highly efficient longitudinal mode selection is achieved in lasers with passive Q switching when a resonant reflector consisting of two or more plane-parallel plates separated by air spaces is employed. With a resonant reflector as the exit mirror of the cavity it is fairly simple to achieve single-frequency generation in ruby[6,7] and Nd-glass[8,9] lasers. The output wavelength of the giant pulse is tuned by varying the pressure in the air spaces. However, a resonant reflector without reflecting coatings has diminished selectivity, and single-frequency lasing cannot be achieved with its use in lasers with active Q switching even near the threshold.

3. The time for linear development of a giant pulse can be artificially increased in lasers with active Q switching by means of slow[10,11] or two-step[12,13] Q switching. In this case, as in a regime with passive Q switching, single-frequency generation of a giant pulse is achieved under the conditions of weak discrimination of the longitudinal modes in ruby[11,12] and Nd:calcium tungstate[13] lasers. However, when the cavity Q factor is switched slowly, the intracavity losses increase dramatically, and the output power of the giant pulses drops. In addition, in the case of slow Q switching it is necessary to work near the threshold; otherwise the output contains several pulses with different spectral characteristics, which also limit the power of the giant pulse. These restrictions are relaxed somewhat when the cavity Q factor is switched in two steps. In the first Q-switching step the shutter is partially opened, and a weak single-frequency output pulse is formed during the period of linear development. At the moment corresponding to the maximum value of the pulse intensity, a second voltage pulse, which completely opens the shutter, is supplied to it, and the giant pulse is emitted without significant power losses. However, in this case, too, it is necessary to work near the threshold.

4. In the case where exact synchronization of the time of appearance of the giant pulse and the process under investigation is necessary, lasers with electrooptic Q switching of the cavity are employed. The most effective method for obtaining single-frequency lasing in such lasers is the injection of an external signal, which was proposed and experimentally implemented with the direct involvement of the author in Ref. 4. After Q switching, single-frequency radiation from a low-power control laser is injected into the cavity of the high-power laser. Since the time of linear development of the giant pulse is inversely proportional to the logarithm of the intensity I of the radiation in the laser cavity,

$$\tau_d \sim \log \frac{C}{I}, \tag{2}$$

in this case the radiation in the giant pulse arises not from spontaneous noise with a broad spectrum, but from the external single-frequency radiation, whose intensity is many orders of magnitude greater than the intensity of the spontaneous noise. In the case of radiation trapping, the spectrum of the giant pulse is completely identical to the spectrum of the external signal. The lack of any selecting elements in the cavity of the high-power laser permits the production of giant pulses with large powers in a single-frequency regime. The gain of a weak monochromatic signal in an injection regime can reach $\sim 10^{10}$. The injection of an external signal has been used in ruby[15] and Nd-glass[16–18] lasers with electrooptic Q switching. The injection of an external signal has found wide application not only in solid-state lasers, but also in liquid[19,20] and gas[21,22] lasers.

RUBY LASER WITH ELECTROOPTIC *Q* SWITCHING OF THE CAVITY

In Refs. 23–27 we thoroughly investigated the use of the injection of an external signal in solid-state lasers with electrooptic Q switching of the cavity. In the three-mirror variant of the scheme for injecting an external signal[23,24] the exit mirror of the control laser served as the semitransparent mirror of the high-power laser. A single-frequency quasistationary ruby laser[28] served as the control laser. In this case there is no need for additional synchronization of the operating regimes of the two lasers. The electrooptic shutter of the high-power laser was opened at the moment when maximum population inversion was achieved, $0.3-0.5$ ms after the onset of lasing in the control laser.

At the edge of the tuning range in the absence of radiation trapping of the external signal, the width of the output spectrum of the high-power laser was of the order of 50 pm. Radiation trapping of the external signal took place in a spectral region with a width of ~ 180 pm near the maximum of the gain line, and the output spectrum of the giant pulse coincided completely with the output spectrum of the external signal. However, in this case the high-power laser generated a bifurcated giant pulse. This is due to the decrease in the reflectivity of the shared mirror as a consequence of the increase in the stored energy in the cavity of the control laser during buildup of the giant pulse intensity, which leads to increases in the losses and the self-excitation threshold of the high-power laser. The power of the giant pulse was restricted to a level of the order of 1 MW in this scheme by the resolution of the reflecting coatings of the Fabry–Pérot etalon selector in the control laser.

In order to increase the power of the giant pulse, a four-mirror scheme for injecting an external signal was investigated in Refs. 23 and 27. The control laser was the same as before, but the cavity of the high-power laser was formed by one end surface of a ruby rod and a semitransparent mirror. Optical isolation between the lasers was effected by an electrooptic shutter consisting of two Archard–Taylor prisms and a Pockels cell, which was switched to transmission of the output of the control laser for a period of the order of 50 ns several nanoseconds before the electrooptic shutter of the high-power laser was opened. Radiation from the giant pulse with a power $\geqslant 10$ MW was extracted through a lateral surface of the polarizing prism, which was often damaged when the power of the giant pulse was very high.

To eliminate this deficiency, a variant of the scheme for injecting an external signal with a ring cavity in the high-power laser was investigated in Refs. 25 and 26. The effect of the giant pulse on the control laser was eliminated in this scheme without the need for electrooptic isolation, because radiation trapping of the external signal occurred not only with respect to the spectrum but also with respect to direction. This significantly simplified the experimental setup. The entry and extraction of radiation took place in the ring laser through a prism with attenuated total internal reflection. Generation of a 50-MW giant pulse on one longitudinal mode (Fig. 1a) was achieved with smooth tuning of the output wavelength over a 250-pm range (Fig. 1c).

Employment of a more powerful single-frequency quasicontinuous ruby control laser with spherical mirrors in a spectrum-inversion regime[29] permitted further simplification of the experimental setup. In this case a linear four-mirror variant of the injection scheme without electrooptic isolation

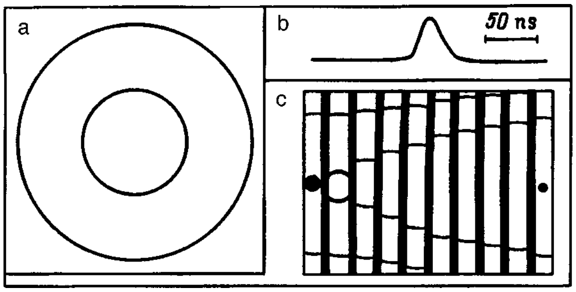

FIG. 1. Output parameters of a high-power single-frequency ruby laser with electrooptic Q switching of the cavity: a — interferogram of the giant-pulse output spectrum (the free spectral range of the interferometer is 1.6 pm); b — oscillogram of the output intensity; c — sequence of interferograms of the giant-pulse output spectrum illustrating the tuning range of the output wavelength (the free spectral range of the interferometer is 240 pm).

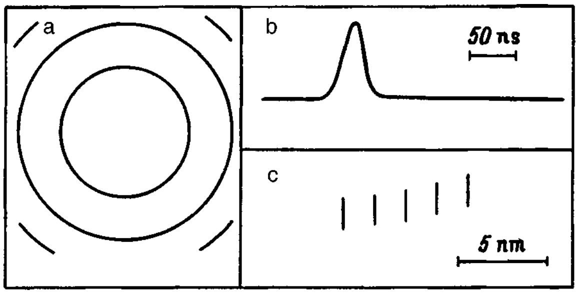

FIG. 2. Output parameters of a high-power single-frequency neodymium-glass laser with electrooptic Q switching of the cavity: a — interferogram of the giant-pulse output spectrum (the free spectral range of the interferometer is 8 pm); b — oscillogram of the output intensity; c — sequence of interferograms of the giant-pulse output spectrum illustrating the tuning range of the output wavelength.

between the lasers was used. The high intensity of the emission from the control laser permitted injection of its output into the cavity of the high-power laser through its semitransparent mirror. In this case the back-effect of the giant pulse on the control laser was insignificant. The spherical front of the control laser was matched to the flat front of the high-power laser using a telescope. Single-frequency generation of a 200-MW giant pulse was achieved when a ruby with sapphire tips of length 240/320 mm and diameter 14 mm was employed in the high-power laser. The output wavelength was tuned by tilting the mirrors of the control laser, and its range was of the order of 120 pm.

The experimentally measured minimum value of the intensity of the external signal at which stable radiation trapping of this signal was achieved at the center of the gain line for the ruby laser was of the order of 0.1 W/cm^2.

Nd-GLASS AND Nd:YTTRIUM ALUMINUM GARNET LASERS WITH ELECTROOPTIC *Q* SWITCHING

The schemes for injecting an external signal in Nd lasers employed fairly powerful single-frequency control lasers, whose output was injected into the cavities of high-power lasers through the semitransparent mirror of the cavity. In this case there was no longer a need for optical isolation between the lasers. A Nd:YAG control laser operating in a regime with single-frequency quasistationary generation of TEM$_{mnq}$ modes was used to create a high-power single-frequency Nd:YAG laser in a scheme for injecting an external signal in Ref. 30. Single-frequency generation of a 40-MW giant pulse by the Nd:YAG laser with tuning of the output wavelength over a 0.35-nm range was obtained with a garnet crystal of length 100 mm and diameter 6 mm in the high-power laser in a regime with radiation trapping of the external signal.

To increase the tuning range of the output wavelength near 1.06 μm, a scheme for injecting an external signal with a neodymium-phosphate glass (GLS-22) laser was investigated. A single-frequency quasistationary Nd-glass laser served as the control laser.[31] An increase in the quasistationary output energy of TEM$_{00q}$ modes in the control laser was achieved by using a neodymium rod in the phosphate glass with a length of 300 mm and increasing the diameter of the diaphragms to 3 mm for a cavity length of 230 cm. This made it possible to raise the single-frequency output energy of the glass control laser to 0.5 J.

In the region with radiation trapping of the external signal, the output spectrum of the giant pulse was identical to the output spectrum of the external signal (Fig. 2a). The absence of intensity beats on the output oscillogram (Fig. 2b) attests to generation of the giant pulse on one longitudinal mode. The power of the giant pulse was 180 MW, and the tuning range of the output wavelength was 5.6 nm (Fig. 2c). The experimentally measured output intensity of the external signal needed for steady radiation trapping of this signal by the giant pulse of a Nd laser was of the order of 0.5 W/cm^2. The radiation-trapping band grew slowly as the pump energy was increased. A sharp increase in the output intensity of the external signal did not lead to appreciable broadening of its radiation-trapping band.

The band $(\omega_s-\omega_0)$ for radiation trapping of the external signal by the giant pulse is estimated under the assumption that the inverted population in the active medium remains practically unchanged $[\partial n(\omega,t)/\partial t=0]$ during the time of linear development of the giant pulse ($t_d\sim 10^{-7}$ s). Then the spectral intensity $I(\omega,t)$ obeys the equation

$$\frac{dI(\omega,t)}{dt}=[\sigma n(\omega)-\tau_p^{-1}]I(\omega,t)+S(\omega)+\tau_p^{-1}U(\omega_s). \tag{3}$$

Here σ is the cross section of the induced transition, τ_p is the lifetime of a photon in the cavity, $S(\omega)$ is the intensity of the spontaneous emission, $U(\omega_s)$ is the intensity of the external signal, and $n(\omega)$ is the inverted population. Since the contour of an inhomogeneously broadened line has a Gaussian shape with a width Γ, for the inverted population and the intensity of the spontaneous noise we have

$$n(\omega,t)=n_0\exp\left[-\left(\frac{\omega-\omega_0}{\Gamma}\right)^2\right], \tag{4}$$

$$S(\omega)=S_0\exp\left[-\left(\frac{\omega-\omega_0}{\Gamma}\right)^2\right]. \tag{5}$$

Setting $I(\omega,0)=0$, we obtain the solution of Eq. (3) in the form

$$I(\omega,t)=\frac{\left\{(\omega_s)+S_0\tau_p\exp\left[-\left(\frac{\omega-\omega_0}{\Gamma}\right)^2\right]\right\}}{\sigma n_0\tau_p\exp\left[-\left(\frac{\omega-\omega_0}{\Gamma}\right)^2\right]-1}\exp[k(\omega)t-1], \tag{6}$$

where

$$k(\omega)=\sigma n_0\exp\left[-\left(\frac{\omega-\omega_0}{\Gamma}\right)^2\right]-\tau_p^{-1}$$

is the gain coefficient.

The order of magnitude of t_d is given by the expression

$$t_d=\frac{1}{k(\omega)}\ln[k(\omega)\tau_p I_d/(U+S\tau_p)+1], \tag{7}$$

where I_d is the spectral intensity at the end of the linear stage of development of the giant pulse.

The condition $I(\omega_s,t_d)=10S(\omega)$ can serve as a criterion for radiation trapping of the external signal. Hence, with consideration of (6) for the radiation-trapping band of the external signal we obtain

$$|\omega_s-\omega_0|<\Gamma\left[\ln\left(\frac{\sigma n_0\tau_p}{1+Q\tau_p t_d^{-1}}\right)\right]^{1/2}, \tag{8}$$

where

$$Q=\ln\left[1+\frac{10Sk\tau_p}{U+S\tau_p}\right].$$

It follows from (8) that the radiation-trapping band of the external signal slowly grows with increasing intensity of the external signal $U(\omega_s)$. At the same time, the duration of the linear development of the giant pulse falls off slowly as the intensity of the external signal rises.

Nd:YAG, Nd:LNB, Nd:Cr:GSGG, AND Nd:LNA LASERS WITH PASSIVE Q SWITCHING BY LIF:F_2^- SHUTTERS

The energy and spectral characteristics of the output of Nd:YAG (yttrium aluminum garnet), Nd:LNB (lanthanum beryllate), Nd:Cr:GSGG (gadolinium scandium gallium garnet), and Nd:LNA (lanthanum hexaaluminate) lasers with passive Q switching by LiF:F_2^- shutters were investigated in Refs. 32–36.

Phototropic shutters have found broad application in the passive Q switching of the cavities of solid-state lasers owing to their simplicity and convenience. However, passive dye shutters have significant losses and do not exhibit high stability. Strontium fluoride crystals with color centers were first used for passive Q switching in a ruby laser,[37] but because of the lower thermal and radiation stability of the color centers, they proved to be ineffective and have not become widely used. Only after the creation of F_2^- color centers that are stable at room temperature in lithium fluoride crystals (F_2^-:LiF)[38] did crystals begin to be employed as nonlinear filters[39] and passive laser shutters.[40] The accessibility of lithium fluoride crystals, the simplicity of their radiation-induced coloration using Co^{60} γ radiation, and their high thermal conductivity and radiation stability have led to their broad use as phototropic shutters for obtaining giant pulses in neodymium-ion lasers.

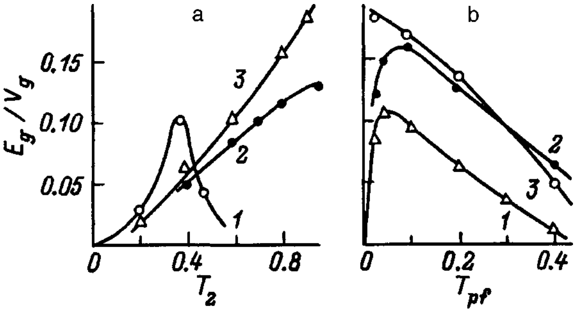

FIG. 3. Dependence of the output energy density E_g/V_g (J/cm^3) on the transmission coefficient of the cavity exit mirror T_2 (a) and the transmission coefficient of the passive filter $P_{\rm pf}$ (b) for a constant pump energy $E_p = 1.2E_t$ of the giant pulses: $L=40$ cm; lasers and generating volumes of the active media V_g, cm^3: *1* — Nd:YAG, 1.24; *2* — Nd:LNB, 0.56; *3* — Nd:Cr:GSGG, 0.5.

The absorption cross section of a phototropic F_2^-:LiF shutter at a wavelength of 1060 nm is 1.7×10^{-17} cm^{-2}, and the relaxation time is 90 ns.[41] The thermal conductivity equals 0.1 W/(cm·K), and the thermal strains do not exceed several percent of the strains in Nd:YAG crystals.[42] The measured absorption coefficient for a concentration of active centers equal to 2×10^{16} cm^{-3} with allowance for Fresnel reflection is 0.41 cm^{-1} (Ref. 43). The degradation threshold of individual crystals is approximately equal to 2 GW/cm^2 for a pulse duration of 20 ns (Ref. 43), but in most crystals the degradation thresholds are $4-5$ times lower. Therefore, preliminary selection of the crystals according to this parameter is needed. In addition, the considerable spread of impurity compositions in the original lithium fluoride crystals makes it difficult to obtain reproducible results.

The high thermal conductivity of F_2^-:LiF crystals permits obtaining stable energy characteristics of the radiation in the giant pulses of neodymium-ion lasers at pulse repetition frequencies up to 100 Hz.[44] Linear polarization of the radiation in giant pulses can be achieved with F_2^-:LiF shutters without a polarizer in the laser cavity.[41] A review of the early research on passive laser shutters with color centers was given in Ref. 45.

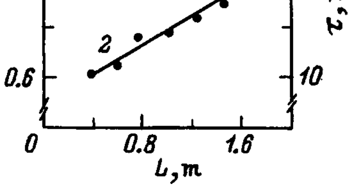

FIG. 4. Dependence of the relative energy E_g (*1*) of giant pulses and their duration τ (*2*) for neodymium-ion lasers on the cavity length L ($E_p = 1.2E_t$).

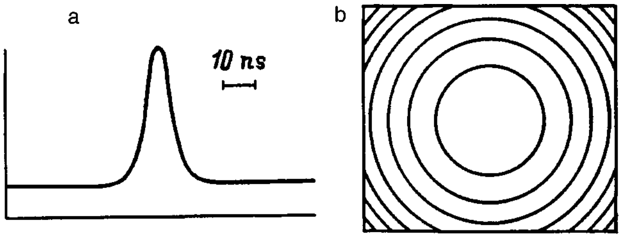

FIG. 5. Output parameters of high-power single-frequency neodymium-ion lasers with passive Q switching of the cavity by LiF:F_2^- shutters: a — oscillogram of the output intensity of the giant pulses, b — interferogram of the output spectrum (the free spectral range of the interferometer is 20 pm).

ENERGY AND SPECTRAL CHARACTERISTICS OF THE LASER RADIATION

The maximum output energy of the giant pulses of lasers was achieved when the transmission coefficients of the cavity exit mirror (T_2) (Fig. 3a) and of the passive filter (T_{pf}) (Fig. 3b) were optimized. The output energy was then measured for a single giant pulse until the appearance of a second pulse, whose lasing threshold was about 1.2 times the lasing threshold of a single pulse. As the pump energy was increased, the number of giant pulses increased linearly, and the output energy grew proportionally to the output energy in a single pulse. At high pump energies the total energy of the giant pulses was of the order of the free-lasing energy. The plots of the output energy in the giant pulses versus the transmission coefficients of the cavity exit mirror and the passive filter had maxima. One exception is the Nd:Cr:GSGG laser, which did not produce such maxima even with the maximum possible values for the transmission coefficients in the experiment: $T_2 = 0.96$ and $T_{\mathrm{pf}} = 0.04$. Because of the very low values of the output energy of the Nd:LNA laser, its energy dependences are not presented in the figures. The following specific powers of the giant pulses were obtained in different lasers with optimal parameters for the cavity and the passive filter: 18 (Nd:Cr:GSGG), 16 (Nd:LNB), 11 (Nd:YAG), and 0.5 MW/cm^3 (Nd:LNA).

As the length of the planar cavity was increased, the output energy in the giant pulses decreased linearly (Fig. 4). Over the entire range of variation of the cavity length, the duration of the giant pulses with a smooth symmetric shape varied from 10 to 15 ns (Fig. 5a). Without longitudinal mode selection, several longitudinal modes participated in the lasing. When a resonant reflector consisting of two sapphire substrates served as the cavity exit mirror, stable single-frequency generation of giant pulses was achieved in all the media (Fig. 5b).

Single-frequency generation of giant pulses (Fig. 5b) with tuning of the output wavelength in ranges of the following widths was achieved in a compound dispersive cavity: 0.2 (Nd:YAG), 0.3 (Nd:Cr:GSGG), 0.4 (Nd:LNB), and 0.5 nm (Nd:LNA).

[1] V. B. Markov, T. Ya. Marusiĭ, M. S. Soskin *et al.*, Zh. Éksp. Teor. Fiz. **64**, 1538 (1973) [Sov. Phys. JETP **37**, 778 (1973)].
[2] O. N. Pogorelyĭ, M. S. Soskin, and B. B. Taranenko, Pis'ma Zh. Tekh. Fiz. **2**(2), 49 (1976) [Sov. Tech. Phys. Lett. **2**(1), 19 (1976)].
[3] L. Girkovic, D. E. Evans, and M. J. Forrest, Appl. Opt. **7**, 981 (1968).
[4] H. G. Danielmeyer, IEEE J. Quantum Electron. **QE-6**, 101 (1970).
[5] Yu. N. Belyaev, A. N. Kiselev, and M. A. Novikov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **13**, 1405 (1970).
[6] M. Hercher, Appl. Phys. Lett. **7**(2), 32 (1965).
[7] E. I. Nikonova, E. N. Pavlovskaya, and G. P. Startsev, Opt. Spektrosk. **22**, 984 (1967) [Opt. Spectrosc. **22**, 535 (1967)].
[8] A. L. Egorov, V. V. Korobkin, and R. V. Serov, Kvantovaya Élektron. (Moscow) **2**, 513 (1975) [Sov. J. Quantum Electron. **5**, 291 (1975)].
[9] J. Morellec, D. Normand, and G. Pitite, Appl. Opt. **8**, 141 (1979).
[10] J. M. McMahon, IEEE J. Quantum Electron. **QE-5**, 489 (1969).
[11] A. L. Mikaélyan, Yu. G. Turkov, and L. N. Razumov, Kvantovaya Élektron. (Moscow), No. 2, 96 (1971) [Sov. J. Quantum Electron. **1**, 189 (1971)].
[12] E. D. Isyanova, A. M. Marugin, and V. M. Ovchinnikov, Zh. Prikl. Spektrosk. **12**, 834 (1970).
[13] D. C. Hanna, B. Luther-Davies, and R. C. Smith, Opto-Electron. **3**(4), 163 (1971).
[14] A. N. Bondarenko, V. A. Smirnov, and V. V. Antsiferov, JETP Lett. **6**, 178 (1967).
[15] Yu. V. Vinogin, L. N. Gnatyuk, V. A. Nikishin *et al.*, Opt. Spektrosk. **28**, 168 (1970) [Opt. Spectrosc. **28**, 85 (1970)].
[16] A. N. Bondarenko, G. V. Krivoshchenkov, and V. A. Smirnov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **12**, 1895 (1969).
[17] A. M. Dukhovnyĭ, D. S. Prilezhaev, B. M. Sedov *et al.*, Opt. Spektrosk. **33**, 733 (1972) [Opt. Spectrosc. **33**, 403 (1972)].
[18] I. G. Zubarev and S. I. Mikhaĭlov, Kvantovaya Élektron. (Moscow) **1**, 625 (1974) [Sov. J. Quantum Electron. **4**, 348 (1974)].
[19] L. E. Erikson and A. Szabo, Appl. Phys. Lett. **18**, 433 (1971).
[20] V. V. Kulikov, L. K. Mikhaĭlov, and S. L. Seregin, Kvantovaya Élektron. (Moscow) **7**, 432 (1980) [Sov. J. Quantum Electron. **10**, 245 (1980)].
[21] R. F. Boĭkova and É. E. Fradkin, Opt. Spektrosk. **22**, 834 (1967) [Opt. Spectrosc. **22**, 452 (1967)].
[22] L. N. Deryugin, B. P. Kulakov, and V. K. Nurmukhametov, Radiotekh. Elektron. **16**, 141 (1971).
[23] V. V. Antsiferov, N. M. Derzhi, and V. S. Pivtsov, Avtometriya, No. 5, 94 (1972).
[24] V. V. Antsiferov, N. M. Derzhi, and V. S. Pivtsov, Zh. Prikl. Spektrosk. **18**, 38 (1973).
[25] V. V. Antsiferov, A. N. Iskol'dskiĭ, and A. S. Kuch'yanov, Avtometriya No. 6, 97 (1974).
[26] V. V. Antsiferov, N. M. Derzhi, and A. S. Kuch'yanov, Kvantovaya Élektron. (Moscow) **2**, 57 (1975) [Sov. J. Quantum Electron. **5**, 30 (1975)].
[27] V. V. Antsiferov, N. S. Erokhin, and A. P. Fadeev, Preprint No. 987 [in Russian], Institute of Space Research, Academy of Sciences of the USSR Moscow (1985), 50 pp.
[28] V. V. Antsiferov, V. S. Pivtsov, and V. D. Ugozhaev, Opt. Spektrosk. **32**, 1159 (1972) [Opt. Spectrosc. **32**, 628 (1972)].
[29] V. V. Antsiferov, A. V. Gaĭner, and K. P. Komarov, Zh. Prikl. Spektrosk. **24**, 18 (1976).
[30] V. V. Antsiferov and Yu. D. Golyaev, Opt. Spektrosk. **52**, 706 (1982) [Opt. Spectrosc. **52**, 421 (1982)].
[31] V. V. Antsiferov, A. V. Chiner, and N. M. Derzi, Opt. Commun. **14**, 388 (1975).
[32] V. V. Antsiferov, Lazernaya Tekh. Optoélektron. No. 3–4(73–74), 23 (1994).
[33] V. V. Antsiferov, A. I. Alimpiev, and E. V. Ivanov, Lazernaya Tekh. Optoélektron. No. 2(58), 3 (1991).
[34] V. V. Antsiferov, E. V. Ivanov, and G. I. Smirnov, Preprint No. 93–105 [in Russian], Nuclear Physics Institute, Russian Academy of Sciences, Siberian Branch, Novosibirsk (1993), 21 pp.
[35] V. V. Antsiferov, E. V. Ivanov, and G. I. Smirnov, Preprint No. 93–106 [in Russian], Nuclear Physics Institute, Russian Academy of Sciences, Siberian Branch, Novosibirsk (1993), 20 pp.
[36] V. V. Antsiferov, E. V. Ivanov, and G. I. Smirnov, Preprint No. 93–1127 [in Russian], Nuclear Physics Institute, Russian Academy of Sciences, Siberian Branch, Novosibirsk (1993), 16 pp.
[37] R. Hultsch, Phys. Status Solidi A **47**, 415 (1978).
[38] Yu. L. Gusev, S. I. Marennikov, and V. P. Chebotaev, Pis'ma Zh. Tekh. Fiz. **3**, 305 (1977) [Sov. Tech. Phys. Lett. **3**, 124 (1977)].
[39] A. P. Maĭorov, V. K. Makukha, V. A. Smirnov *et al.*, Pis'ma Zh. Tekh. Fiz. **6**, 941 (1980) [Sov. Tech. Phys. Lett. **6**, 407 (1980)].
[40] A. P. Maĭorov, V. K. Makukha, V. A. Smirnov *et al.*, Zh. Tekh. Fiz. **51**, 2391 (1981) [Sov. Phys. Tech. Phys. **26**, 1409 (1981)].

[41] V. A. Buchenkov, A. G. Kalintsev, A. A. Mak *et al.*, Kvantovaya Élektron. (Moscow) **8**, 2239 (1981) [Sov. J. Quantum Electron. **11**, 1367 (1981)].

[42] V. A. Arkhangel'skaya, A. A. Mak, V. P. Pokrovskiĭ *et al.*, Izv. Akad. Nauk SSSR Ser. Fiz. **46**, 2012 (1982) [Bull. Acad. Sci. USSR, Phys. Ser. **46**(10), 147 (1982)].

[43] V. S. Burakov, V. A. Kononov, L. S. Korochkin *et al.*, Zh. Prikl. Spektrosk. **36**, 494 (1982).

[44] S. G. Vasil'ev, E. D. Isyanova, and V. M. Ovchinnikov, Pis'ma Zh. Tekh. Fiz. **7**, 217 (1981) [Sov. Tech. Phys. Lett. **7**, 93 (1981)].

[45] N. A. Ivanov, I. A. Parfianovich, V. M. Khulugurov *et al.*, Izv. Akad. Nauk SSSR Ser. Fiz. **46**, 1985 (1982) [Bull. Acad. Sci. USSR, Phys. Ser. **46**(10), 121 (1982)].

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Spatial solitons in photorefractive semiconductors with the Franz–Keldysh effect

V. N. Belyĭ, N. A. Khilo, and P. A. Khilo

Institute of Physics, Belarus Academy of Sciences, 220072 Minsk, Belarus

(Submitted July 26, 1996; resubmitted July 2, 1997)
Zh. Tekh. Fiz. **68**, 80–84 (October 1998)

The features of the nondiffractive propagation of light beams in photorefractive semiconductors under conditions for manifestation of the Franz–Keldysh effect are investigated. It is shown that the combined influence of electrorefraction and electroabsorption in such materials leads to the formation of inhomogeneous spatial solitons, for which the directions of energy propagation and the phase velocity are different. The dependence of the soliton shape on the parameters of the photorefractive semiconductor and the external electric field strength is established. © *1998 American Institute of Physics.* [S1063-7842(98)01410-X]

INTRODUCTION

Nonivergent light beams, or spatial solitons, are a subject of intensive theoretical and experimental research.[1–5] Spatial solitons appear as a result of nonlinear variation of the refractive index of a material, under which the diffractive divergence of a beam is compensated exactly by the nonlinear refraction. Most of the research on spatial solitons has been performed in reference to media with Kerr nonlinearity. This nonlinearity directly influences the phase of the light field and thus also permits compensation of the phase diffraction effect. However, the nonlinear susceptibility χ_3, which is responsible for the Kerr effect, is relatively small; therefore, light intensities $\geqslant 1$ MW/cm^2 are needed to form nonivergent beams. In this respect the spatial solitons recently discovered and investigated in photorefractive (PR) crystals[1–9] have considerable advantages. In contrast to the case of media with Kerr nonlinearity, the formation of solitons by the photorefraction mechanism takes place already at electric field strengths <1 W/cm^2.

The existence of three different types of PR spatial solitons has been established. The first type is due to the nonlocal nature of the PR effect, which is manifested by a dependence of the perturbation δn of the refractive index on the transverse gradient of the intensity $I(r)$ of the light beam: $\delta n \propto \nabla_\perp I(r)/I(r)$. The treatment of solitons of this type is based on utilization of the response function of two-wave mixing as applied to each pair of plane-wave components comprising the beam.[1] Solitons of the first type are quasistationary, since they exist during the temporal window from the moment when the PR gratings are formed to the moment when the external electric field is shorted by the conduction currents. In addition, quasistationary solitons exist within a strictly defined range of external electric field strengths, departure from which leads to their destruction.[1,2]

The photorefractive solitons of the second type, which have been called screening solitons,[3] are local with respect to the mechanism for their formation and exist under stationary conditions.[4,5] The conduction currents play a decisive role in the formation of a screening soliton, and the contribution of the diffusion current can be neglected. Physically, the formation of a screening soliton is attributed to the fact that a beam intensity which is nonuniform over a transverse section also excites a nonuniform concentration of charge carriers (electrons), which drift in the external electric field applied to the crystal and are captured in deep traps. This leads to the formation of a nonuniform space-charge field in a transverse section of the beam and, consequently, to nonuniform shorting of the external electric field (nonuniform screening). Ultimately, a self-consistent optical waveguide is induced in the crystal as a result of the linear electrooptic effect.[3]

A third type of soliton exists in PR crystals with a strong photovoltaic effect, such as lithium niobate.[6,7] The light field in these crystals induces photovoltaic currents, which cause variation of the refractive index that coincides in functional form with the nonlinearity of the saturating absorber. Photovoltaic solitons are sensitive to the ratio I/I_d (of the light intensity I to the equivalent dark intensity I_d). This sets them apart from solitons of the first type, which, as was noted above, are governed by nonlinearity of the gradient type.

The formation of spatial solitons of the first type in PR crystals with nonlocal nonlinearity is hampered in the general case by the transfer of energy between plane-wave components of the light beam upon diffraction on phase-shifted diffraction gratings. Therefore, in Refs. 1 and 2 spatial solitons were investigated under conditions with suppression of the phase shift of the PR gratings by an external electric field. However, as was shown in Refs. 8 and 9, spatial solitons exist in PR crystals with a drift mechanism of nonlinearity even when there is a phase shift of the optical gratings and the PR gratings induced by them. In this case the energy-exchange processes lead to variation of the spatial shape of the soliton and, more specifically, to the appearance of transverse asymmetry.[8]

The types of solitons considered in Refs. 1–5 and 7–9 exist in PR ferroelectrics with a strong linear electrooptic effect. This does not apply to the extensive class of optical materials known as PR semiconductors, in which a two-field interaction is possible due to the Franz–Keldysh effect.[10] If the frequency ω of the light beam lies near the fundamental absorption edge ω_g of the semiconductor and if the inequality $\omega < \omega_g$ holds, the space-charge field E_{sc} induced by the interference light field forms both a refractive-index grating (electrorefraction) and an absorption grating (electroabsorption).

This raises the important question of the possibility of the existence of spatial solitons in PR semiconductors with an amplitude–phase type of photoinduced diffraction gratings. Thus, the purpose of the present work is to investigate spatial solitons in PR semiconductors under conditions for manifestation of the Franz–Keldysh effect.

BASIC EQUATIONS

Following Ref. 1, we consider the model of a two-dimensional light beam propagating in a PR semiconductor along the Z axis and diffracting in the direction of the X axis. We neglect the diffraction effects along the Y axis. An external electric field E is applied to enhance the PR effect along the X axis. The XYZ coordinate system is chosen in such a manner that no linear electrooptic effect is manifested for the assigned beam polarization and the electrorefraction effect is wholly responsible for soliton formation. For example, $Z \| [110]$, $X \| [001]$, and a beam polarization coinciding with the $[001]$ direction should be chosen for a cubic GaAs crystal.

The equation describing the evolution of the light field in a PR semiconductor with a nonlocal response has the form

$$\left(\frac{\partial}{\partial z} - \frac{i}{2k_0 \overline{n}} \frac{\partial^2}{\partial x^2} \right) A(x,z) = i k_0 \delta\overline{n}(x,z) A(x,z), \tag{1}$$

where $A(x,z)$ is the two-dimensional field amplitude, $k_0 = 2\pi/\lambda$, and $\overline{n} = n' + i n''$ is the unperturbed complex refractive index.

Equation (1) with a zero right-hand side describes the diffractive and diffusive spreading of a light beam in a linear absorbing medium. The nonlinear addition $\delta\overline{n}(x,z)$ is obtained by considering the pair-mixing processes of the plane-wave beam components[1,8]

$$\delta\overline{n}(x,y) = \frac{1}{|A(x,z)|^2} \int A(x-\rho,z) A^* \times (x+\rho',z) \chi(\rho,\rho')\, d\rho\, d\rho', \tag{2}$$

where

$$\chi(\rho,\rho') = \frac{1}{(2\pi)^2} \int \hat{\chi}(q,q') \exp[i(q\rho + q'\rho')]\, d\rho\, d\rho'$$

is the nonlinear nonlocal response function, which is determined by the parameters of the PR crystal, as well as by the magnitude and direction of the external electric field. The two-dimensional Fourier transform $\hat{\chi}(q,q')$ of the nonlocal response function is proportional to the usually employed photorefractive coupling constant in the theory of two-wave mixing.

Let us switch from an integrodifferential equation to a differential equation in (1) by expanding the amplitudes $A(x-\rho,z)$ and $A^*(x+\rho',z)$ in series in the small displacements ρ and ρ' and confining ourselves to second-order terms. Then

$$\begin{aligned}
\delta\overline{n}(x,z) = {} & \int \chi(\rho,\rho') d\rho d\rho' + \frac{\partial A(x,z)}{\partial x} \frac{1}{A(x,z)} \\
& \times \int \rho \chi(\rho,\rho')\, d\rho\, d\rho' - \frac{\partial A^*(x,z)}{\partial x} \frac{1}{A^*(x,z)} \\
& \times \int \rho \chi(\rho,\rho')\, d\rho\, d\rho' + \frac{\partial^2 A(x,z)}{\partial x^2} \frac{1}{2A(x,z)} \\
& \times \int (\rho')^2 \chi(\rho,\rho')\, d\rho\, d\rho' + \frac{\partial^2 A^*(x,z)}{\partial x^2} \\
& \times \frac{1}{2A^*(x,z)} \int \rho^2 \chi(\rho,\rho')\, d\rho\, d\rho' - \frac{\partial A(x,z)}{\partial x} \\
& \times \frac{\partial A^*(x,z)}{\partial x} \frac{1}{|A(x,z)|^2} \int \rho\rho' \chi(\rho,\rho')\, d\rho\, d\rho'.
\end{aligned} \tag{3}$$

The values of the integrals appearing in (3) depend significantly on the symmetry of $\chi(\rho,\rho')$. To determine the symmetry it is convenient to start out from an analysis of the Fourier components $\hat{\chi}(q,q')$, which describe the nonlocal response of the crystal to the interaction of a pair of plane waves. For the Franz–Keldysh semiconductor crystals under consideration here, we can write[11,12]

$$\hat{\chi}(q,q') = r' E^2(q,q') + i r'' E^m(q,q'), \tag{4}$$

where r' and r'' are constants, which describe electrorefraction and electroabsorption; $E(q,q') = -(E_1(q,q') + iE_2(q,q'))$; $E_1(q,q')$ and $E_2(q,q')$ are the strengths of the electric fields in the crystal responsible for the formation of cophasal and $\pi/2$-shifted diffraction gratings, which depend on the transverse components q and q' of the wave vectors of the pair of interacting waves.

For most semiconductors the electrorefraction is quadratic with respect to the electric field over a broad range of variation of its strength. In particular, the empirical value for GaAs in the range of fields up to 600 kV/cm is $r' \approx 3.45 \times 10^{-16} \exp(3/\lambda^3)$ in units of $(\mathrm{cm/V})^2$ at the wavelengths $0.9 < \lambda < 1.55$ μm (Ref. 11). At the same time, the field dependence of the electroabsorption is more complicated.[11] For GaAs in the range of variation of the field from several tens of kV/cm to ~300 kV/cm the exponent in (4) varies from $m \approx 7$ to 3.

The symmetry of $\hat{\chi}(q,q')$ in (4) can easily be determined for any value of m on the basis of the known symmetry properties of the fields $E_1(q,q')$ and $E_2(q,q')$:

$$\begin{aligned}
& E_1(q,q') = E_1(q',q) = E_1(-q,-q'), \\
& E_2(q,q') = E_2(q',q) = -E_2(-q,-q').
\end{aligned} \tag{5}$$

Allowing for the relations (5), we can represent $\hat{\chi}(q,q')$ in the form of a sum of two terms of different symmetry: $\hat{\chi}(q,q') = \hat{\chi}_c(q,q') + \hat{\chi}_a(q,q')$, where the symmetry of

$\hat{\chi}_c(q,q')$ and $\hat{\chi}_a(q,q')$ coincides with the symmetry of $E_1(q,q')$ and $E_2(q,q')$, respectively. For example, in the case of $m=3$ we have

$$\hat{\chi}_c(q,q')=r'(E_1^2(q,q')-E_2^2(q,q'))$$
$$+ir''(E_1^3(q,q')-3E_2^2(q,q')E_1(q,q')),$$
$$\hat{\chi}_a(q,q')=r''(E_2^3(q,q')-3E_1^2(q,q')E_2(q,q'))$$
$$+2ir'E_1(q,q')E_2(q,q').$$

Returning to the coordinate representation of the response, in analogy to the foregoing we obtain $\chi(\rho,\rho')=\chi_c(\rho,\rho')+\chi_a(\rho,\rho')$, where

$$\chi_c(\rho,\rho')=(2\pi)^{-2}\int\hat{\chi}_c(q,q')\cos(q\rho+q'\rho')\,dq\,dq',$$

$$\chi_a(\rho,\rho')=i(2\pi)^{-2}\int\hat{\chi}_a(q,q')\sin(q\rho+q'\rho')\,dq\,dq', \quad (6)$$

Next, using (6), we can find the values of the integrals appearing in (3) for each concrete case of the dependence of $\hat{\chi}(q,q')$ on $E(q,q')$.

SOLUTION AND ANALYSIS

Moving on to the solution of Eq. (1), we note that the case of weak absorption is of greatest practical interest. When the characteristic length of the PR crystal $L\sim 1$ cm, to maintain an acceptable level for the output signal it must be assumed that the absorption coefficient $2k_0n''\leqslant 1$ cm^{-1}. Furthermore, when the characteristic transverse dimension of the beam $d\sim 10^{-2}$ cm, the influence of absorption on the diffractive spreading can be estimated by the multiplicative factor $\exp(-k_0n''d^2/L)\approx 1-2.5\times 10^{-5}$. Thus, under the typical conditions considered for the propagation of a spatial soliton, its diffusive spreading is smaller. A similar conclusion can also be drawn regarding the influence of electric-field-induced absorption on beam propagation. On the basis of this estimate we shall henceforth neglect the imaginary parts of the integrals in (3), which are quadratic in ρ and ρ'. We seek a solution of Eq. (1) in the form

$$A(\rho,z)=\alpha[(\rho-\tan(\Theta)z)]\exp(i\gamma z), \quad (7)$$

where $\gamma=\gamma'+i\gamma''$ is the complex soliton propagation constant and $\tan(\Theta)$ takes into account the possible disparity between the directions of the energy flow and the phase velocity.

Substituting (7) into (1) and separating the real and imaginary parts, we obtain two equations:

$$\frac{1}{k_0}\frac{\partial\alpha}{\partial\rho}-I'_{00}-\left(\frac{1}{2k_0^2n'}+I'_{20}\right)\frac{1}{\alpha}\frac{\partial^2\alpha}{\partial\rho^2}$$
$$-\frac{2I'_{01}}{\alpha}\frac{\partial\alpha}{\partial\rho}+\frac{I'_{11}}{\alpha^2}\left(\frac{\partial\alpha}{\partial\rho}\right)^2=0, \quad (8)$$

$$\frac{\gamma''}{k_0}-I''_{00}+\left(\frac{\tan(\Theta)}{k_0}+2I''_{01}\right)\frac{1}{\alpha}\frac{\partial\alpha}{\partial\rho}=0, \quad (9)$$

where

$$I_{00}=\int\chi_c(\rho,\rho')\,d\rho\,d\rho',$$

$$I_{01}=\int\rho'\chi_a(\rho,\rho')\,d\rho\,d\rho',$$

$$I_{20}=\int\rho^2\chi_c(\rho,\rho')\,d\rho\,d\rho',$$

$$I_{11}=\int\rho\rho'\chi_c(\rho,\rho')\,d\rho\,d\rho', \quad (10)$$

$I'_{mn}=\mathrm{Re}(I_{mn})$, and $I''_{mn}=\mathrm{Im}(I_{mn})$.

The solution of Eq. (8) has the form

$$\alpha(x,z)=\alpha_0\exp\left[\frac{P'_{01}(x-\tan(\Theta)z)}{2(P'_{11}-P'_{20})}\right]$$
$$\times[\sec h(P'(x-\tan(\Theta)z))]^D$$
$$\times\exp(i\gamma'z-\gamma''z), \quad (11)$$

where

$$P'_{01}=2k_0I'_{01},\quad P''_{11}=k_0I'_{11},$$

$$P'_{20}=k_0I'_{20}+\frac{1}{2k_0n'},\quad D=\frac{P'_{20}}{P'_{11}-P'_{20}},$$

$$P'=\frac{1}{2P'_{20}}\sqrt{(P'_{01})^2+4(P'_{11}-P'_{20})\gamma'},$$

$$\overline{\gamma'}=k_0I'_{00}-\gamma'. \quad (12)$$

In addition, from (9) we find the parameters for energy absorption (γ'') and drift $[\tan(\Theta)]$:

$$\gamma''=k_0I''_{00},\quad \tan(\Theta)=2k_0I''_{10}. \quad (13)$$

It follows from (11) and (13) that the occurrence of electroabsorption leads to two features in the behavior of a spatial soliton. First, additional damping of the soliton with the coefficient γ'' and, second, beam energy drift up to the angle $\arctan(2k_0I''_{10})$ to the direction of the phase velocity are induced. It is important here that neither of these effects, which influence absorption, destroy the soliton.

Let us ascertain the origin of soliton energy drift on the physical level. For this purpose, we isolate the addition $\delta n''$ to the absorption coefficient, which is proportional to I''_{10}, from (3). Then, with allowance for the explicit form of the field $A(x,z)$ we obtain

$$\delta n''(x,z)=\left(\frac{P'_{01}}{2(P'_{11}-P'_{20})}+\frac{P'}{2(P'_{11}-P'_{20})}\right.$$
$$\left.\times\tanh[p'(x-z\tan(\Theta))]\right)I''_{01}. \quad (14)$$

It is seen from (14) that $\delta n''(x,z)$ contains a contribution $\sim\tanh[p'(x-z\tan(\Theta))]$, which is spatially asymmetric with respect to the beam axis. Because of this term, a decrease in the energy of the beam in one of its halves is compensated by an increase in energy in the other. Thus, there is energy transfer within the beam accompanied by drift as a whole. To estimate tan(Θ), from (10) and (6) we find $I''_{10} =(\partial\hat{\chi}_a(q,q')/\partial q)_{q,q'=0}$. Substituting therein $\hat{\chi}_a(q,q')$ from (4) and the explicit form of $E_{1,2}(q,q')$ (see, for example, Refs. 1 and 13) and performing the differentiation, we obtain $\tan(\Theta)=4\pi r'' E_0^{m+1}/P\lambda$, where E_0 is the external electric field strength, $P=eN/\varepsilon_0\varepsilon_r$, N is the concentration of traps in the PR crystal, and ε_r is the dielectric constant. For GaAs:EL2 we have $\varepsilon_r=12$, and we take $N=10^{15}$ cm^{-3}. Then, for $E_0=10^5$ V/cm we have $\Delta n''\approx 1\times 10^{-4}$ (Ref. 11) and an energy drift angle $\Theta\approx 1°$. The value of the drift angle depends strongly on applied field. For example, already when $E_0=5\times 10^4$ V/cm, we obtain $\Theta\approx 0.01°$. As we see, at relatively low fields energy drift is insignificant and is of interest in the investigation of the stability of a soliton toward small losses. In the case where $\Theta\sim 1°$ this effect is of practical interest for controlling the direction of soliton motion by varying the external field.

Returning to the solution (11), we note that, apart from energy drift, another characteristic feature of a spatial soliton in a PR semiconductor is its spatial asymmetry. The appearance of asymmetry is caused by the self-interaction in the beam accompanying diffraction on the phase-shifted gratings. Under conditions where the influence of the phase-shifted diffraction gratings can be neglected (see the experimental studies in Ref. 2) $P'_{01}=0$, and, according to (11), the soliton shape is a symmetric function of the transverse coordinate.

CONCLUSIONS

We have investigated the features of the nondiffractive propagation of light beams (spatial solitons) in photorefractive semiconductors under conditions for manifestation of the Franz–Keldysh effect. We have shown that, despite the presence of absorption gratings, i.e., an amplitude–phase character of the nonlinear nonlocal response function, electrorefractive spatial solitons form in the semiconductor. The occurrence of electroabsorption causes additional damping of the soliton as a whole, as well as transverse drift of its energy flow direction. Apart from energy drift, another characteristic feature of an electrorefractive soliton is its spatial asymmetry, which is caused by a self-effect in the beam accompanying diffraction on the phase-shifted gratings.

Under conditions for manifestation of the Franz–Keldysh effect, solitons are efficiently formed in the near-IR region of the spectrum. This is important for the optical processing of information using semiconductor lasers, since it expands the possibilities for the realization of optical interconnections in information-processing and communication systems. It should be stressed that electrorefractive solitons exist in semiconductors of arbitrary symmetry, including an extensive group of centrosymmetric crystals and polycrystalline materials, provided these materials contain traps for sustaining photoinduced space-charge fields. The other important merits of solitons in semiconductors include: a) their high rate of formation in comparison to ferroelectric crystals, b) their electrorefractive nonlinearity, which is larger than the linear electrooptic effect.

[1] B. Crosignani, M. Segev, D. Engin *et al.*, J. Opt. Soc. Am. B **10**, 446 (1993).

[2] G. Duree, J. L. Shultz, G. Salamo *et al.*, Phys. Rev. Lett. **71**, 533 (1993); Opt. Lett. **74**, 1978 (1995).

[3] M. Segev, G. C. Valley, B. Crosignani *et al.*, Phys. Rev. Lett. **73**, 3211 (1994).

[4] D. N. Christodoulides and M. I. Carvalho, J. Opt. Soc. Am. B **12**, 1628 (1995).

[5] M. D. Iturbe Castillo, P. A. Marquez Anguilar, J. J. Sanchez-Mondragon *et al.*, Appl. Phys. Lett. **62**, 408 (1994).

[6] C. Gu, J. Hong, Hsin-Yu *et al.*, J. Appl. Phys. **69**, 1167 (1991).

[7] G. C. Valley, M. Segev, B. Crosignani *et al.*, Phys. Rev. A **50**, R4457 (1994).

[8] V. N. Belyĭ and N. A. Khilo, Pis'ma Zh. Tekh. Fiz. **20**(18), 40 (1994) [Tech. Phys. Lett. **20**(9), 740 (1994)].

[9] V. N. Belyi, N. S. Kazak, V. K. Pavlenko *et al.*, in *15th International Conference on Coherent and Nonlinear Optics*, St. Petersburg (1995), Vol. 1, pp. 385–386.

[10] A. Partovi, A. Kost, E. M. Garmire *et al.*, Appl. Phys. Lett. **56**, 1089 (1990).

[11] J. G. Mendoza, L. A. Coldren, A. Apling *et al.*, J. Lightwave Technol. **6**, 793 (1983).

[12] B. R. Bennett and R. A. Soref, IEEE J. Quantum Electron. **QE-23**, 2159 (1987).

[13] M. P. Petrov, S. I. Stepanov, and A. V. Khomenko, *Photosensitive Electrooptic Media in Holography and Optical Information Processing* [in Russian], Leningrad (1983), 270 pp.

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Magnetostatic analog of Love surface elastic waves

I. A. Kaĭbichev and V. G. Shavrov

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 103907 Moscow, Russia

(Submitted December 9, 1996; resubmitted June 3, 1997)
Zh. Tekh. Fiz. **68**, 85–90 (October 1998)

The possibility of the existence of a magnetostatic analog of Love surface elastic waves is predicted. They appear in situations where the conditions for the existence of magnetostatic volume waves hold in the upper layer of a ferromagnetic bilayer, but not in the lower layer. © *1998 American Institute of Physics.* [S1063-7842(98)01510-4]

INTRODUCTION

The existence of Love elastic waves in a solid-layer/space system was first predicted theoretically by Bromwich for the long-wavelength region,[1] and a more detailed investigation for arbitrary frequencies was conducted in Ref. 2. They have been observed experimentally[3] and utilized in acoustoelectronics.[4–6] In recent years there has been rapid development of the research in the physics and technology of microwave devices.[7,8] A great deal of attention has been focused on the investigation of the properties of magnetostatic surface waves (MSSWs) in ferromagnetic bilayer and multilayer films containing layers with different saturation magnetizations for the purpose of using them in delay lines, filters, directional couplers, and nonreciprocal devices.[9–13] For a long time there have been differences in opinion among researchers regarding the number of branches in the MSSW spectrum. Some investigators[9,10] assumed that there is only one branch, the existence of one or two branches, depending on the propagation direction, was assumed in Refs. 14 and 15, and the existence of two branches was demonstrated in Refs. 16–18. There were also disputes regarding the type of wave: in Refs. 9–11 and 15 they were forward waves, while in Refs. 14, 16, 17, and 19 they were forward for some wave vectors k and backward for others. These inconsistencies were eliminated in Ref. 19, where it was shown that the number of branches in the spectrum is equal to the number of ferromagnetic layers and does not depend on the wave propagation direction. In addition, it was shown that the MSSWs propagating over the external surfaces of a film are forward waves, while those propagating over the internal surfaces are forward waves when $|k|<k_0$ and backward waves when $|k|>k_0$ (k_0 is the limiting value of the wave number, which equals $\sim 10^3$ cm^{-1}). Bilayer films can also have different magnetic anisotropy constants. This situation has escaped the attention of researchers, and new results can naturally be expected for MSSWs in them. A difference between the magnetic anisotropy constants in the layers with an appropriate choice of parameters permits, in principle, the realization of a case which is analogous to the conditions for the existence of Love surface elastic waves: only magnetostatic volume waves (MSVWs) can exist in the upper layer taken separately, while they cannot exist in the lower layer taken separately. Then, magnetostatic waves can propagate in a film consisting of such layers in a definite frequency range, the magnetic potential oscillating in the upper layer and decaying exponentially in the lower layer. The MSSWs thus obtained are magnetostatic analogs of Love surface elastic waves.

STATEMENT OF THE PROBLEM AND BASIC EQUATIONS

Let us consider a ferromagnetic bilayer film with different values of the magnetic anisotropy field $H_{Aj}=2K_j/M_{0j}$ ($j=1$ for $0\leqslant z\leqslant d_l$ and $j=2$ for $d_2\leqslant z\leqslant 0$), where K_j is the magnetic anisotropy constant and M_{0j} is the saturation magnetization in the respective layer. We assume that the easy magnetization axes of the ferromagnets coincide and are directed parallel to the Z axis, which is perpendicular to their surfaces (Fig. 1). The external static magnetic field H_0 is oriented parallel to the Z axis and perpendicularly to the surfaces of the ferromagnets. In the ground state the magnetization vectors are inclined at an angle Ψ_{0j} from the normal n to the surface. We assume that the magnetic anisotropy field $H_{Aj}<4\pi M_{0j}$. Then in the weak magnetic fields $H_0<\min(4\pi M_{01}-H_{A1},\ 4\pi M_{02}-H_{A2})$ the state with magnetization vectors inclined from the easy axes by

$$\Psi_{0j}=\pm\arccos(H_0/[4\pi M_{0j}-H_{Aj}]),\quad j=1,2,\tag{1a}$$

is stable, while in the strong fields $H_0\geqslant\max(4\pi M_{0j}-H_{Aj})$ the ground state with magnetization vectors directed along the easy axis is stable and

$$\Psi_{0j}=0.\tag{1b}$$

The latter configuration is also observed in the case of $H_{Aj}\geqslant 4\pi M_{0j}$ and $H_0\geqslant 0$. When $H_{Aj}\geqslant 4\pi M_{0j}$ and the condition on the strength of the external magnetic field $\max(4\pi M_{0j}-H_{Aj})\leqslant H_0\leqslant 0$ holds, phase (1b) is metastable. Therefore, in the latter case the results for the MSSW spectrum are valid only if the energy of the magnetic excitations of the ferromagnet is small compared with the potential barrier preventing passage of the ferromagnet into the homogeneous stable state with $\Psi_0=\pi$. The exchange interaction was not taken into account in determining the ground state. This is possible, if the dimensions of the ferromagnetic film are greater than the exchange length. The boundary conditions for continuity of the tangential components of the magnetic field and the normal components of the magnetic

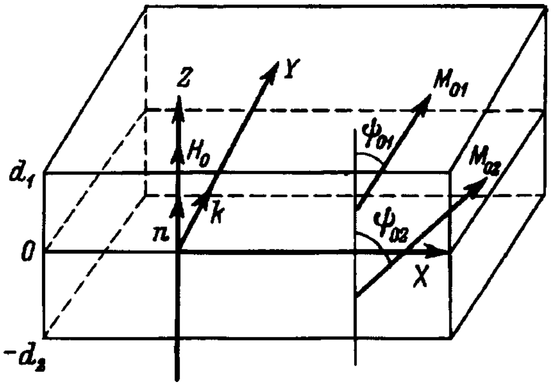

FIG. 1. Ferromagnetic bilayer film. It is assumed that the saturation magnetizations M_{0j} of the layers lie in the ZX plane; in the general case they form an angle Ψ_{0j} with a normal to the surface. The MSSWs propagate along the Y axis with the wave vector k.

induction vector must hold on the contact plane ($z=0$) between the two ferromagnetic layers. This calls for fulfillment of the conditions

$$M_{01}\cos\Psi_{01}=M_{02}\cos\Psi_{02}\rightarrow\frac{M_{01}}{4\pi M_{01}-H_{A1}}=\frac{M_{02}}{4\pi M_{02}-H_{A2}} \tag{2a}$$

for the ground states with magnetization vectors inclined from the easy axes (1a) and

$$M_{01}=M_{02}=M_0 \tag{2b}$$

for the ground state with magnetization vectors directed along the easy axes (1b). If the condition (2) does not hold, an inhomogeneous transitional region appears in the bilayer film near the contact plane between the layers, but such a situation will not be considered.

We assume that the MSSWs propagate along the Y axis; therefore, all the variables of the problem are proportional to $\exp(i\omega t-iky)$, where ω is the frequency. We shall henceforth consider the frequency range up to several gigahertz, since it is usually used in practice.[20–23] At such frequencies the wave vector $k\leqslant10^5\ \mathrm{cm}^{-1}$. In this range the contribution of the exchange interaction is small compared with the other terms in the magnetic energy, viz., the dipole–dipole and Zeeman terms. When magnetostatic waves are considered, it can be disregarded. In determining the equations describing the propagation of MSSWs, we shall consider only one of the ferromagnetic layers. The derivation procedure for the other layer is similar, the difference being confined to the values of the magnetic anisotropy field and the magnetization. We start out from the system of magnetostatic equations

$$\mathrm{curl}H_M=0,\quad \mathrm{div}(H_M+4\pi M)=0. \tag{3}$$

Here H_M is the demagnetizing field. The magnetization M satisfies the Landau–Lifshitz equation

$$\partial M/\partial t=-\gamma[M\cdot H_{\mathrm{eff}}], \tag{4}$$

where γ is the gyromagnetic ratio, $H_{\mathrm{eff}}=H_0+H_M+H_A(Mn)n/M_0$ is the effective magnetic field, and n is the unit vector characterizing the direction of the anisotropy axis of the ferromagnetic crystal.

In the ground state the magnetization vector has the components $(M_0\sin\Psi_0,\ 0,\ M_0\cos\Psi_0)$, and the demagnetizing field has the components $(0,\ 0,\ -4\pi M_0\cos\Psi_0)$. The deviations of the magnetization vector m and the demagnetizing field h from these equilibrium values are assumed to be small. We linearize the Landau–Lifshitz equation (4), and, as a result, we obtain the relationship between the components of m and h, which we write in the form

$$m_i=\chi_{ij}h_j,\quad i,j=x,y,z, \tag{5}$$

where χ_{ij} is the high-frequency magnetic susceptibility tensor of the ferromagnet.

Its elements have the form

$$\chi_{xx}=\Gamma\Omega_1\cos^2\Psi_0,\quad \chi_{xy}=-\chi_{yx}=\Gamma i\omega\cos\Psi_0,$$

$$\chi_{yy}=\Gamma\Omega_2,\quad \chi_{yz}=-\chi_{zy}=\Gamma i\omega\sin\Psi_0,$$

$$\chi_{zx}=\chi_{xz}=-\Gamma\Omega_1\cos\Psi_0\sin\Psi_0,\quad \chi_{zz}=\Gamma\Omega_1\sin^2\Psi_0.$$

Here we have used the notation

$$\Gamma=\gamma M_0/[\Omega_1\Omega_2-\omega^2],$$

$$\Omega_1=\gamma[H_0^{(i)}+H_A\cos\Psi_0]\cos\Psi_0,$$

$$\Omega_2=\gamma[H_0^{(i)}\cos\Psi_0+H_A\cos2\Psi_0],$$

and $H_0^{(i)}=H_0-4\pi M_0\cos\Psi_0$ is the internal magnetic field. We note that after going over to the coordinate system with the Z axis coinciding with the magnetization of the ground state of the ferromagnet, we obtain the previously known form of the high-frequency magnetic susceptibility tensor.[24] After substituting (5) into the magnetostatic equations and introducing the magnetic scalar potential $\Phi(h=-\nabla\Phi)$, we obtain the second-order differential equation

$$D^2\Phi(z)-k^2Q(\omega)\Phi(z)=0, \tag{6}$$

where $Q(\omega)=1+4\pi\gamma M_0\Omega_2/[\Omega_1\Omega_2-\omega^2]$ is a function which specifies the character of the solutions and is termed the spectral function and $D^2=\partial^2/\partial z^2$.

After calculating the frequencies Ω_1 and Ω_2 for the ground states (1), we find that

$$Q(\omega)=[\omega^2-\omega_{vU}^2]/\omega^2 \tag{7a}$$

for the ground state with magnetization inclined from the easy axis (1a) and

$$Q(\omega)=[\omega_{vU}^2-\omega^2]/[\omega_{vL}^2-\omega^2] \tag{7b}$$

for the phase with magnetization along the easy axis (1b). Here ω_{vL} and ω_{vU} are the lower and upper limits of the MSVW spectrum of a homogeneous plate. Their form depends on the choice of the ground state. For the ground state with a magnetization vector inclined from the easy axis (1a)

$$\omega_{vL}=0,\quad \omega_{vU}=2\gamma|\sin\Psi_0|\sqrt{-\pi M_0H_A}, \tag{8a}$$

and for phase (1b)

$$\omega_{vL}=\Omega_0,\quad \omega_{vU}=\sqrt{\Omega_0[\Omega_0+4\pi\gamma M_0]}, \tag{8b}$$

where $\Omega_0=\gamma[H_0-4\pi M_0+H_A]$ is the homogeneous ferromagnetic resonance frequency.

Let us determine the form of the boundary conditions with consideration of two ferromagnetic layers. The variables referring to the upper layer are marked by a subscript 1 and those referring to the lower layer are labeled by a subscript 2. The lower layer will be treated below as a half space (i.e., $|k|d_2\gg 1$). Then the boundary conditions of the problem reduce to continuity of the normal component of the magnetic induction and the tangential component of the magnetic field intensity on the surface of the upper layer $(z=d_1)$ and the interface $(z=0)$. This is equivalent to the following boundary conditions on the magnetic potential:

$$\Phi_1(d_1)=\Phi_B(d_1),\quad \Phi_1(0)=\Phi_2(0),$$

$$-D\Phi_1(d_1)+4\pi m_{z1}(d_1)=-D\Phi_B(d_1),$$

$$-D\Phi_1(0)+4\pi m_{z1}(0)=-D\Phi_2(0)+4\pi m_{z2}(0). \tag{9}$$

Here $\Phi_j(z)$ is the magnetic potential in the jth layer, $D=\partial/\partial z$ is the derivative with respect to z, $\Phi_B(z)$ is the magnetic potential in the vacuum region $(z\geqslant d_1)$

$$\Phi_B(z)=\Phi_0\exp[-|k|(z-d_1)], \tag{10}$$

where Φ_0 is a constant.

Thus, to investigate the propagation of MSSWs in a two-layer ferromagnetic film we must solve the second-order differential equation for the magnetic potential (6) with the boundary conditions (9).

DISPERSION RELATION FOR MSSWS

We assume that MSVWs can propagate, i.e., that the spectral function $Q_1(\omega)<0$, in the upper layer and that they cannot, i.e., that $Q_2(\omega)>0$, in the lower layer. Then Eq. (6) has the solutions

$$\Phi_1(z)=A\cos[|k|(z-d_1/2)\sqrt{-Q_i(\omega)}]$$
$$+B\sin[|k|(z-d_1/2)\sqrt{-Q_i(\omega)}] \tag{11a}$$

for the upper layer $(0\leqslant z\leqslant d_1)$ and

$$\Phi_2(z)=C\exp[|k|z\sqrt{Q_2(\omega)}] \tag{11b}$$

for the lower layer $(-d_1\leqslant z\leqslant 0)$. The lower layer can be regarded here as a half space.

After substituting the solutions (11) into the boundary conditions (9), we obtain the dispersion relation

$$|k|d_1\sqrt{-Q_i(\omega)}$$
$$=\operatorname{arccot}\left\{-\frac{Q_1(\omega)+\beta_1(\omega)[\sqrt{Q_2(\omega)}+\beta_2(\omega)-\beta_1(\omega)]}{\sqrt{-Q_1(\omega)}[\beta_2(\omega)+\sqrt{Q_2(\omega)}]}\right\}$$
$$+\pi N,$$
$$N=0,1,2, \tag{12}$$

where $\beta_j(\omega)=1+4\pi\gamma M_{0j}\sigma\sin\Psi_{0j}/\omega$; $j=1,2$ for the upper and lower layers, respectively, and $\sigma=\operatorname{sgn}k$.

It can be seen from (12) that there is a whole frequency band for waves corresponding to different modes with the numbers N. Such wave resemble Love surface elastic waves,[25,26] and we shall call them magnetostatic surface waves. The distribution of the magnetic potential in such MSSWs oscillates in the upper layer $(0\leqslant z\leqslant d_1)$ according to the law

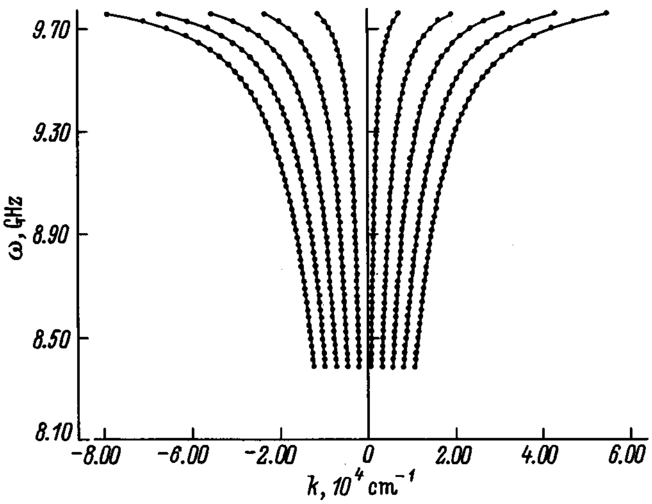

FIG. 2. Spectrum of the magnetostatic analog of Love waves (modes with the numbers $N=0,1,2,3,4,5$) when the ferromagnetic layers contain ground states inclined from the easy axis. Film parameters: $K_1=-1.24\times 10^4$ erg/cm^3, $K_2=-9.007\times 10^3$ erg/cm^3, $H_0=100$ Oe, $M_{01}=140.06$ G, $M_{02}=119.37$ G, $\omega_{vU1}=9.8432\times 10^9$ Hz, $\omega_{vU2}=8.3849\times 10^9$ Hz, $d_1=2\times 10^{-3}$ cm.

$$\Phi_1(z)=A\{\cos[|k|(z-d_1/2)\sqrt{-Q_1(\omega)}]$$
$$+E\sin[|k|(z-d_1/2)\sqrt{-Q_1(\omega)}]\}, \tag{13a}$$

and decays exponentially in the lower layer $(-d_2\leqslant z\leqslant 0)$ according to the law

$$\Phi_2(z)=A\{\cos\alpha-E\sin\alpha\}\exp[|k|z\sqrt{Q_2(\omega)}] \tag{13b}$$

and in the vacuum $(z\leqslant d_1)$ according to the law

$$\Phi_B(z)=A\{\cos\alpha+E\sin\alpha\}\exp[-|k|(z-d_1)]. \tag{13c}$$

Here

$$E=\frac{\sqrt{-Q_1(\omega)}\sin\alpha-\beta_1(\omega)\cos\alpha}{\sqrt{-Q_1(\omega)}\cos\alpha+\beta_1(\omega)\sin\alpha},$$

$$\alpha=|k|\frac{d_1}{2}\sqrt{-Q_1(\omega)}.$$

Thus the dispersion relation (12) describes a MSSW, in which the magnetic potential oscillates in the upper layer and decays exponentially in the lower layer (13). The form of the dispersion curves depends on the choice of the ground state.

MSSW SPECTRUM IN A FERROMAGNETIC BILAYER FILM WITH MAGNETIZATION VECTORS INCLINED FROM THE EASY AXIS

Let us assume that both layers of the film contain states with magnetization vectors inclined from the easy axis (1a). Then the real roots of the dispersion relation (12) can exist only in the case of $\omega_{vU2}<\omega_{vU1}$. The MSSW spectrum (Fig. 2) resembles the MSVW spectrum in a ferromagnetic plate, the differences being confined to the low-frequency cutoff at the frequency ω_{vU2} and the wave vectors

$$|k_N|=\frac{\omega_{vU2}}{d_1\sqrt{\omega_{vU1}^2-\omega_{vU2}^2}}\times\left\{\operatorname{arccot}\left[\frac{\omega_{vU1}^2-\omega_{vU2}^2+\omega_{vU2}^2\beta_1(\omega_{vU2})S(\omega_{vU2})}{\omega_{vU2}\sqrt{\omega_{vU1}^2-\omega_{vU2}^2}\beta_2(\omega_{vU2})}\right]+\pi N\right\},\quad S(\omega)=\beta_1(\omega)-\beta_2(\omega),\tag{14}$$

as well as to the nonreciprocity.

In the short-wavelength range $|k|d_1\gg 1$ the dispersion relation (12) admits an analytical solution in the form of MSSWs with a frequency

$$\omega_S=\frac{\omega_{vU1}}{\sqrt{1+\left(\frac{\pi[N+1/2]}{|k|d_1}\right)^2}}\tag{15a}$$

and a magnetic-potential penetration depth in the lower layer

$$L\approx\frac{\omega_s}{|k|\sqrt{\omega_S^2-\omega_{vU2}^2}}\ll d_1,\tag{15b}$$

which is much smaller than the thickness of the upper layer.

Thus the magnetostatic analog of Love elastic waves[25,26] in a ferromagnet with a ground state inclined from the easy axis exists only if the upper limit of the MSVW spectrum in the lower layer is smaller than the corresponding value for the upper layer. The form of the MSSW spectrum resembles the form of the MSVW spectrum of a ferromagnetic plate. The differences are confined to the low-frequency cutoff at the frequency of the upper limit of the MSVW spectrum of the lower layer and to the nonreciprocity. In the short-wavelength range the frequency of the waves investigated is slightly less than the upper limit of the MSVW spectrum of the upper layer, and the magnetic-potential penetration depth in the lower layer is much smaller than the thickness of the upper layer.

MSSW SPECTRUM IN A FERROMAGNETIC BILAYER WITH MAGNETIZATION VECTORS DIRECTED ALONG THE EASY AXIS

Let us consider the MSSW spectrum in the case of the realization of ground states with magnetization vectors directed along the easy axis (1b) in both layers. Real roots of the dispersion relation (12) can exist only in the frequency ranges $\omega_{vL1}<\omega<\omega_{vU1}$ and $\omega<\omega_{vL2}$ or $\omega>\omega_{vU2}$. If $\omega_{vU1}<\omega_{vL2}$, the MSSW spectrum (Fig. 3a) resembles the MSVW spectrum of a ferromagnetic plate. A similar picture is observed in the case of $\omega_{vU2}<\omega_{vL1}$. In these two situations the dispersion relation (12) in the long-wavelength region $|k|d_1\ll 1$ has an analytical solution in the form of MSSWs with a frequency

$$\omega_S=\omega_{vL1}+2\pi\gamma M_{01}\left[\frac{-\pi N+\sqrt{(\pi N)^2+4|k|d_1\rho}}{2\rho}\right]^2,\tag{16a}$$

where $\rho=1+\sqrt{[\omega_{vU2}^2-\omega_{vL1}^2]/[\omega_{vL2}^2-\omega_{vL1}^2]}$.

The magnetic-potential penetration depth in the lower layer

$$L\approx\frac{1}{|k|}\sqrt{\frac{\omega_{vL2}^2-\omega_{vL1}^2}{\omega_{vU2}^2-\omega_{vL1}^2}}>d_1$$

is greater than the thickness of the upper layer. The solution of the dispersion relation (12) for the short-wavelength range $|k|d_l\gg 1$ consists of MSSWs with a frequency

$$\omega_S=\omega_{vL1}\sqrt{1+\frac{4\pi\gamma M_{01}/\Omega_{01}}{1+\left[\frac{\pi(N+1/2)}{|k|d_1}\right]^2}}\tag{17a}$$

and a magnetic-potential penetration depth in the lower layer

$$L\approx\frac{1}{|k|}\sqrt{\frac{\omega_{vL2}^2-\omega_S^2}{\omega_{vU2}^2-\omega_S^2}}<d_1,\tag{17b}$$

which is smaller than the thickness of the upper layer. In the case of $\omega_{vL1}<\omega_{vL2}<\omega_{vU1}$ the MSSW spectrum has a high-frequency ‘‘cutoff’’ (Fig. 3) at the frequency $\omega=\omega_{vL2}$ and the wave vectors

$$|k_N|=\frac{1}{d_1}\sqrt{\frac{\omega_{vL2}^2-\omega_{vL1}^2}{\omega_{vU1}^2-\omega_{vL2}^2}}\times\left[\operatorname{arccot}\left\{-\sqrt{\frac{\omega_{vL2}^2-\omega_{vL1}^2}{\omega_{vU1}^2-\omega_{vL2}^2}}\right\}+\pi N\right].\tag{18}$$

Then for the long-wavelength range $|k|d_1\ll 1$ the dispersion relation has the solution (16a) with the magnetic-potential penetration depth in the lower layer (16b), which is greater than the thickness of the upper layer. If $\omega_{vU2}<\omega_{vU1}$, the MSSW spectrum (Fig. 3c) has a low-frequency ‘‘cutoff’’ at the frequency $\omega=\omega_{vU2}$ and the wave vectors

$$|k_N|=\frac{1}{d_1}\sqrt{\frac{\omega_{vU2}^2-\omega_{vL1}^2}{\omega_{vU1}-\omega_{vU2}}}\times\left[\operatorname{arccot}\left\{\sqrt{\frac{\omega_{vU1}^2-\omega_{vU2}^2}{\omega_{vU2}^2-\omega_{vL1}^2}}\right\}+\pi N\right].\tag{19}$$

In the short-wavelength range $|k|d_1\gg 1$ the dispersion relation (12) has a solution in the form of MSSWs with a frequency

$$\omega_S=\omega_{vU1}\left[1-\frac{\omega_{vU1}^2-\omega_{vL1}^2}{2\omega_{vU1}^2}\left\{\frac{\pi[N+1]}{|k|d_1}\right\}^2\right]\tag{20}$$

and a magnetic-potential penetration depth in the lower layer

$$L\approx\frac{1}{|k|}\sqrt{\frac{\omega_{vL2}^2-\omega_S^2}{\omega_{vU2}^2-\omega_S^2}}\ll d_1,\tag{21}$$

which is much smaller than the thickness of the upper layer.

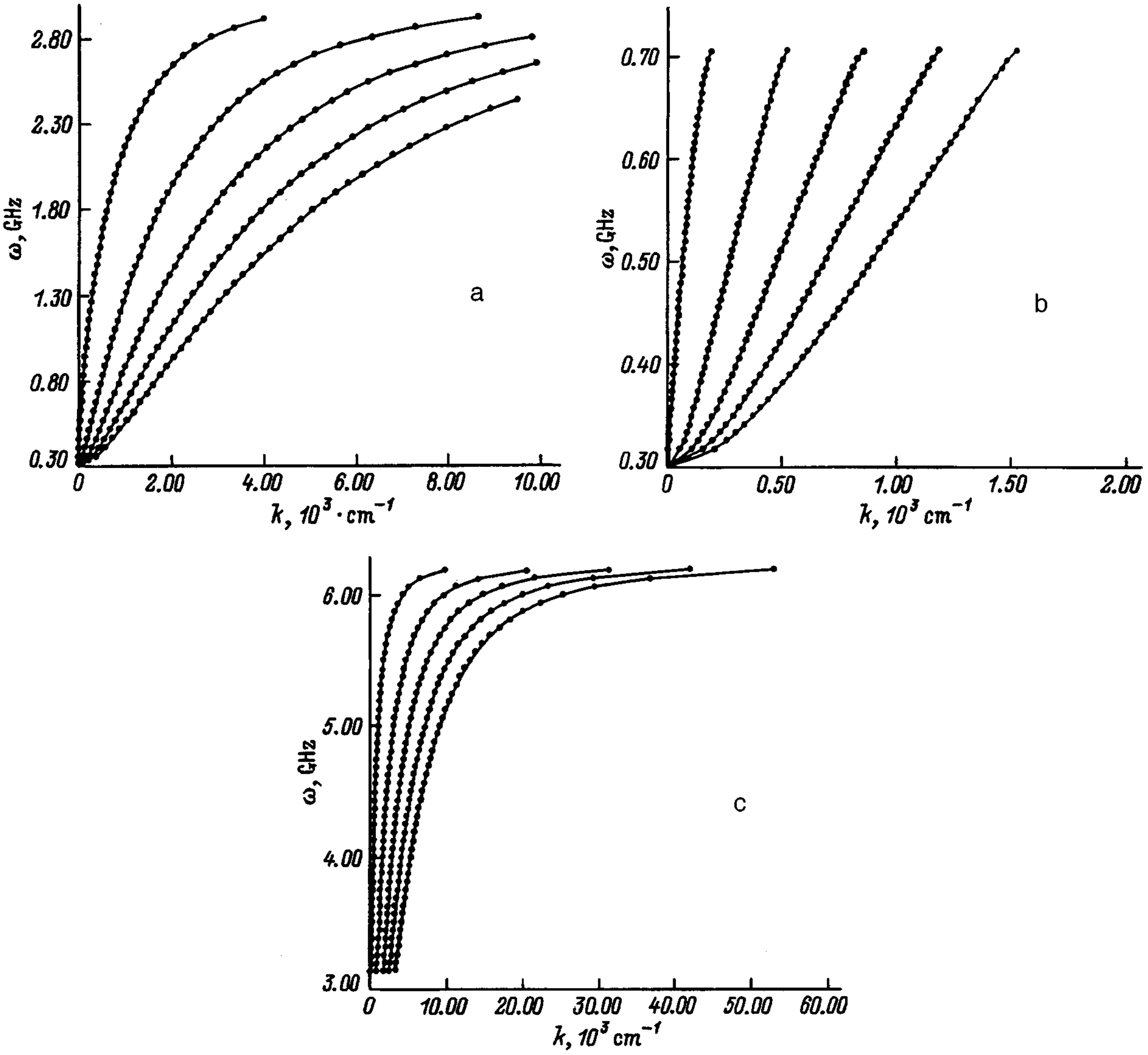

FIG. 3. Same as in Fig. 2 for the case of the realization of states with saturation magnetization vectors directed along the easy axis in the ferromagnetic layers: a — $\omega_{vU1}<\omega_{vL2}$ (similarly for $\omega_{vU2}<\omega_{vL1}$). Film parameters: $K_1=1.24\times10^4$ erg/cm^3, $K_2=3\times10^4$ erg/cm^3, $H_0=1600$ Oe, $M_{01}=M_{02}=M_0=140.06$ G, $d_1=2\times10^{-3}$ cm; b — $\omega_{vL1}<\omega_{vL2}<\omega_{vU1}$; only the value $K_2=1.4\times10^5$ erg/cm^3 differs from the parameters in Fig. 3a; c — $\omega_{vU2}<\omega_{vU1}$; film parameters: $K_1=1.6\times10^4$ erg/cm^3, $K_2=1.24\times10^4$ erg/cm^3, $H_0=1600$ Oe, $M_{01}=M_{02}=M_0=140.06$ G, $d_1=2\times10^{-3}$ cm.

Thus we have demonstrated that a magnetostatic analog of Love elastic waves[25,26] can exist in a bilayer ferromagnet magnetized along the easy axis. We have established that in cases where the frequency regions in which MSVWs exist in the upper and lower layers do not intersect, the MSSW spectrum resembles the MSVW spectrum of the upper layer. Then the frequencies of the MSSWs in the long-wavelength region are somewhat greater than the frequency of the lower limit of the MSVW spectrum, and the localization depth in the lower layer exceeds the thickness of the upper layer. In the short-wavelength limit the frequencies of the MSSWs are somewhat less than the frequency of the upper limit of the MSVW spectrum, and the localization depth in the lower layer is less than the thickness of the upper layer. If the frequency of the lower limit of the MSVW spectrum of the lower layer falls in the frequency range for the existence of MSVWs in the upper layer, the spectrum has a high-frequency ''cutoff.'' There is a low-frequency ''cutoff'' in the case where the frequency of the upper limit of the MSVW spectrum of the lower layer falls in the frequency range for the existence of MSVWs in the upper layer.

CONCLUSIONS

Thus, the possibility of the existence of MSSWs in a bilayer film due to the difference between the magnetic anisotropy constants has been demonstrated. The magnetic potential in the waves predicted oscillates in the upper layer and decays exponentially in the lower layer. Therefore, they can be regarded as a magnetostatic analog of Love surface elastic waves.[25,26]

[1] T. J. I'A. Bromwich, Proc. London Math. Soc. **30**, 98 (1898).
[2] A. E. H. Love, *Some Problems of Geodynamics*, Cambridge University Press, Cambridge (1911), 180 pp.
[3] H. Matthews and H. van de Vaart, Appl. Phys. Lett. **14**(5), 171 (1969).

[4] P. Tornois and C. Lardat, in *Reports to the 6th International Congress on Acoustics*, Tokyo, (1968), Vol. 6, pp. 37–40.

[5] Special Issue on Surface Acoustic Waves, Proc. IEEE **64**(5), 324, 1976 [Mir, Moscow (1976)].

[6] S. S. Karinskiĭ, *Devices for Processing Ultrasonic Surface Wave Signals* [in Russian], Sov. Radio, Moscow (1975).

[7] G. M. Vapné, in *Microwave Devices Using Magnetostatic Waves. Reviews on Electronics, Ser. 1* [in Russian], TsNII ''Élektronika,'' Moscow (1984), No. 8, p. 1060.

[8] V. A. Nikitov and S. A. Nikitov, Zarubezhn. Radioélektron. No. 2, 41 (1981).

[9] L. R. Adkins and H. L. Glass, Electron. Lett. **16**, 590 (1980).

[10] N. S. Chang and Y. Matsuo, Proc. IEEE **66**, 1577 (1978).

[11] H. Sasaki and N. Mikoshiba, J. Appl. Phys. **52**, 2546 (1981).

[12] R. L. Carter, J. M. Owens, C. V. Smith, and K. W. Reed, J. Appl. Phys. **53**, 2633 (1982).

[13] A. V. Vashkovskiĭ, V. S. Stal'makhov, and Yu. P. Sharaevskiĭ, *Magnetostatic Waves in Microwave Electronics* [in Russian], Izd. Saratovskogo Universiteta, Saratov (1993), 312 pp.

[14] F. Wolfram, J. Appl. Phys. **41**, 4748 (1970).

[15] A. V. Vashkovskiĭ and A. V. Stal'makhov, Radiotekh. Elektron. **29**, 901 (1984).

[16] Yu. I. Bespyatykh and V. I. Zubkov, Zh. Tekh. Fiz. **45**, 2386 (1975) [Sov. Phys. Tech. Phys. **20**, 1485 (1975)].

[17] I. A. Gilinskiĭ, Fiz. Met. Metalloved. **55**, 455 (1983).

[18] A. V. Vashkovskiĭ and A. A. Stal'makhov, Radiotekh. Elektron. **29**, 2409 (1984).

[19] V. I. Zubkov and V. A. Epanechnikov, Pis'ma Zh. Tekh. Fiz. **11**(23), 1419 (1985) [Sov. Tech. Phys. Lett. **11**(12), 585 (1985)].

[20] P. Hartemann and D. Fontaine, IEEE Trans. Magn. MAG **-18**, 1595 (1982).

[21] N. I. Lyashchenko and V. M. Talalaevskiĭ, Ukr. Fiz. Zh. **31**, 1716 (1986).

[22] Yu. V. Gulyaev, I. A. Ignat'ev, A. F. Popkov, and V. M. Shabunin, in *Abstracts of the 10th All-Union Seminar ''New Magnetic Materials for Microelectronics''* [in Russian], Riga, (1986), Part 1, pp. 176–177.

[23] I. G. Kudryashkin, D. G. Krutogin, E. A. Ladygin *et al.*, Zh. Tekh. Fiz. **59**(3), 70 (1989) [Sov. Phys. Tech. Phys. **34**, 294 (1989)].

[24] A. I. Akhiezer, V. G. Bar'yakhtar, and S. V. Peletminskii, *Spin Waves*, North-Holland, Amsterdam (1968), 368 pp.

[25] I. A. Viktorov, *Acoustic Surface Waves in Solids* [in Russian], Nauka, Moscow (1981), 288 pp.

[26] E. Dieulesaint and D. Royer, *Elastic Waves in Solids*, (Wiley, New York, 1981; Nauka, Moscow 1982), 424 pp.

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Theory of gyro devices with thin electron beams (large-orbit gyrotrons)

V. L. Bratman, Yu. K. Kalynov, and A. É. Fedotov

Institute of Applied Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia

(Submitted June 3, 1997)
Fiz. Tekh. Poluprovodn. **68**, 91–98 (October 1998)

It is shown for a precisely aligned electron beam of small thickness in a large-orbit gyrotron that high cyclotron harmonics (up to $s\leqslant 5$) can be selectively excited with an electronic efficiency that is acceptable for some applications (1–10%). When the quality of the electron beam is high, the selective properties are maintained even for modes with high radial indices, raising hopes that this method can be used to obtain coherent radiation not only in the millimeter but also in the entire submillimeter wavelength range. A method for taking into account the space-charge effects is developed, and it is shown that the corresponding effects can be important at relatively small values of the electron pitch factor. The results of the analysis and preliminary experiments reveal the possibility of creating high-power compact sources of submillimeter radiation for the spectroscopy of various media, the diagnostics of dense plasmas, and some other applications.
[S1063-7842(98)01610-9]

INTRODUCTION

Owing to the selective mechanism of the interaction of electrons with the radio-frequency (rf) field in modern high-power gyrotrons, spatially developed modes with very high azimuthal and radial indices can be employed as working modes.[1,2] Such modes can be excited by hollow electron beams of large diameter (close to the diameter of the internal caustic of the mode), which is many times greater than the Larmor diameter, and consist of a continuous set of uniformly filled Larmor tubes that are uniformly distributed over the azimuth. Along with this main topic, several studies in the last few years have explored the possibility of employing, in various modifications of the cyclotron resonance maser, thin beams (with a thickness of one Larmor diameter) in which the particles circulate around a common axis coinciding with the axis of an axially symmetric electrodynamic system (one Larmor tube or even one Larmor helix coaxial to a circular waveguide). Because of the symmetry of such a system, only modes whose azimuthal index coincides with the number of the resonant cyclotron harmonic can be excited in it. Accordingly, there is additional thinning of the spectrum, which is especially important in attempting to excite high cyclotron harmonics. This situation is realized in so-called large-orbit gyrotrons,[3–8] where, in fact, as in an ordinary gyrotron, the ‘‘orbit’’ of one electron coincides with the Larmor orbit and is large only in comparison to the relatively small diameter of the electrodynamic system. In addition, some cyclotron autoresonance masers have a similar geometry.[9,10] Helical electron beams, which are similar, but gyrate as a whole, are also used in so-called harmonic converters,[11,12] which are essentially a type of gyro multiplier with a novel system for the simultaneous formation and modulation of the beam.

The main mechanisms for the interaction of thin electron beams with a rf field do not differ from those which operate in ordinary gyrotrons and can be described within the general theory.[13–19] At the same time, the spectrum of excitable modes and the effects of the rf space charge, which are especially important for interactions at high cyclotron harmonics, have definite specific features. These problems, which are important for creating high-power short-wavelength devices, have not yet found solutions in published studies. In this paper we construct a general theory for a relativistic large-orbit gyrotron, which will enable us to analyze, in particular, the two problems indicated.

1. SELECTIVE EXCITATION OF CYCLOTRON HARMONICS IN A CIRCULAR WAVEGUIDE

A gyrotron with a thin electron beam circulating around the axis of an axially symmetric waveguide has enhanced selective properties in comparison to an ordinary gyrotron, in which the axes of the electron trajectories do not coincide with the axis of the waveguide. In fact, when a particle orbits the axis of a circular waveguide exactly, the field of the azimuthally traveling $\mathrm{TE}_{n,p}$ mode in its orbit is a single gyrating multipole ($2n$-pole, where n is the azimuthal mode index of the wave; Fig. 1), rather than a set of field multipoles, as in an ordinary gyrotron.[15] This single multipole, which gyrates with a frequency ω/n, where ω is the frequency of the mode, can interact resonantly with an electron orbiting with the cyclotron frequency ω_H at only one cyclotron harmonic, whose number s coincides with the azimuthal index of the mode ($s=n$) when the synchronism condition is satisfied:

$$\omega\approx s\omega_H . \qquad (1)$$

If the synchronism condition for another harmonic with the number n' is simultaneously satisfied in (1), a resonant interaction of an electron can take place at this harmonic only with a wave having the azimuthal index n'. This prop-

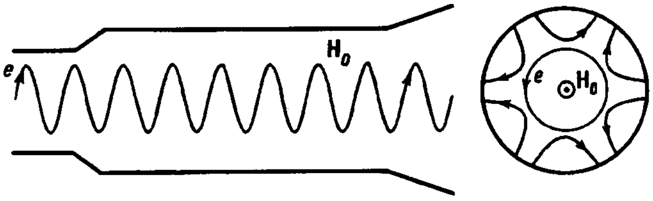

FIG. 1. Working space of a large-orbit gyrotron.

erty of cyclotron resonance masers was clear already from the early theoretical studies,[13–16] but its systematic use in experiment began only after the publication of Ref. 3, where the special term ''large-orbit gyrotron'' was proposed for this type of gyrotron.

The enhanced selectivity is clear from the following expression, which was derived in Ref. 13 and is well known in the theory of gyrotrons, for the structure factor appearing in the impedance of the interaction of a hollow beam of electrons having leading centers of the same radius R_b with the field of a cylindrical waveguide:

$$G=\frac{J_{n-s}^2(k_\perp R_b)}{(1-n^2/\nu_{n,p}^2)J_n^2(\nu_{n,p})}. \tag{2}$$

Here J_n is a Bessel function, $\nu_{n,p}=k_\perp a$ is the pth positive root of the derivative of the Bessel function $J_n'(\nu)$, and a is the radius of the waveguide. When $R_b=0$ and the argument of the Bessel function in the numerator in (2) thus vanishes, it is easy to see that G is nonzero only if $n=s$.

In order to compose the initial representation of the possibility of the selective excitation of individual modes, we take into account that the generation frequency is close to the corresponding critical frequency of the waveguide:

$$\omega\approx\nu_{n,p}c/a. \tag{3}$$

Hence, for the resonant magnetic field we obtain

$$H[\mathrm{kOe}]=1.7\frac{\gamma}{a[\mathrm{cm}]}\frac{\nu_{n,p}}{s}, \tag{4}$$

where γ is the relativistic factor of the electrons.

The coincidence between the harmonic number s and the azimuthal mode index n obviously greatly thins the spectrum of the resonant values (4) and the modes corresponding to them which an electron beam is capable of exciting when it is ideally injected into the waveguide. Owing to the nonequidistant arrangement of the roots $\nu_{n,p}$, the resonant values of the magnetic field for the first harmonics have a fairly sparse distribution (Fig. 2). For example, for modes with the radial index 1 the values corresponding to the fourth, fifth, and sixth harmonics differ from one another by 4 and 2.5%.

The corresponding magnetic-field resonance band can be estimated, as usual, from the condition that the values of the drift angle $(\omega-s\omega_H)T$ (the phase shift of the cyclotron gyration relative to the wave occurring during the time of flight T of electrons through the interaction space) at the edges of the band differ from one another by a value of the order of 2π:

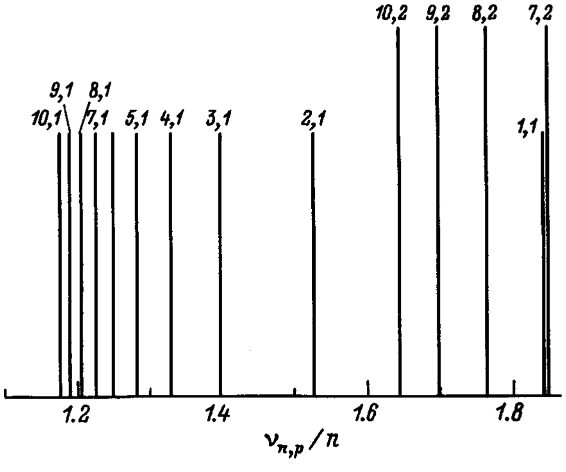

FIG. 2. Dimensionless values of the resonant magnetic field for harmonics $n=1-10$ corresponding to modes with the radial index $p=1$ (the analogous modes with the radial index $p=2$ are also shown).

$$\frac{\Delta H}{H}=\frac{1}{sN}, \tag{5}$$

where $N=\omega_H T/2\pi$ is the number of gyrations of an electron in the interaction space.

According to this estimate, the main zone for generation of the fifth harmonic is isolated from the neighboring zones already when the number of gyrations $N\geqslant 5$.

The excitation efficiency of a particular mode is determined not only by the structure factor G, but also by the so-called electron-wave coupling coefficient, which is proportional to the power of the individual emission of one particle and varies with the number of the harmonic approximately as $\beta_\perp^{2s}$, where $\beta_\perp=v_\perp/c$ is the normalized orbital velocity of the particles. Hence it is clear that when the orbital velocities of the particles are increased to relativistic values, the coupling with high harmonics becomes significantly more effective and their excitation is markedly facilitated.

2. EQUATIONS OF A RELATIVISTIC GYROTRON

In the working space of a gyrotron electrons travel along helical trajectories in a homogeneous (or weakly inhomogeneous) magnetic field $\mathbf{H}_0=H_0\mathbf{z}_0$ and excite rf oscillations in an electrodynamic system having the form of a portion of a weakly inhomogeneous cylindrical waveguide (Fig. 1). We shall assume that all particles entering the space for interaction with the rf field have identical orbital ($v_{\perp 0}$) and translational ($v_{\parallel 0}$) velocities and are uniformly distributed among the cyclotron gyration phases. For simplicity, we also assume that there is no spread in the positions of the leading centers and that they all lie on the waveguide axis. Thus, at the entrance to the interaction space all the particles are located on the surface of a single Larmor tube.

We assume that a stationary generation regime operates and that the electrons interact with one (working) transverse-electric mode ($\mathrm{TE}_{n,p}$)

$$\mathbf{E}=\mathrm{Re}\{E_0 f(z)\mathbf{E}_{n,p}\exp(i\omega t)\},$$

$$\mathbf{H}=\mathrm{Re}\{E_0 f(z)\mathbf{H}_{n,p}\exp(i\omega t)\}, \tag{6}$$

whose frequency ω, on the one hand, is close to the corresponding critical frequency and, on the other hand, satisfies the cyclotron-resonance condition (1) at the nth harmonic. In (6) the function $f(z)$ describes the longitudinal structure, and the vector functions $\mathbf{E}_{n,p}$ and $\mathbf{H}_{n,p}$, which are expressed in terms of a membrane function in a certain manner, describe the transverse structure of the variable field at a frequency ω in an empty waveguide. The cavity is assumed to be high-Q, and the longitudinal structure of the field is assumed to be fixed.

The equations of motion averaged over the cyclotron oscillations for a relativistic electron in a magnetostatic field and the wave field (6) under resonance conditions at the nth harmonic have the form[18]

$$\frac{du}{d\zeta}=-\frac{\partial}{\partial\vartheta}\mathrm{Re}\{Ff(\zeta)p_\perp J_n'(np_\perp)\exp(i\vartheta)\}, \tag{7}$$

$$\frac{d\vartheta}{d\zeta}=\Delta-u+\frac{\partial}{\partial u}\mathrm{Re}\{Ff(\zeta)p_\perp J_n'(np_\perp)\exp(i\vartheta)\}. \tag{8}$$

Here $u=2(1-\gamma/\gamma_0)/\beta_{\perp 0}^2$ is the energy variable, which is equal to the normalized current change in the total energy of the particle; $\vartheta=(\omega t-n\varphi)$ is its phase relative to the rf field; φ is the angle characterizing the position of the electron in the cyclotron orbit; $\zeta=\beta_{\perp 0}^2\omega z/2\beta_{\parallel 0}c$, $F=4eE_0/(m\gamma_0\times c\omega\beta_{\perp 0}^4)$, and $p_\perp=\beta_{\perp 0}\gamma/\gamma_0$ are the normalized longitudinal coordinate, amplitude of the cavity field, and transverse momentum of the electron; $\Delta=2(1-n\omega_{H_0}/\omega)/\beta_{\perp 0}^2$ is the initial resonance mismatch between the electrons and the wave. The normalized transverse momentum is related to the energy by the expression $p_\perp^2\approx\beta_{\perp 0}^2(1-u)$.

For a generator excited by a preliminarily unmodulated electron beam, the initial conditions for the equations of motion at the entrance to the interaction space (at $\zeta=0$) have the form

$$u(0)=0,\quad \vartheta(0)=\vartheta_0,\quad 0\leqslant\vartheta_0<2\pi. \tag{9}$$

In the steady-state regime a gyrotron obeys the relation

$$\frac{\omega W}{Q}=\eta IU, \tag{10}$$

which describes the balance between the power extracted from the cavity by the wave and the power imparted to the rf field by the electron beam. Here W is the energy stored in the cavity;

$$Q=\frac{Q_{\mathrm{ohm}}Q_d}{Q_{\mathrm{ohm}}+Q_d} \tag{11}$$

is its loaded Q factor, which is determined by the diffractional extraction of radiation (the diffractional quality factor Q_d) and the Ohmic losses in the cavity walls (the Ohmic quality factor Q_{ohm}); η is the electronic efficiency of the gyrotron; and I and U are the current and voltage of the electron beam. The electronic efficiency can be represented in the form of a product of two factors

$$\eta=\eta_{\mathrm{s.p}}\eta_\perp, \tag{12}$$

where

$$\eta_{\mathrm{s.p}}=\frac{\beta_{\perp 0}^2}{2(1-\gamma_0^{-1})}\quad\text{and}\quad \eta_\perp=\langle u(\zeta=\mu)\rangle_{\Theta_0}$$

are the so-called single-particle and transverse efficiencies, respectively, and $\mu=\beta_{\perp 0}^2\omega L/2\beta_{\parallel 0}c$ is the dimensionless length of the interaction space.

For interactions at high harmonics in a homogeneous magnetostatic field, it should be stressed that relativistic electrons are capable of imparting only a relatively small part of their initial energy to the monochromatic wave due to the strong dependence of the cyclotron frequency on energy. In fact, an electron goes out of resonance with the wave when the dynamic change in its drift angle caused by the energy loss is of the order of 2π. Hence for the change in the energy of the electron we obtain the estimate

$$\frac{|\delta\gamma|}{\gamma_0}\leqslant\frac{1}{nN}. \tag{13}$$

Utilizing the small value of the energy change, we can[19] significantly simplify the equations of motion and reduce them to the standard asymptotic form for all devices with inertial bunching of particles.[20] To this end the change in the energy of the particles on the right-hand side of Eq. (7) should be neglected, the right-hand side of the equation for the phase (8) should be linearized with respect to the small energy change, and the term in it responsible for the force bunching of particles, which is proportional to the field amplitude, should be discarded. As a result, the equations of motion can be written in new variables in the following form:[19]

$$\frac{d\hat{u}}{d\hat{\zeta}}=\mathrm{Im}\{f(\hat{\zeta})\exp(i\vartheta)\}, \tag{14}$$

$$\frac{d\vartheta}{d\hat{\zeta}}=\hat{\delta}-\hat{u}. \tag{15}$$

Here we have used the notation $\hat{\zeta}=\zeta\sqrt{\hat{F}}$, $\hat{\delta}=\Delta/\sqrt{\hat{F}}$, $\hat{u}=u/\sqrt{\hat{F}}$, and $\hat{F}=\beta_{\perp 0}J_n'(n\beta_{\perp 0})F$. The gyrotron efficiency can now be expressed in the following manner:

$$\eta=\frac{\hat{\mu}\hat{\eta}_\perp}{(1-\gamma_0^{-1})(\omega L/v_{\parallel 0})},\quad \hat{\eta}_\perp=\langle\hat{u}(\hat{\mu})\rangle_{\vartheta_0}, \tag{16}$$

where $\hat{\mu}=\beta_{\perp 0}^2\omega L\sqrt{\hat{F}}/2\beta_{\parallel 0}c$.

The amplitude of the rf field in the cavity (and thus the dimensionless length $\hat{\mu}$) can be adjusted by varying the current so that the maximum of the function $\hat{\mu}\hat{\eta}(\hat{\mu})$ optimized with respect to the frequency mismatch, would be achieved. Then, irrespective of the particle velocities, the efficiency is inversely proportional to the wave frequency. If the number of the harmonic and, accordingly, the azimuthal mode index n are increased, while the radial index p of the working mode is fixed, the efficiency decreases with the harmonic number as $1/n$ in accordance with the estimate (13).

3. STATIONARY GENERATION REGIME

For simplicity in this section we shall assume that the Q factor of the gyrotron cavity is fairly high, and we shall accordingly neglect the space-charge effects (the next section is devoted to a treatment of the latter). Solving Eqs. (7) and (8) in the small-signal approximation and using the simplifying sinusoidal approximation

$$f(\zeta)=\sin\pi\zeta/\mu \tag{17}$$

for the longitudinal structure, we obtain[18] the following expression for the start-up current of a relativistic gyrotron:

$$I_{\mathrm{st}}=\frac{mc^3}{e}\gamma_0\beta_{\|0}^2\beta_{\perp 0}\frac{na^2}{2\lambda L}\frac{1}{QG\chi}. \tag{18}$$

In this expression $\lambda=2\pi c/\omega$ is the wavelength of the radiation generated;

$$\chi=-n(1-\beta_{\perp 0}^2)J_n'(n\beta_{\perp 0})J_n(n\beta_{\perp 0})\phi$$
$$-\beta_{\perp 0}\mu J_n'^2(n\beta_{\perp 0})\phi' \tag{19}$$

is the susceptibility imparted to the cavity field by the electron beam; the function

$$\phi=4\pi^2\frac{1+\cos\Theta}{(\Theta^2-\pi^2)^2} \tag{20}$$

characterizes the spatial spectrum of the rf force acting on the electrons (in the present case the spectrum of one arc of the sinusoid); and $\Theta=\Delta\mu$ is the drift angle, which describes the kinematic displacement of the electrons relative to the wave.

For the asymptotic regime of large interaction distances and small changes in the particle energies considered at the end of the preceding section, where only the inertial bunching of the particles caused by nonisochronism of their cyclotron gyration is significant, only the second term in the expression for the susceptibility, which is proportional to the dimensionless length μ, is significant:

$$\chi=-\beta_{\perp 0}\mu J_n'^2(n\beta_{\perp 0})\phi. \tag{21}$$

The expression (18) for the start-up current permits a more detailed analysis than in Sec. 1 of the selective properties of a gyrotron with a thin electron beam for different regimes, including the excitation of high harmonics, and estimation of the danger of the excitation of parasitic modes for them. In relatively long systems ($\mu\gg 1$), where a regime of dominant inertial bunching of the particles is realized, the dependence of the efficiency on beam current and on the adjustment of the magnetostatic field corresponds to the soft self-excitation of auto-oscillations of individual modes. Under the conditions of overlap of the resonance curves and interaction of the working mode with competing modes at other cyclotron harmonics, a situation of ‘‘strong’’ mode coupling should be realized[21] because of the identical longitudinal structure of these modes. According to Ref. 21, such conditions provide for the establishment of the stable generation of a single mode, which has a small start-up current and is previously excited during application of the working voltage.

It is clear from (18) that the decrease in the susceptibility χ with increasing harmonic number and, accordingly, increasing frequency can be compensated at least partially by the increase in the cavity Q factor with increasing frequency of the mode. In fact, for efficient mode selection with respect to the longitudinal index and the efficient extraction of radiation from the cavity, the loaded Q factor of the cavity must be of the order of the diffractional Q factor (i.e., the Ohmic losses must not be excessively large). At the same time, the diffractional Q factor

$$Q_d=4\pi\left(\frac{L}{\lambda}\right)^2\frac{1}{1-R}, \tag{22}$$

where R, which is the reflectivity of the exit end of the cavity for the wave, increases quadratically with increasing frequency of the working mode. The Ohmic Q factor

$$Q_{\mathrm{ohm}}=\frac{a}{d}(1-m^2/\nu_{m,p}^2) \tag{23}$$

is determined to within a factor of 1/2 by the ratio of the effective volume $\pi a^2(1-m^2/\nu_{m,p}^2)L$ occupied by the rf field within the cavity (between the wall and the caustic surface) to the volume of the skin layer $2\pi aLd$, where d is the skin depth. For a fixed radial mode index p, as the harmonic number, which is equal to the azimuthal mode index, increases, the effective volume occupied by the field decreases approximately in proportion to the volume of the skin layer, causing a weak dependence of the Ohmic Q factor on the azimuthal mode index. As the radial mode index increases, the effective volume of the field rises, and the volume of the skin layer falls, causing a rapid rise in the Ohmic Q factor (approximately proportional to p). Estimates show (see below) that, owing to the large orbital velocity of the particles, the start-up currents of a relativistic gyrotron are very large even for small cavity lengths and reflectivities of the exit end. Under such conditions the Ohmic Q factor is, as a rule, significantly greater than the diffractional Q factor over the entire range of millimeter waves.

As we have already stated, the structure factor G for an electron beam strictly encircling the cavity axis ($R_b=0$) is nonzero only for modes with an azimuthal index equal to the number of the resonant cyclotron harmonic. At the same time, real electron beams have a definite thickness, which leads to a spread of leading centers for the electron trajectories. Assuming that the leading centers evenly fill a circular area of radius b and using the known integral of the square of a Bessel function, we can easily obtain the following expression for the structure factor with consideration of the beam thickness:

$$G=\frac{J_0^2(kb)+J_1^2(kb)}{J_n^2(\nu_{n,p})(1-n^2/\nu_{n,p}^2)}. \tag{24}$$

It follows from this expression that even a fairly considerable beam thickness, for which $kb=1$, leads to only a 22% decrease in the structure factor. At the same time, inexact beam alignment, under which the radius of a leading center satisfies the condition $kR_b=1$, produces an almost doubled effect. Both of these factors have an even greater effect on

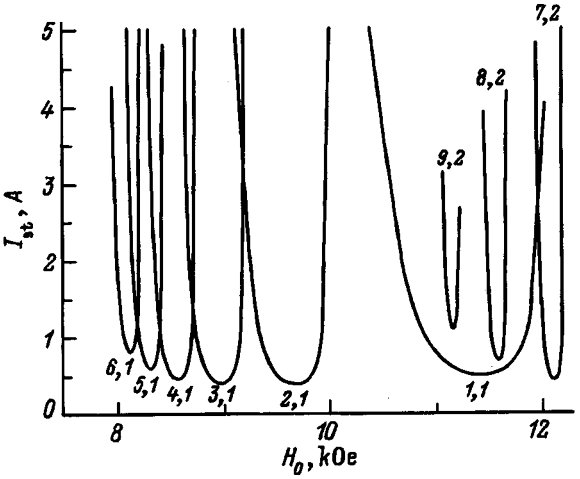

FIG. 3. Principal self-excitation zones for higher cyclotron harmonics in a large-orbit gyrotron (the effects of the rf space charge are not taken into account).

the danger of exciting parasitic modes, primarily those whose azimuthal index differs from the harmonic number by unity. As an illustration, let us consider the example of a moderately relativistic electron beam with an energy of 300 keV and a pitch factor $g=\beta_\perp/\beta_\|=1.5$, where the increase in the cavity Q factor for high harmonics already partially compensates the decrease in the susceptibility χ. For a fixed cavity with a diameter of the cylindrical part equal to 8 mm, a length of 40 mm, and tapering and exit expansion angles equal to 1.5 and 3° (such a cavity was used in Ref. 22 to generate the fifth harmonic with a wavelength of about 4 mm) the start-up currents for the first to fifth harmonics are close to one another (Fig. 3). The generation zones for $s=1-4$ are adequately separated, but the zones for the fifth and higher harmonics overlap fairly strongly. It is noteworthy that the generation zone of the fundamental $\mathrm{TE}_{1,1}$ mode covers the zones for the high harmonics with the radial index 2, which correspond to essentially the same start-up currents (a similar effect was noted in Ref. 23 for so-called gyroharmonic converters).

4. INFLUENCE OF THE RADIO-FREQUENCY SPACE CHARGE ON THE START-UP CONDITIONS

For interaction spaces with relatively short lengths and cavities with low diffractional Q factors, where the working currents of the gyrotrons are fairly high, the effects of the rf space charge can play a significant role (see, for example, Refs. 24–29). This circumstance is especially important for gyrotrons with thin electron beams, where the entire current is concentrated within one Larmor tube or one Larmor helix. For simplicity, we shall assume that in the undisturbed state the electron beam is a uniformly filled Larmor tube with a surface charge density $I/2\pi r_H v_\|$. Allowance should be made for the specific features of a gyrotron, in which it propagates almost transversely to the electron beam and creates disturbances in it that vary slowly with the longitudinal coordinate. Then, for calculating the space-charge field it is convenient to divide the Larmor tube into gyrating filaments consisting of electrons with the same initial phase relative to the rf field (similarly, in gyrotrons with polyhelical electron beams, elementary sources in the form of oscillating charged planes are used to calculate the space-charge field[25–27,29]). If the energies and phases of the particles vary weakly over a length of the order of the Larmor radius, these filaments can be considered locally straight and parallel to the external magnetic field. The quasistatic electric field of a straight gyrating filament containing electrons from the interval of initial phases $[\tilde{\vartheta}_0, \tilde{\vartheta}_0+d\tilde{\vartheta}_0]$ equals

$$E_+=-\frac{I}{\tilde{v}_\|}\frac{d\tilde{\vartheta}_0}{2\pi}\frac{r_+-\tilde{r}_+}{|r_+-\tilde{r}_+|^2}, \tag{25}$$

where the following complex combinations of transverse vector components are used: $E_+=E_x+iE_y$, $r_+=x+iy$, and $\tilde{r}_+=\tilde{x}+i\tilde{y}$. Here r_+ corresponds to the coordinates of the observation point where the subject filament with the initial phase ϑ_0 is located, and $\tilde{r}_+$ is the coordinate of the filament creating the field.

Associating r_+ and $\tilde{r}_+$ with the corresponding current phases of the electrons in the expressions $r_+=r_H\exp[i(\omega t-\vartheta)/n]$ and $\tilde{r}_+=\tilde{r}_H\exp[i(\omega t-\tilde{\vartheta})/n]$ and summing over all the filaments creating the field, we obtain an additional quasistatic force acting on an electron of the subject filament. In the case of dominant inertial bunching of the particles, where the variation of the gyroradius of the electrons can be neglected, the expression for the additional force has the form

$$F_\varphi=-\frac{S}{2\pi}\int_0^{2\pi}\cot\frac{\tilde{\vartheta}-\vartheta}{2n}\,d\tilde{\vartheta}_0, \tag{26}$$

where

$$S=\frac{4eI_0}{mc^3}\frac{1}{n\gamma_0\beta_{\|0}\beta_{\perp 0}^4}$$

is the space-charge parameter.

For the same currents and electron velocities, this parameter is $2\pi R_0/r_H$ times greater than the analogous quantity for a gyrotron with a conventional beam having the leading-center radius R_b, where the current is distributed over a large area. The force F_φ should be added to the right-hand side of the equation for the particle energy

$$\frac{du}{d\zeta}=\mathrm{Im}\left\{\hat{F}f(\zeta)\exp(i\vartheta)-S\left\langle\cot\frac{\vartheta-\tilde{\vartheta}}{2}\right\rangle_{\tilde{\vartheta}}\right\}, \tag{27}$$

$$\frac{d\vartheta}{d\zeta}=\Delta-u. \tag{28}$$

The start-up current of a gyrotron can be found with consideration of the space-charge effects from Eqs. (27) and (28) and the power-balance equation (10) by the method of successive approximations with respect to the amplitude of the rf field $\hat{F}$. Then the linear solution of Eqs. (27) and (28), which takes into account the space-charge effects, can be found by varying an arbitrary constant in analogy to Ref. 28. For example, for the linear correction to the electron phase we obtain $\vartheta^{(1)}=\vartheta_1+\vartheta_2$, where

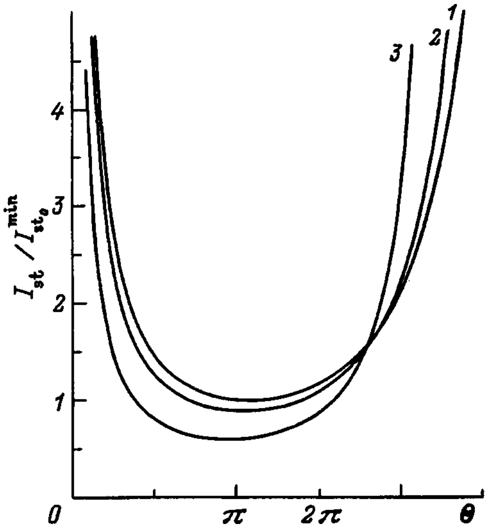

FIG. 4. Influence of the rf space charge on the start-up current of a large-orbit gyrotron. ν: *1* — 0, *2* — 2.0, *3* — 4.0; $I_{st_0}^{min}$ is the minimum value of the start-up current in the absence of space charge.

$$\vartheta_{1,2}=\pm\frac{F}{4q}\left[\frac{\exp[i(\Delta-h)\zeta]}{\Delta\mp q-h}-\frac{\exp[i(\Delta+h)\zeta]}{\Delta\mp q+h}-\frac{2h\exp(iq\zeta)}{(\Delta\mp q)^2-h^2}\right], \tag{29}$$

$h=\pi/\mu$, and $q=\sqrt{-Q}$ is a purely imaginary quantity.

Substituting (29) into Eq. (27) for the energy, we find the efficiency in the small-signal approximation, and then using the balance equation (10), we find the start-up current of the gyrotron. The expression found in this manner for the start-up current differs from the analogous expression (18) with neglect of the space-charge effects in the form of the spectral function appearing in the expression for the susceptibility of the electron flux

$$\chi=-\beta_{\perp 0}\mu J_n'^2(n\beta_{\perp 0})\psi,$$

$$\psi=\frac{2\pi^2}{i\nu}\left(\frac{1-\cos(\theta-\pi+i\nu)}{((\theta+i\nu)^2-\pi^2)^2}-\frac{1-\cos(\theta-\pi-i\nu)}{((\theta-i\nu)^2-\pi^2)^2}\right). \tag{30}$$

Here $i\nu=q\mu$ is the addition to the drift angle associated with the space charge. When the space charge is vanishingly small $(S=0)$, the function ψ tends to the derivative ϕ'.

The expression for ψ coincides in form with the expression for the spectral function in an orotron,[30] but in contrast to the latter, in a gyrotron the change in the drift angle under the space-charge effects is a purely imaginary quantity. This is because the instability of the negative mass renders each eigenwave in the electron beam of gyrotron a growing and a decaying wave (a fast and a slow wave in an orotron).

The space-charge effects become significant when ν takes values appreciably exceeding unity. For example, when $\nu=4$, the minimum start-up current decreases to about half in comparison to its value in the absence of space charge (Fig. 4). Applying the method of successive approximations, we use the values of the start-up current (18) found in the absence of space charge to calculate ν. Then

$$\nu^2=\frac{1}{2nG\beta_{\perp 0}^2 J_n'^2(n\beta_{\perp 0})Q\psi}. \tag{31}$$

For the example considered in Sec. 3 we obtain $\nu\approx 0.4$, and the influence of the space charge is small (the start-up current at the fifth harmonic decreases by only a few percent when the space charge is taken into account). At the same time, when the orbital velocity of the electrons and the length of the interaction space are small and the Q factor is close to the minimum diffractional Q factor, the space charge can have a much stronger influence at the high harmonics. For example, when the orbital velocity of the particles $\beta_\perp=0.5$ $(g=1)$ and $L=5\lambda$, we obtain $\nu\approx 4$ for the fifth harmonic. It can be seen from (31) that the influence of the space charge increases with the harmonic number, since the same absolute phase shift of the electrons corresponds to stronger bunching for higher harmonics, and, in addition, higher harmonics have larger currents. It should, of course, be borne in mind that if the electrons have a positional or, even more so, a velocity spread, the influence of the space charge is reduced because of the phase mixing of the particles,[31] which has a stronger effect at the high harmonics.

CONCLUSIONS

A gyrotron with a moderately relativistic, thin electron beam encircling the axis of an axially symmetric electrodynamic system provides for the selective excitation of high cyclotron harmonics even when conventional cavities are used and no additional methods for discriminating the parasitic modes are employed. In this way we can hope to obtain high-power coherent radiation not only in the millimeter wavelength range, but also in the entire submillimeter range, although the difficulties in forming and aligning the electron beam increase dramatically at short wavelengths. According to calculations, the electronic efficiency of a large-orbit gyrotron at the harmonics with numbers $s\lesssim 5$ can reach a value of the order of 10% when the magnetic field is uniformly distributed in the working space. When an inhomogeneous magnetic field, which provides for a particle trapping regime,[32] is employed, the efficiency can be increased severalfold.

When the density of the particles in a large-orbit gyrotron is high and their pitch factor is relatively small, the effects of the rf space charge, which lead, in particular, to appreciable lowering of the start-up currents play a significant role.

According to a calculation, a large-orbit gyrotron with a nanosecond pulse duration and a particle energy of 200–300 keV can be created on the basis of existing, very compact electron accelerators. Probably, generators of a short-pulse magnetic field with a strength up to 300 kOe can also be not very cumbersome. This opens up possibilities for the creation of high-power compact sources of submillimeter radiation, which would be very attractive for many applications. To extend the list of possible applications of such sources, methods should be developed for more broad-band tuning of the frequency at a fixed cyclotron harmonic than is achievable with the use of conventional gyrotron cavities.

[1] A. L. Goldenberg and A. G Litvak, Phys. Plasmas **2**, 2562 (1995).

[2] V. E. Njasnikov, M. V. Agapova, V. V. Alikaev *et al.*, in *Proceedings of the 21st International Conference on Infrared and Millimeter Waves*, (Berlin 1996), ATh1.

[3] D. B. McDermott, N. C. Luhmann Jr., A. Kupiszewski, and H. R. Jory, Phys. Fluids **26**, 1936 (1983).

[4] W. Lawson, W. W. Destler, and C. D. Shtriffler, IEEE Trans. Plasma Sci. **13**, 444 (1985).

[5] D. S. Furuno, D. B. McDermott, C. S. Kou *et al.*, IEEE Trans. Plasma Sci. **18**, 313 (1990).

[6] A. Singh, W. W. Destler, V. L. Granatstein, and W. R. Hix, Int. J. Electron. **72**, 827 (1992).

[7] W. Lawson and C. D. Striffler, Phys. Fluids **28**, 2868 (1986).

[8] G. S. Nusinovich, Int. J. Electron. **72**, 959 (1992).

[9] G. Bekefi, A. DiRienzo, C. Leibovitch, and B. G. Danly, Appl. Phys. Lett. **54**, 1302 (1989).

[10] V. L. Bratman, G. G. Denisov, B. D. Kol'chugin *et al.*, Phys. Rev. Lett. **75**, 3102 (1995).

[11] G. S. Kou, D. B. McDermott, N. C. Luhmann Jr., and K. R. Chu, IEEE Trans. Plasma Sci. **18**, 343 (1990).

[12] A. K. Ganguly and J. L. Hirshfield, Phys. Rev. E **47**, 1993 (1993).

[13] I. I. Antakov, A. V. Gaponov, and V. K. Yulpatov, Vopr. Radioélektron. Ser. 1: Élektron., No. 12, 33 (1965).

[14] M. I. Petelin and V. K. Yulpatov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **18**, 290 (1975).

[15] M. I. Petelin, in *Gyrotrons* [in Russian], Inst. Prikl. Fiz. Akad. Nauk SSSR, Gor'kiĭ (1981), pp. 5–25.

[16] V. K. Yulpatov, in *Gyrotrons* [in Russian], Inst. Prikl. Fiz. Akad. Nauk SSSR, Gor'kiĭ (1981), pp. 26–40.

[17] V. L. Bratman, N. S. Ginzburg, and G. S. Nusinovich, Pis'ma Zh. Tekh. Fiz. **3**, 961 (1977) [Sov. Tech. Phys. Lett. **3**, 395 (1977)].

[18] V. L. Bratman, N. S. Ginzburg, G. S. Nusinovich *et al.*, Int. J. Electron. **51**, 541 (1981).

[19] N. S. Ginzburg and G. S. Nusinovich, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **22**, 754 (1979).

[20] V. L. Bratman, N. S. Ginzburg, G. S. Nusinovich *et al.*, in *Relativistic Radio-Frequency Electronics* [in Russian], Inst. Prikl. Fiz. Akad. Nauk SSSR, Gor'kiĭ (1979), pp. 249–275.

[21] G. S. Nusinovich, Int. J. Electron. **51**, 457 (1981).

[22] V. L. Bratman, Yu. K. Kalynov, V. N. Manuilov *et al.*, in *Abstracts of the 11th International Conference on High Power Particle Beams*, Prague (1996), pp. 1–43.

[23] C. Wang, J. L. Hirshfield, and A. K. Ganguly, Phys. Rev. Lett. **77**, 3819 (1996).

[24] I. S. Kovalev, A. A. Kuraev, S. V. Kolosov, and G. Ya. Slepyan, Dokl. Akad. Nauk BSSR **17**, 416 (1973).

[25] V. L. Bratman and M. I. Petelin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **18**, 1538 (1975).

[26] V. I. Kanavets and V. A. Cherepenin, Radiotekh. Elektron. **20**, 2539 (1975).

[27] V. L. Bratman, Zh. Tekh. Fiz. **46**, 2030 (1976) [Sov. Phys. Tech. Phys. **21**, 1188 (1976)].

[28] M. I. Petelin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **28**, 354 (1985).

[29] V. L. Bratman and A. V. Pylin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **32**, 1408 (1989).

[30] F. S. Rusin, in *High-Power Electronics* [in Russian], Nauka, Moscow (1968), Vol. 5, pp. 9–37.

[31] V. L. Bratman and A. V. Savilov, Phys. Plasmas **2**, 557 (1995).

[32] V. L. Bratman, N. S. Ginzburg, and A. V. Savilov, in *Relativistic Radio-Frequency Electronics, No. 7* [in Russian], Inst. Prikl. Fiz. Akad. Nauk SSSR, Gor'kiĭ (1992), pp. 22–39.

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Influence of cubic anisotropy on the temperature characteristics of magnetostatic waves in ferrite films magnetized in the film plane

V. V. Shagaev

Institute of Materials for the Electronics Industry, 248650 Kaluga, Russia

(Submitted March 12, 1997; resubmitted August 4, 1997)
Zh. Tekh. Fiz. **68**, 99–103 (October 1998)

Expressions for the temperature coefficients of frequency for magnetostatic waves in tangentially magnetized ferrite films having cubic anisotropy are obtained and analyzed. It is shown that in cases where the cubic axis lies in the plane of the film, temperature-induced changes in both the magnitude and direction of the external field can be used for thermostabilization of the frequencies. The problem of two-frequency thermostabilization is considered. The results of experimental investigations of the temperature characteristics of magnetization waves in a tangentially magnetized film of yttrium iron garnet with a {100} surface are presented.

[S1063-7842(98)01710-3]

INTRODUCTION

Ferrite films are employed in investigations of magnetostatic waves (MSWs) and the creation of devices for the superhigh-frequency range.[1,2] The temperature dependence of MSW frequencies arises as a result of temperature-induced changes in the saturation magnetization and anisotropy fields.[3] The temperature characteristics of MSWs have been studied most thoroughly within isotropic film models.[4–6] In particular, explicit approximate expressions have been obtained for the temperature coefficients of frequency at any values of the wave vector.[6] A cubic anisotropy field was taken into account in the investigation of the conditions for thermostabilization of the uniform ferromagnetic resonance (FMR) frequency in Refs. 7 and 8. As it turned out, anisotropy can significantly influence the choice of thermostabilization parameters. In anisotropic films the characteristics of MSWs also depend on the orientation of the magnetization relative to the crystallographic axes.

In this paper we obtain and investigate explicit expressions for the temperature coefficients of frequency of MSWs in tangentially magnetized ferrite films having cubic anisotropy and {100} or {110} surfaces.

STATEMENT OF THE PROBLEM

Let us consider a ferrite film of thickness d magnetized tangentially to saturation by an external static magnetic field **H**. The z axis is oriented in the equilibrium direction of the magnetization vector **M**. The y axis is located on the film surface. The x, y, and z axes form a right-handed orthogonal coordinate system. Let the film have cubic and uniaxial magnetic anisotropy. In deriving the expression for the temperature coefficient of frequency we shall take into account the influence of the anisotropy fields on the equilibrium orientation of the magnetization vector exactly. The calculations are simplified, if the temperature-induced changes in the direction of **M** reduce to rotation in the plane of the film. The magnetization changes will clearly be of just this type, if one of the cubic symmetry planes coincides with the film surface and the uniaxial anisotropy axis coincides with a normal. We shall consider only these cases.

The final geometry of the problem is shown in Fig. 1. The vectors **M** and **H** and the [001] axis lie on the film surface, which coincides with either a (100) or a (110) plane, and the angle γ between the [010] axis and the film surface takes values of 0 and $\pi/4$, respectively. In addition, ψ and ϕ_h are the angles between the [001] axis and the vectors **M** and **H**, respectively.

The nonzero components of the effective demagnetizing anisotropy factor tensor[9] have the form

$$MN_{xx}=-\frac{3}{2}H_c r(1-\cos 2\psi)-H_u\,, \tag{1a}$$

$$MN_{yy}=-\frac{3}{8}H_c(4-r)(1-\cos 4\psi), \tag{1b}$$

$$MN_{zz}=-\frac{1}{8}H_c[(4-r)(3+\cos 4\psi)+4r\cos 2\psi], \tag{1c}$$

$$MN_{yz}=-\frac{1}{8}H_c[2r\sin 2\psi+(4-r)\sin 4\psi], \tag{1d}$$

where $r=(1/2)(1-\cos 4\gamma)$, and H_c and H_u are the strengths of the cubic and uniaxial anisotropy fields, respectively.

The temperature-induced changes in the equilibrium orientation of **M** are given by the derivative $d\psi/dT$ (T is the temperature). It is calculated from the equation for the projection of the external field onto the y axis

$$H\sin(\psi-\psi_h)=MN_{yz}\,. \tag{2}$$

Differentiating with respect to T, we obtain

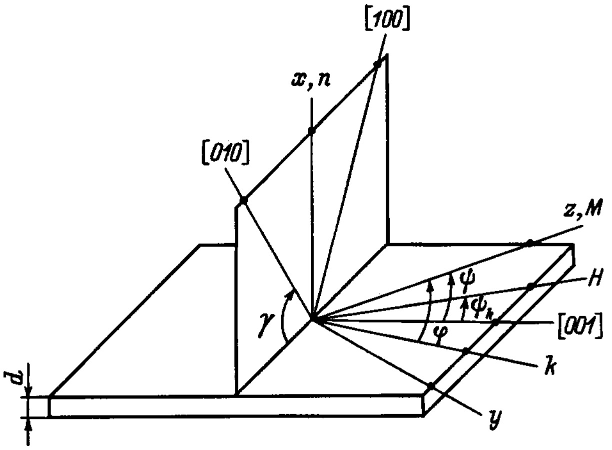

FIG. 1. Geometry of the problem.

$$\frac{d\psi}{dT}=-\frac{1}{8}\left\{[2r\sin 2\psi+(4-r)\sin 4\psi]\left(\frac{dH_c}{dT}-\frac{H_c}{H}\frac{dH}{dT}\right)+H_z\frac{d\psi_h}{dT}\right\}\left\{H_z+\frac{1}{2}H_c[r\cos 2\psi+(4-r)\cos 4\psi]\right\}^{-1}, \tag{3a}$$

where

$$H_z=H\cos(\psi-\psi_h). \tag{3b}$$

The temperature derivative of H_z is needed for the calculations. Differentiating Eq. (3b) and allowing for (2) and (1d), we obtain

$$\frac{dH_z}{dT}=\frac{H_z}{H}\frac{dH}{dT}+\frac{1}{8}H_c[2r\sin 2\psi+(4-r)\sin 4\psi]\times\left(\frac{d\psi}{dT}-\frac{d\psi_h}{dT}\right). \tag{4}$$

CALCULATION OF TEMPERATURE COEFFICIENTS OF FREQUENCY

Let us consider an MSW with a wave vector **k**, which forms an angle φ with the vector **M**. The dispersion relations are derived in the zero-exchange approximation from the magnetostatic equations with consideration of the electrodynamic boundary conditions on the film surface.[10] Dispersion relations of two types are obtained as a result.

Magnetostatic volume waves (MSVWs) can propagate in a film at any angle ψ. The dispersion relations for them have the form

$$kd=\left|\frac{f^2-f_l^2}{f^2-f_t^2}\right|^{1/2}\left\{L\arctan\left[\frac{2|(f^2-f_l^2)(f^2-f_t^2)|^{1/2}}{f_l^2+f_t^2+f_m^2-2f^2}\right]+Lp+\frac{\pi}{2}(1-L)+n\pi\right\}\equiv U_v(f,f_l,f_t,f_m), \tag{5}$$

where

$$f_l^2=g^2[H_z+M(N_{xx}-N_{zz})+4\pi M][H_z+M(N_{yy}-N_{zz})], \tag{6}$$

$$f_t^2=g^2[H_z+M(N_{xx}-N_{zz})]\times[H_z+M(N_{yy}-N_{zz})+4\pi M\sin^2\varphi], \tag{7}$$

$$f_m=4\pi M\,g\sin\varphi, \tag{8}$$

f denotes the frequencies of the wave modes with the numbers $n=0,1,2,\ldots$; $g=2.8$ MHz/Oe is the gyromagnetic ratio; $L=1$ if $f_t>f_l$ (magnetostatic forward volume waves — MSFVWs), and $L=-1$ if $f_t<f_l$ (magnetostatic backward volume waves — MSBVWs); $p=0$ if $2f^2\leqslant f_l^2+f_t^2+f_m^2$, and $p=\pi$ when the reverse inequality holds; the continuous branch in the arctan function with a range of values from $-\pi/2$ to $\pi/2$ is used; the MSVW frequencies lie in the interval between f_l and f_t; and the frequency $f=f_l$ corresponds to uniform FMR (even when $f_l=f_t$).

Magnetostatic surface waves (MSSWs) exist in the range of angles $0\leqslant\varphi<\varphi_0$, where φ_0 is the root of the equation $|f_t^2-f_l^2|=f_m^2$ solved with respect to φ. Here f_l, f_t, and f_m are defined by (6)–(8), in which the MN_{qq} $(q=x,y,z)$ depend on φ according to Eqs. (1a)–(1c), and the dependence of H_z on φ is specified by Eqs. (2) and (3b).

The dispersion relation of the MSSWs has the form

$$kd=\frac{1}{2}\left|\frac{f^2-f_l^2}{f^2-f_t^2}\right|^{1/2}\times\ln\left\{1+\frac{4[(f^2-f_l^2)(f^2-f_t^2)]^{1/2}}{f_m^2-[(f^2-f_l^2)+(f^2-f_t^2)^{1/2}]^2}\right\}\equiv U_s(f,f_l,f_t,f_m). \tag{9}$$

The MSSW frequencies lie in the interval $\max\{f_l,f_t\}<f<(1/2f_m)[(f_l^2+f_t^2+f_m^2)^2-4f_l^2f_t^2]^{1/2}$. If $f_t>f_l$, the dispersion curve of the fundamental mode of the MSFVWs transforms smoothly into the dispersion curve of the MSSWs at f_t and the wave vector value $kd=2(f_t^2-f_l^2)/(f_m^2-f_t^2+f_l^2)$.

The temperature coefficient α for the frequency f corresponding to an assigned value of the wave vector is defined as $\alpha=(1/f)(df/dT)$. The temperature coefficients of frequency of the MSVWs and MSSWs can be calculated using the formula, which follows from the form of the dispersion relation,

$$\alpha_j=-\frac{i}{f^2(\partial U_j/\partial f^2)}\left(\frac{\partial U_j}{\partial f_l^2}\frac{df_l^2}{dT}+\frac{\partial U_j}{\partial f_t^2}\frac{df_t^2}{dT}+\frac{\partial U_j}{\partial f_m^2}\frac{df_m^2}{dT}\right).$$

Here $j=v,s$ for the MSVWs and MSSWs respectively. For the sake of brevity, the arguments in the functions U_j are not indicated. The partial derivatives of U_j are calculated from the explicit dependences (5) and (9). Analytic expressions for the temperature derivatives are obtained by differentiating (6)–(8) and replacing $d\psi/dT$ and dH_z/dT according to formulas (3a) and (4) (with allowance for the obvious equality $d\varphi/dT=d\psi/dt$). Substitution of the expression for the derivatives into the formula for α_j yields the explicit dependences of the temperature coefficient of frequency on the frequency of the MSW and on the field and film parameters. The dependence of the temperature coefficient of frequency on kd is assigned parametrically in terms of f and the dispersion relation.

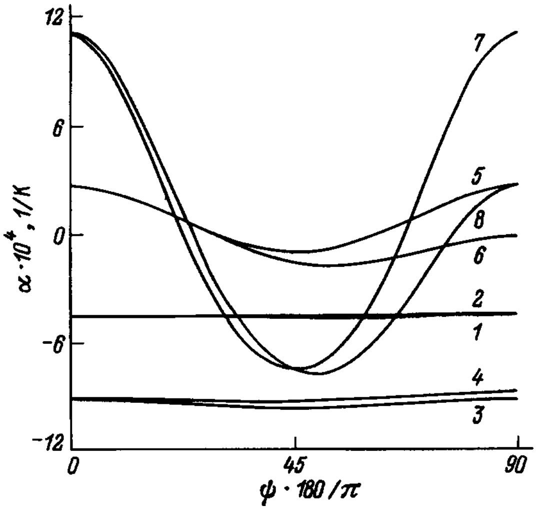

FIG. 2. Calculated plots of the angular dependence of the contributions of the temperature-induced changes in the saturation magnetization and the cubic anisotropy field to the temperature coefficient of frequency for $f_l=3$ GHz (*3,4,7,8*) and 10 GHz (*1,2,5,6*) and for $\gamma=0$ (*1,3,5,7*) and $\pi/4$ (*2,4,6,8*): *1–4* — $dH_c/dT=0$, *5–8* — $d(4\pi M)/dT=0$.

ANALYSIS OF THE ANGULAR DEPENDENCES OF THE TEMPERATURE COEFFICIENT OF FREQUENCY

The exact expressions for the temperature coefficient of frequency obtained by the method described above were investigated numerically. The formula for calculating α_j can clearly be brought into the form

$$\alpha_j=A_j\frac{d(4\pi M)}{dT}+B_j\frac{dH_c}{dT}+C_j\frac{dH_u}{dT}+D_j\frac{dH}{dT}+E_j\frac{d\psi_h}{dT}. \tag{10}$$

Let us first discuss the dependence of the temperature coefficient of frequency on the angle ψ in the case where the external field does not depend on temperature $(dH/dT=0$, $d\psi_h/dT=0)$. The film parameters in the calculations were as follows: $4\pi M=1750$ G, $H_c=-42$ Oe, $H_u=0$, $d(4\pi M)/dT=-4.15$ G/K, $dH_c/dT=0.46$ Oe/K, $dH_u/dT=0$. The angular dependences were calculated for constant values of f and kd, and the values of the field projection H_z were determined from the dispersion relation. For this purpose, the expressions for f_l, f_t, and f_m from Eqs. (6)–(8) were substituted into Eq. (5) or (9), and the equation was solved in such a form for H_z.

As an example, Fig. 2 presents calculated plots of the first two terms in (10) as functions of ψ for the long-wavelength edge of the MSW spectrum ($kd=0$, $f=f_l$). The following values of the parameters were used: $f_l=3.10$ GHz; $\gamma=0,\pi/4$. The dependences on ψ are symmetric relative to the values $\psi=(\pi/2)n$, where $n=1,2,\dots$, and the curves can easily be continued for $\psi>\pi/2$.

An analysis of the angular dependences reveals that the films considered exhibit considerable anisotropy of the temperature coefficient of frequency. The main contribution to the dependence on ψ is made by the second term in (10). Its maximum values, which are achieved near $\psi=0$, are positive and significantly influence the magnitude and sign of the temperature coefficient of frequency.

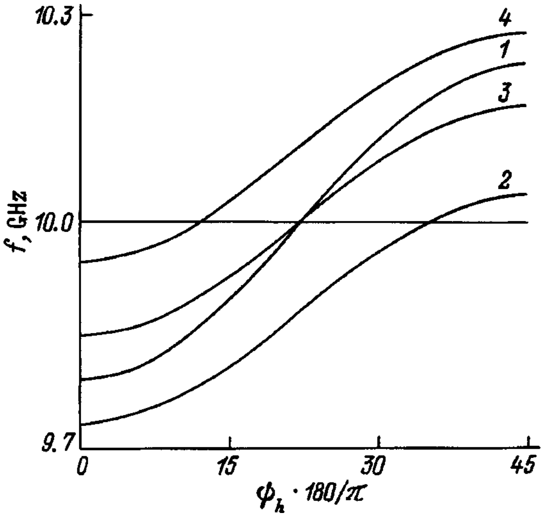

FIG. 3.

Let us consider the problem of compensation of the temperature-induced changes in the MSW frequencies by temperature-induced changes in the magnetizing field.[11,12] The anisotropic properties of films with a tangential cubic axis permit the use of changes in both the magnitude and direction of the external field for solving this problem. We shall illustrate this in the example of a film with a $\{100\}$ surface.

Each of the four curves in Fig. 3 was obtained from a calculation and shows the dependence of the frequency of the long-wavelength edge of the MSW spectrum on the angle between the external field and the $[001]$ axis. This dependence is obtained from Eqs. (2) and (6) in parametric form with parameter ψ. Curve *1* was calculated for the following initial values of the parameters: $4\pi M=1750$ G, $H_c=-42$ Oe, and $H=2826$ Oe. Curves *2–4* were obtained for $4\pi M=1646.25$ G and $H_c=-30.5$ Oe, which correspond to the changes in the original values for a 25 K rise in temperature when $d(4\pi M)/dT=-4.15$ G/K and $dH_c/dT=0.46$ Oe/K. The external field strengths were 2826, 2860, and 2888.5 Oe for curves *2–4*, respectively. All the curves pass through the frequency value $f=10$ GHz, which is marked in the figure by a straight line. It can be seen that thermal compensation can be achieved at this frequency by altering the magnitude (curve *3*), the direction (curve *2*), or both the magnitude and direction (curve *4*) of the external field. The last two techniques clearly stem from the angular dependence of the frequency displayed when the cubic anisotropy is taken into account. The frequency range on this dependence for yttrium iron garnet (YIG) films with a tangential cubic axis amounts to hundreds of megahertz, while the range for the widely used YIG$\{111\}$ films amounts to only a few megahertz, and thermostabilization of the frequencies in them by turning the magnetizing field is possible only for slight temperature changes.

In the example considered thermostabilization of the frequency could be accomplished using different pairs of values of dH/dT and $d\psi_h/dT$. To uniquely specify these parameters, two conditions, rather than one, must be imposed on

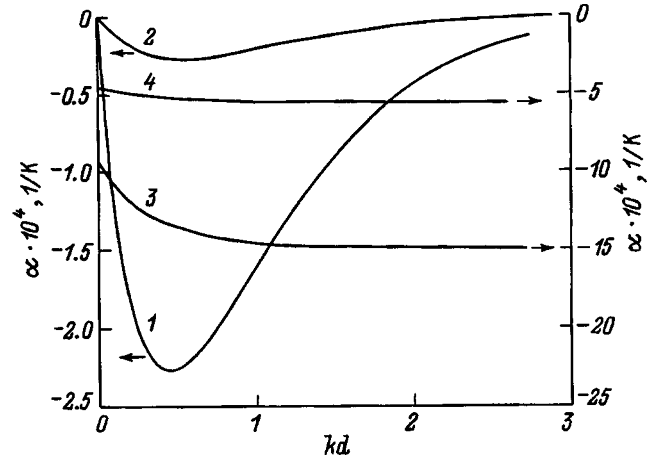

FIG. 4. Dependence of the temperature coefficient of frequency on kd for an isotropic film (*3,4*) and for a film with cubic anisotropy and thermostabilized frequencies at the beginning and end of the MSSW spectrum (*1,2*).

the temperature coefficient of frequency. We assign the conditions for a vanishing temperature coefficient of frequency at two frequency values from the MSW spectrum. This problem corresponds to a system of two equations with the two unknowns dH/dT and $d\psi_h/dT$:

$$\alpha_j(f_1,H_z,dH/dT,d\psi_h/dT)=0,$$

$$\alpha_i(f_2,H_z,dH/dT,d\psi_h/dT)=0. \qquad (11)$$

Here $j=\nu,s$ and $i=\nu,s$. The functional dependences on the left-hand sides of the equations follow from the explicit expressions for the temperature coefficient of frequency: f_1 and f_2 are the assigned values of the MSW frequencies. The value of H_z is calculated from the dispersion relation for an assigned value of the wave vector and one of the frequencies.

It follows from expression (10) that the equations of the system (11) are linear with respect to dH/dT and $d\psi_h/dT$ and that solutions can be obtained in an explicit form. We present an example of the solution of the system (11). We consider a film with a $\{100\}$ surface and the same values for the magnetic parameters as in the preceding examples. We choose the frequencies of the beginning and end of the MSSW spectrum for the angles $\psi=\pi/2$ and $\psi=\pi/8$ as the frequencies being thermostabilized. When $H_z=520$ Oe, the spectrum lies in the frequency range from 3 to 3.818 GHz, and when $H_z=2826$ Oe, it lies in the range from 10 to 10.275 GHz. In the former case the solutions are $dH/dT=4.6$ Oe/K and $(d\psi_h/dT)(180/\pi)=-0.741$ K^{-1}, and in the latter case they are $dH/dT=4.611$ Oe/K and $(d\psi_h/dT)\times(180/\pi)=-0.781$ K^{-1}. Figure 4 shows plots of the temperature coefficient of frequency as a function of kd for both cases: curves *1* and *2* are for 3 and 10 GHz, respectively. For comparison, the figure includes curves *3* and *4*, which were calculated for an isotropic film ($H_c=0$, $dH_c/dT=0$). For curve *3*, $H_z=508.3$ Oe, and the MSSW spectrum lies in the frequency range from 3 to 3.873 GHz. For curve *4*, $H_z=2802$ Oe, and the spectrum lies in the range from 10 to 10.296 GHz. It is seen that the absolute values of the temperature coefficient of frequency on the curves with two-frequency thermostabilization are significantly smaller than those on the curves for an isotropic film in the respective frequency ranges.

EXPERIMENT

The anisotropic properties of a tangentially magnetized YIG film associated with the dependence of the temperature coefficients of frequency of MSWs on the orientation of the magnetization vector relative to the cubic axes were revealed in the experiment described below.

A YIG film of thickness 9.8 μm on a $\{100\}$ gadolinium gallium garnet substrate measuring 13×13 was used. The magnetic parameters were as follows (at $T=273$ K): $4\pi M_{\mathrm{eff}}=4\pi M-H_u=1863$ G, $H_c=-59$ Oe, $d(4\pi M_{\mathrm{eff}})/dT=-4.1$ G/K, $dH_c/dT=0.44$ Oe/K. The MSWs were excited and received using microstrip transducers (antennas) in analogy to Ref. 13. The interference of the signals received in two identical antennas at different distances from identical transmitting antennas was used. The antennas were fabricated on a Polikor (crystalline leucosapphire) substrate and connected to microstrip summing circuits. This design permits construction of the $f(kd)$ curve from the amplitude-frequency characteristic.[13] The film was placed between the poles of an electromagnet in a tangential magnetic field. The orientation of the magnetization vector relative to the tangential cubic axes was assigned by rotating the film about a normal. The temperature changes were caused by cooling in nitrogen vapor and heating with the aid of a flat heating element. A thermocouple was used to measure the temperature. The temperature coefficients of frequency for the assigned values of kd were calculated from the $f(kd)$ dispersion curves using the formula

$$\alpha(kd)=[1/f_e(kd)][f_2(kd)-f_1(kd)]/(T_2-T_1),$$

where

$$f_e(kd)=[f_1(kd)+f_2(kd)]/2,$$

and the frequencies $f_1(kd)$ and $f_2(kd)$ correspond to T_1 and T_2.

Measurements were performed for two film orientations. In one case the direction of the magnetization coincided with the $\langle 100\rangle$ axis, and in the other case it coincided with the $\langle 110\rangle$ axis. These orientations were achieved by rotating the film about a normal and monitoring the position of the frequency range of the MSW spectrum on the amplitude-frequency characteristic. The smallest and largest values of the MSW frequencies are achieved when the direction of the magnetization vector coincides with the $\langle 100\rangle$ and $\langle 110\rangle$ axes, respectively.

Figure 5 presents experimental and theoretical plots of $f(kd)$ and $\alpha(kd)$ (at $T=273$ K). It was assumed in the calculations that $H_u=0$ and $dH_u/dT=0$. The observed differences for the two directions of the magnetization vector are clearly attributable to the cubic anisotropy of the film. It follows from the calculations that the values of the contribution of the temperature-induced changes in the cubic anisotropy field to the temperature coefficient of frequency for the directions considered differ significantly [the values on curve

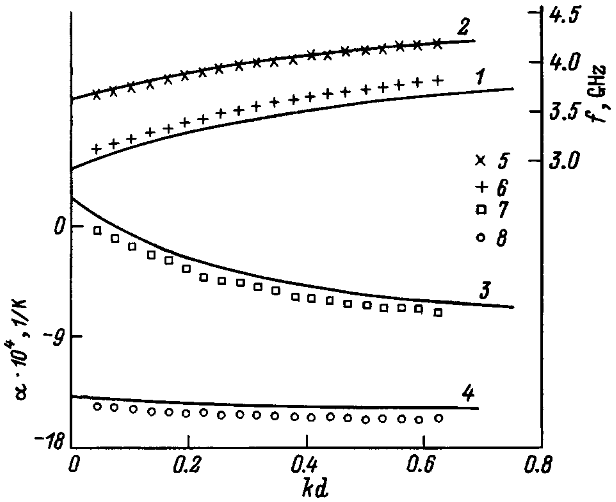

FIG. 5. Plots of $f(kd)$ (*1,2,5,6*) and $\alpha(kd)$ (*3,4,7,8*) in a YIG{100} film: *1,3,6,7* — **M**‖[001]; *2,4,5,8* — **M**‖[011]. Solid curves — theory.

7 (Fig. 2) at $\psi = 0$ and $\pi/4$ differ similarly]. The contributions of the temperature-induced changes in the magnetization differ only slightly [as do the values on curve *3* (Fig. 2) at the same values of ψ].

CONCLUSION

The cubic anisotropy of a ferrite film can have a significant influence on the temperature characteristics of MSW devices. Doped YIG films can be used to both enhance and weaken its influence. The methods that have been developed for theoretically analyzing the temperature coefficient of frequency are applicable in this case, too, since they take into account the cubic anisotropy field exactly regardless of its magnitude. The anisotropic properties of tangentially magnetized films with a cubic axis lying on the surface can be used to design MSW instruments with assigned temperature characteristics.

We thank A. V. Maryakhin for preparing the YIG film.

[1] J. P. Castera, J. Appl. Phys. **55**, 2506 (1984).

[2] P. Hartemann, IEEE Trans. Magn. Mag-20, 1272 (1984).

[3] Yu. M. Yakovlev and S. Sh. Gendelev, *Ferrite Single Crystals in Radio Electronics* [in Russian], Sov. Radio, Moscow (1975), 360 pp.

[4] A. S. Beregov, E. V. Kudinov, and I. N. Ereshchenko, Élektron. Tekh., Ser. 1: Élektron. SVCh No. 1(395), 19 (1987).

[5] Yu. K. Fetisov, Zh. Tekh. Fiz. **57**, 2393 (1987) [Sov. Phys. Tech. Phys. **32**, 1451 (1987)].

[6] A. Slavin and Yu. K. Fetisov, Zh. Tekh. Fiz. **58**, 2210 (1988) [Sov. Phys. Tech. Phys. **33**, 1343 (1988)].

[7] L. V. Lutsev and I. L. Berezin, Élektron. Tekh. Ser. 1: Élektron. SVCh, No. 6(420), pp. 3–8 (1989).

[8] S. V. Belyakov and O. A. Gorodaĭkina, Élektron. Tekh. Ser. 1: Élektron. SVCh, No. 7(391), pp. 28–33 (1986).

[9] A. G. Gurevich, *Magnetic Resonance in Ferrites and Antiferromagnets* [in Russian], Nauka, Moscow (1973), 591 pp.

[10] A. S. Beregov, Izv. Vyssh. Uchebn. Zaved. Ser. Radioélektron. **27**(10), 9 (1984).

[11] J. D. Adam and M. R. Daniel, IEEE Trans. Magn. Mag-17, 2951 (1981).

[12] J. D. Adam, in *IEEE MTT-S International Microwave Symposium Digest* (1979), pp. 160–161.

[13] B. N. Gusev, O. A. Chivileva, A. G. Gurevich *et al.*, Pis'ma Zh. Tekh. Fiz. **9**, 159 (1983) [Sov. Tech. Phys. Lett. **9**, 70 (1983)].

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Photostimulated diffusion of Na atoms in the alloy NaAu. I. Kinetic characteristics

M. V. Knat'ko, V. I. Paleev, and M. N. Lapushkin

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

(Submitted November 18, 1997)
Zh. Tekh. Fiz. **68**, 104–107 (October 1998)

The results of an investigation of the photoactivation of the diffusion of Na atoms in the subsurface region of heated gold are presented. The dependences of the rate of this process on substrate temperature and the concentration of alkali-metal atoms in the adlayer and in the subsurface layer of gold are obtained. The kinetic characteristics of photostimulated diffusion are found. It is shown that this process is caused by the action of light on the alloy NaAu, which leads to an increase in the rate of the diffusive motion of Na atoms in the subsurface layer of the solid. © *1998 American Institute of Physics.* [S1063-7842(98)01810-8]

INTRODUCTION

The adsorption of alkali-metal atoms on a gold substrate at room or elevated temperatures is accompanied by the formation of the alloy AuM (M is an alkali-metal atom) in the subsurface layer. The compounds formed have properties which differ significantly from the properties of the metals comprising the alloy. For example, the Cs–Au alloy is a semiconductor with a 2.6 eV band gap.[1] A semiconductor electronic structure has also been predicted for equiatomic alloys of gold with rubidium.[2]

An investigation of photochemical reactions involving alkali-metal and halogen atoms on a polycrystalline gold surface at high temperatures revealed that irradiation of the surface by low-intensity visible light leads to an increase in the rate of the diffusive transport of the alkali-metal atoms.[3] It was shown that the irradiation of a gold ribbon by light with $h\nu > 2.6$ eV stimulates the emergence of Cs, Na, and K impurity atoms implanted in the subsurface region of the solid on its surface and that, conversely, at high surface coverages of gold with an alkali metal illumination causes the passage of M atoms from the surface into the bulk of the metal.

This paper presents the results of further investigations of the effect of light on the diffusion of atoms in the subsurface layers of gold when sodium serves as an adsorbate. As in Ref. 3, a method based on surface-ionization detection of the alkali-metal atoms was employed. The current of Na^+ ions in the flow of particles thermodesorbed from the surface was measured. Since the fraction of ions in the flow of desorbed particles of the same chemical composition during surface ionization at a constant emitter temperature T, a stable work function φ, and low surface concentrations of the adsorbate was determined unequivocally, the variation of the concentration of particles on the emitter surface n_c can be traced by monitoring the variation of the ion current. The surface-ionization detection of alkali-metal atoms having low ionization energies can be performed with very high efficiency and has been widely used[4] since the discovery of this phenomenon. In this case the source of supply of the particles being ionized to the emitter surface does not play any role. This method is applicable both for detecting intrinsic thermionic emission and for detecting particles supplied from without.

EXPERIMENTAL PROCEDURE

The mass-spectrometric system used in the experiments to investigate the photostimulation of diffusion was previously described in Ref. 5; the gold samples were precleaned just as in Ref. 3.

The kinetics of the formation of the alloy NaAu in the subsurface region of a heated gold substrate when Na atoms are adsorbed on it, as well as the removal of an alkali-metal impurity, were considered in detail in Ref. 6.

Gold emitters prepared in two ways were used to study the effect of light on the diffusion of Na in gold. In some experiments a flow of Na atoms with a density at the surface $i = 10^{12} - 10^{13}$ $cm^{-2}s^{-1}$ was directed onto a cleaned gold surface, and the ribbon was held at a constant temperature T in the range $600 < T < 1200$ K. The deposition of Na was continued under these conditions for 24 h, and then the flow of the alkali metal onto the surface was stopped, and the ribbon was annealed for $2-3$ h at $1150 < T < 1250$ K. After such treatment of the ribbon, the intrinsic thermionic emission of Na^+, which is a consequence of the emergence of implanted Na atoms from the bulk on the surface, was recorded. The value of the dark ion current I_b gradually decreased in accordance with the variation of the bulk concentration of the alkali-metal impurity in the gold, but after a 150-h anneal, the emission of Na^+ ions did not cease completely.

In another series of experiments a flow of Na atoms was directed onto a heated gold ribbon freed of alkali-metal impurities. The value of the dark current for the surface ionization of Na (I_b) depended on the emitter temperature T, the work function φ, and the flux density of Na impinging on the surface, as well as on the rates of the diffusive exchange of sodium between the surface and subsurface layers of the substrate.

Photostimulation of the diffusive emergence of Na atoms on the surface (or of their passage from the substrate

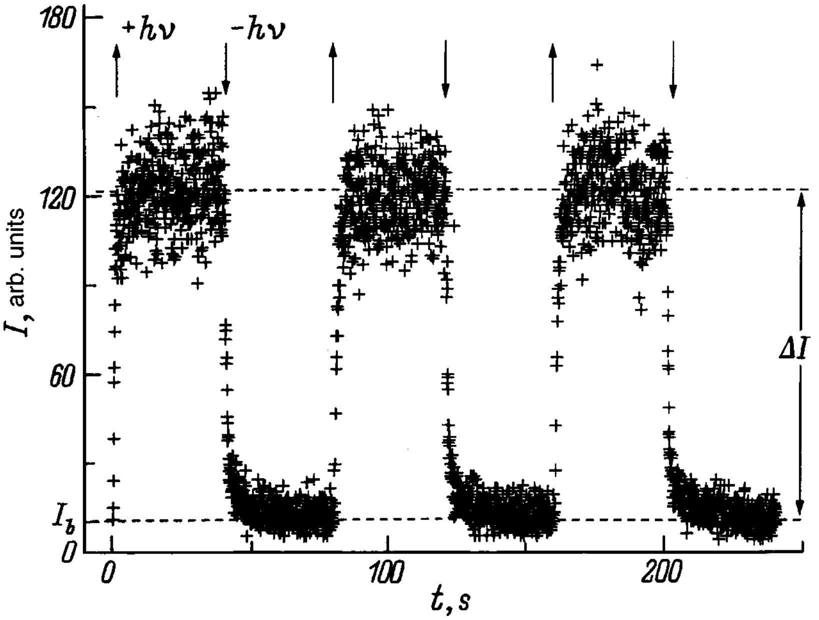

FIG. 1. Dependence of the current of Na^+ ions on the exposure time of a gold ribbon preliminarily doped with sodium and then annealed. $T=1150$ K, $h\nu=3.52$ eV.

surface into the subsurface region) was determined from the change in current $\Delta I=|I-I_b|$, where I is the current of Na^+ ions when the substrate is illuminated.

As an example, Fig. 1 shows the characteristic form of the dependence of the current of Na^+ ions on exposure time during irradiation of a heated gold ribbon preliminarily doped with sodium and then annealed in the absence of an external flow of Na by monochromatic light with an intensity <10 mW/cm^2. In order to cause a similar increase in intrinsic thermionic emission by heating the ribbon, its temperature would have to be raised by $\Delta T=45-50°$, which would require a large expenditure of energy and would be easily determined pyrometrically. In the case under consideration the increase in T as a consequence of radiation heating amounted to $\Delta T \ll I_0$ and could not have a measurable influence on ΔI.

As the emitter temperature T is increased, the value of $\Delta I/I_b$, which characterizes the efficiency of the photoactivation of the diffusive emergence of Na on the surface, decreases monotonically over the entire temperature range investigated $1000<T<1200$ K, as is shown in Fig. 2.

The measured dependence of the photostimulated diffusion of Na on the concentration of these atoms in the subsurface region of gold has a complicated character. The change in current ΔI is associated with the photoinduced changes in the surface concentration of Na atoms on the substrate. The diffusive flux density of the atoms is determined by their concentration gradient between the surface and the subsurface region, as well as by the diffusion coefficient, which is related to the height of the energy barrier on the surface of the solid. The structure of the surface barrier and the thermionic-emission properties of the adsorption system depend both on the concentration of sodium in the subsurface region and on the density of Na on the substrate surface.

In the case of gold preliminarily saturated with sodium and partially freed of the alkali-metal impurity by a high-temperature anneal, photostimulation of the emergence of Na on the surface followed by its desorption is clearly recorded (Fig. 1). In the case where a flux i of Na atoms is directed from an evaporator onto the gold surface, the surface concentration of sodium is determined not only by the supply of these atoms from without, but also by their diffusive exchange between the surface and subsurface layers.

By varying the external flux density i of Na we can influence the magnitude and direction of the concentration gradient of these atoms between the upper and subbarrier layers of the substrate and thereby alter the magnitude and direction of the diffusive flux. A plot of the variation of the photostimulated part of the diffusive flux of Na atoms between the adlayer and the interior of the gold sample as a function of the intensity of the atomic flow impinging on the surface as i systematically increases is plotted on a logarithmic scale in Fig. 3.

The fraction of particles in the photostimulated flow for which the deposited Na is responsible was defined here as

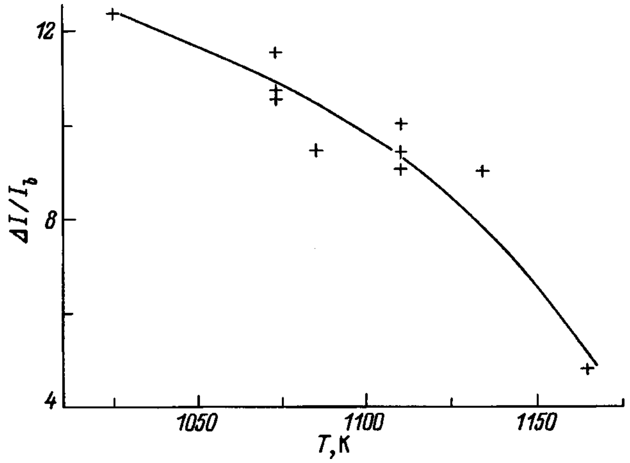

FIG. 2. Temperature dependence of the efficiency of the photoactivation of diffusion in a gold ribbon preliminarily doped with sodium and then annealed. $h\nu=3.52$ eV.

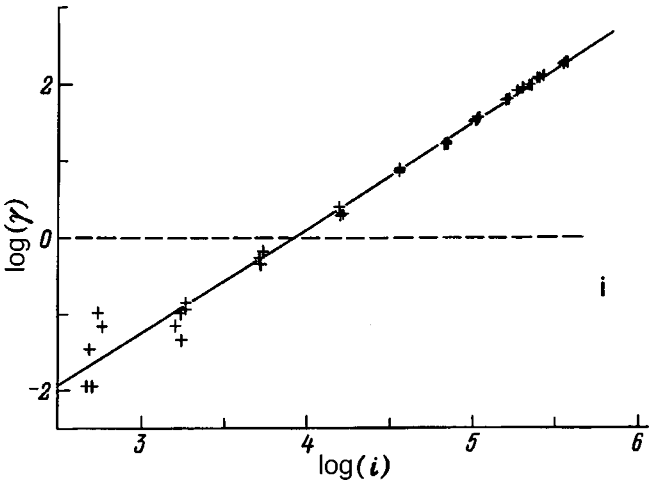

FIG. 3. Dependence of γ on the flux density of Na being deposited.

$$\gamma=|(\Delta I_p-\Delta I_0)/\Delta I_0|,$$

where ΔI_p is the photoresponse of the ion current when the flow of Na is supplied and ΔI_0 is the photoresponse when the flow is stopped.

At some fairly large values of i illumination causes a decrease in n_c for sodium due to the stimulation of its passage from the surface into the bulk of the solid.

The plot shown has a hysteretic character, and high-temperature annealing of the gold is needed to restore the original value of γ.

Figure 4 shows the variation of the current of Na^+ ions with time after illumination of the surface is switched on or switched off in the case where the resultant diffusive flux is directed into the interior of the substrate. The form of the $I(t)$ curve in Fig. 4 is determined by the kinetics of the variation of the coverage of the substrate surface with sodium, which, in turn, reflects the variation of the diffusive fluxes of the alkali-metal atoms determining n_c.

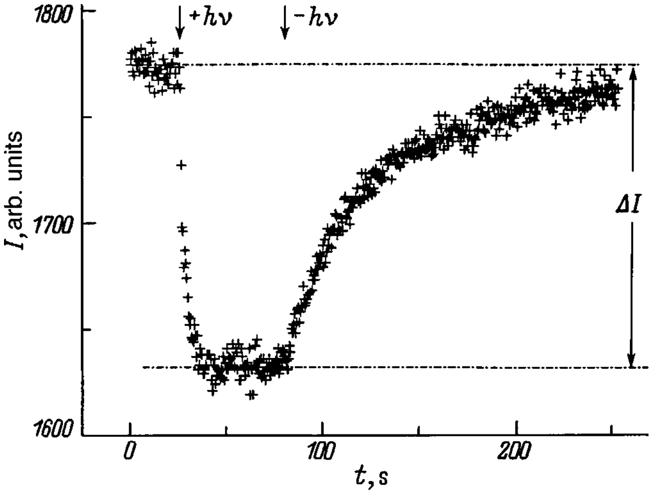

FIG. 4. Dependence of the current of Na^+ ions on the exposure time during the deposition of Na atoms on gold. $T=1150$ K, $h\nu=3.52$ eV, $i=10^{12}\ \mathrm{cm}^{-2}\mathrm{s}^{-1}$.

Several conclusions regarding the reasons for the appearance of the photostimulated diffusion of Na atoms in gold and the parameters of the process can be drawn on the basis of the experimental results presented.

DISCUSSION OF EXPERIMENTAL RESULTS

The photostimulated diffusion of Na atoms in heated gold requires the preliminary creation of a concentration of the alkali metal sufficient for forming the alloy NaAu in the subsurface region of metallic gold. This is indicated by the absence of photoactivation of the diffusion of microscopic alkali-metal impurities in pure gold. On the other hand, photostimulation of the diffusion of alkali-metal atoms appears simultaneously with a decrease in the rate of sublimation of Au,[7] which shows that a layer differing in chemical behavior from the pure metal is formed on the gold surface. Thus, the process observed is associated with the primary effect of light on AuNa accompanied by the diffusion of Na atoms in this compound. As Fig. 1 shows, exposure of the alloy to light significantly increases the rate of diffusion in comparison to the dark rate.

Since in the absence of illumination only an insignificant portion of the flow of atoms directed at the surface passes into the substrate, it can be assumed that at low substrate coverages, which do not alter φ, the value of n_c is directly proportional to i. As follows from the plot of $\log\gamma=f(\log i)$ in Fig. 3, the light-induced diffusive flux of Na is related to n_c by the expression $\gamma=Cn_c^{1.32}$. It can be presumed that this power law is specified by such factors as the variation of the diffusion coefficient of Na in NaAu and the variation of the optical characteristics of the alloy with variation of the Na concentration, as well as, possibly, with variation of the phase composition of $\mathrm{Na}_x\mathrm{Au}_y$. Additional investigations are needed to ascertain which of these factors plays the leading role. However, several conclusions can be drawn from a comparison of the fronts of the photoinduced current change ΔI for the two substrate states represented in Figs. 1 and 4.

The kinetics of the growth and decline of ΔI depend on the rate of the slowest of the processes determining the variation of the concentration n_c when illumination of the substrate is switched on and switched off. These processes are the diffusion of Na (to the surface from the interior of the substrate or in the reverse direction) and the thermal desorption of these atoms accompanied by surface ionization.

Let us consider the case represented in Fig. 1, which shows the effect of light on the rate of the desorption of Na from preliminarily doped and annealed gold. In the absence of light there is a slow decrease in I. Periodic illumination of the surface causes modulation of the current with the time constants for increases and decreases $\tau^{h\nu}=\tau^0=0.1$ s, respectively. For all practical purposes, these values of τ coincide with the lifetime of Na atoms on the surface of a gold substrate at $T=1130$ K if the activation energy for the desorption of Na from a gold surface obtained in Ref. 8, $l=230$ kJ/mol, is used to calculate it. Therefore, the rate of the photoactivated diffusion of Na in the subsurface layers of a ribbon partially depleted of alkali-metal atoms is not the rate-limiting factor in the observed kinetics of the desorption of Na^+ ions.

In the case represented in Fig. 4 a surface of strongly doped Au is supplied with a flow of atoms that maintains a concentration n_c, which is sufficient for growth of the alloy into the interior of the substrate as a result of the diffusion of Na from the surface, at the same values of T. When light is switched on, the time constant of the drop in I as a result of photoactivation of the diffusion of Na from the surface $\tau^{h\nu} = 4.2$ s, and when the illumination is disrupted, the characteristic time for the restoration of I to the stationary value for dark emission $\tau^0 = 55$ s.

In our opinion, one possible explanation for the observed differences in the effect of light and its relaxation in the cases of weakly and strongly doped Au is as follows. As estimates from experimental data[6] show, the activation energies for the diffusion of Na atoms in NaAu and in metallic gold differ and amount to $\Delta E(\mathrm{NaAu}) = 3.15 - 3.20$ eV and $\Delta E(\mathrm{Au}) = 5.1$ eV, respectively.

During the annealing of a ribbon, there is significant depletion of alkali-metal atoms from its surface layer, which results in ''conservation'' of the alloy in the low-lying subsurface layers. Thermal decomposition of the alloy is accompanied by the weak dark emission of alkali-metal atoms which have succeeded in crossing the layer depleted of the alloy. Light causes degradation of the alloy, which is accompanied by an increase in the concentration of free Na on its boundary near the surface and a corresponding increase in the flux of these particles onto the surface. On the basis of the fact that the rate-limiting step of the observed photoprocess is thermodesorption of the particles, it can be assumed that Na is supplied to the surface through a narrow depleted layer as a result of 1–2 jumps.

In the case of gold strongly doped with sodium, the activation energy for the diffusion of Na in NaAu is apparently smaller than the activation energy in annealed gold and depends on the concentration of Na in the subsurface region. In this case exposure to light causes an increase in the transport of Na into the interior of the substrate toward the reaction zone for formation of the alloy. Since the photoactivation efficiency decreases with the penetration depth of light into the solid and since the alloy formed extends to a greater depth, the resultant influence of illumination on the variation of the diffusive flux of Na in the subsurface layer, which determines the concentration of these atoms on the substrate surface, is complicated and affects a large number of layers. In this case the recorded rates of growth of the effect of exposure to light and of its decline after the illumination is disrupted are determined by the kinetics of the diffusion processes in the alloy and the rate of restoration of the distribution of the Na concentration after illumination.

As the investigations have shown, the photostimulation of diffusion is associated with excitation of the electronic system of the alkali-metal/gold alloy, and this question will be the subject of our forthcoming publications.[9]

This work was performed with the support of a grant from the Russian Fund for Fundamental Research (Project No. 96-02-16908) and the ''Surface Atomic Structures'' State Program of the Ministry of Science and Technology of the Russian Federation (Project No. 95-2.13).

[1] W. E. Spicer, A. H. Sommer, and J. G. White, Phys. Rev. **115**, 57 (1959).
[2] C. Koenig, N. E. Christensen, and J. Kollar, Phys. Rev. B **29**, 6481 (1984).
[3] É. Ya. Zandberg, M. V. Knat'ko, V. I. Paleev *et al.*, Pis'ma Zh. Tekh. Fiz. **21**(19), 15 (1995) [Sov. Tech. Phys. Lett. **21**(10), 774 (1995)].
[4] É. Ya. Zandberg and N. I. Ionov, *Surface Ionization* [in Russian], Nauka, Moscow (1969), 432 pp.
[5] É. Ya. Zandberg, M. V. Knat'ko, and V. I. Paleev, Pis'ma Zh. Tekh. Fiz. **12**, 979 (1986) [Sov. Tech. Phys. Lett. **12**, 404 (1986)].
[6] M. V. Knat'ko, V. I. Paleev, and M. N. Lapushkin, Phys. Low-Dimens. Struct., Nos. 5–6, 85 (1998).
[7] M. V. Knat'ko, V. I. Paleev, and E. Ya. Zandberg, Phys. Low-Dimens. Struct., Nos. 7–8, 27 (1996).
[8] A. Neumann, S. L. M. Schroeder, and K. Christmann, Phys. Rev. B **51**, 17 007 (1995).
[9] M. V. Knat'ko, M. N. Lapushkin, and V. I. Paleev, Zh. Tekh. Fiz. **68**(10), 108 (1998) [Tech. Phys. **43**, 1235 (1998)].

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Photostimulated diffusion of Na atoms in the alloy NaAu. II. Spectral and polarization dependences

M. V. Knat'ko, M. N. Lapushkin, and V. I. Paleev

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

(Submitted November 18, 1997)
Zh. Tekh. Fiz. **68**, 108–111 (October 1998)

The results of an investigation of the spectral and polarization dependences of the photostimulated diffusion of sodium in heated gold are presented. It is shown that the light-induced increase in the diffusion rate is associated with optical excitation of the electronic structure of the alloy NaAu, which forms in the subsurface layer of gold during the adsorption of the alkali-metal atoms. The semiconductor properties of this alloy are revealed, and the gap width is estimated. © *1998 American Institute of Physics.* [S1063-7842(98)01910-2]

INTRODUCTION

Our study of the effect of light on chemical reactions involving alkali-metal atoms on metal surfaces revealed[1] that the irradiation of a gold surface by photons with $h\nu > 2.6$ eV causes acceleration of the diffusive exchange of Cs, K, and Na atoms between the adlayer on the surface and the subsurface region of the gold substrate. The photostimulated diffusion of alkali-metal atoms (M) was recorded using surface-ionization detection of the alkali-metal atoms by measuring the change in the current of thermodesorbed M^+ ions upon irradiation of the surface with low-intensity light, which does not lead to appreciable heating of the solid. It was found[1] that the irradiation of a gold substrate preliminarily doped with M atoms leads to an increase in the emission of M^+ ions in comparison to the dark level as a result of the photostimulated diffusion of M atoms to the emitter surface and their surface ionization upon thermodesorption. When flows of M of large flux density are supplied to the gold surface from an external source, illumination of the gold causes the photostimulated diffusion of adatoms into the interior of the substrate. As a result, the surface concentration of M atoms decreases in comparison to the dark value, and the current due to their surface ionization consequently decreases. As was shown in Ref. 2, photostimulated diffusion is associated with the presence of an M_xAu_y alloy in the subsurface region of metallic gold. The formation of such a coating dramatically alters the absorption spectrum of visible light by the surface layer[1] and alters the work function of the solid. In addition, the evaporation rate of Au decreases significantly in comparison to the rate for the pure metal.[3]

The interaction of a gold adsorbent with adsorbed alkali-metal atoms is known to lead to reconstruction of the surface of single-crystal gold and the appearance of M_xAu_y alloys in the subsurface region even for adsorbent temperatures $T < 300$ K.[4–6] The calculations in Ref. 7 showed that the alloys CsAu and RbAu should have semiconductor properties and that the compounds of Au with K, Na, and Li should have metallic properties if these compounds have a crystal structure similar to CsCl. The experiments in Ref. 8 showed that the alloy CsAu does, in fact, have semiconductor properties.

In Ref. 1 we attributed the appearance of the photostimulated diffusion of alkali-metal atoms in gold to features of the interatomic bond in M^+Au^- molecules, which is broken when light is absorbed. The question of the properties of the system which cause photostimulated diffusion, however, calls for additional study.

This paper presents the results of investigations of the changes in the optical properties of gold caused by the adsorption of Na on its surface and characterized by the appearance of photostimulated diffusion of Na atoms.

EXPERIMENTAL TECHNIQUES AND MEASUREMENT METHODS

The experimental techniques that we used to investigate photostimulated diffusion, the procedure of the preliminary preparation of the gold ribbon adsorbent, and the methods for forming an alloy on its surface were described in Ref. 2. The methods used to obtain the optical characteristics of the surface are considered in greater detail below.

The investigations were carried out in the energy range $2.0 < h\nu < 3.6$ eV. A beam of monochromatized light from high-pressure mercury and xenon arc lamps was directed toward the ribbon surface at the angle $\alpha = 74°$. The intensity of the light at the surface did not exceed 50 mW/cm^2.

The optical absorption coefficient A of the substrate was determined calorimetrically from the change in the resistance of long, thin ribbons caused by their illumination. A bridge setup, which makes it possible to detect relative changes in the resistance of a ribbon at the level of $\Delta R/R > 10^{-5}$, was employed for this purpose. By comparing the changes in the resistance of a ribbon caused by illumination in the case of a clean surface and after holding the ribbon in a flow of Na atoms, we could find the changes in A associated with the formation of NaAu in the subsurface region of Au if the light-irradiation intensities are identical. Known data on the spectral dependence of the absorption of pure metallic gold $A_{Au}(h\nu)$ (Ref. 9) were used.

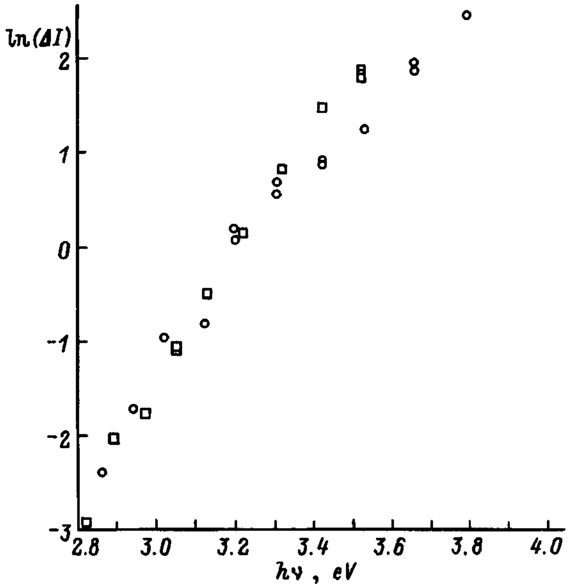

FIG. 1. Spectral dependence of the yield of Na^+ ions from a ribbon doped with sodium and then annealed for various annealing times. At $T = 1200$ K: □ — 2-h anneal, ○ — 120-h anneal. The absolute values of the currents have been shifted into coincidence.

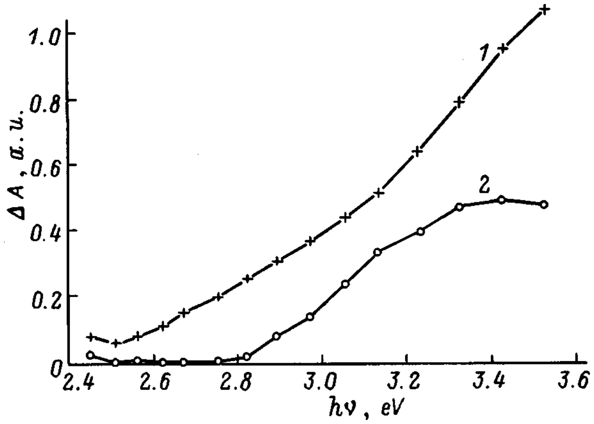

FIG. 2. Variation of the spectral dependence of the absorption of gold due to doping with sodium: *1* — 10-sec anneal, *2* — 120-h anneal at $T = 1200$ K.

The effects of light on sodium-doped gold for two methods of forming NaAu were compared. In one of them pure gold heated to $T = 700 - 800$ K was subjected to prolonged exposure to a flow of Na atoms with a flux density $i = 10^{13}$ $cm^{-2}s^{-1}$ at the surface. Then the ribbon was annealed at $T = 1150 - 1250$ K, and the characteristics of the interaction of light with the adsorbent were measured for different annealing times. The variation of the efficiency of the photostimulated diffusive emergence of Na atoms on the substrate surface was monitored. In the other method the optical characteristics of the surface and the effect of light on the diffusion of Na were obtained during the deposition of Na with $i = 10^9 - 10^{13}$ $cm^{-2}s^{-1}$. To determine the optical properties of the substrate, the flow of Na was stopped, and the ribbon was cooled to $T = 300$ K.

The efficiency of the photostimulation of diffusion was determined from the difference between the surface-ionization currents accompanying the thermal desorption of Na from the surface $\Delta I = I - I_b$, where I is the current of Na^+ ions when the surface is illuminated, and I_b is the dark emission current of these ions.

RESULTS AND DISCUSSION

Figure 1 shows the spectral dependence of the efficiency of the photostimulation of the diffusion of Na atoms for the case of a ribbon preliminarily doped with the alkali metal and then annealed. It is seen that the form of the $\Delta I(h\nu)$ curve does not vary with the annealing time and is determined by the presence of the alloy phase in the subsurface region of the substrate. A similar dependence of photostimulation is observed in the case of the deposition of Na on a heated gold surface, where the diffusive passage of Na atoms from the substrate surface into its subsurface region is stimulated.

The form of the $A(h\nu)$ curves varies with the extent of doping of Au by Na atoms, as was shown in Ref. 1, since the measured values of A are determined by the superposition of the spectral dependences of the absorption of metallic Au and the NaAu film on its surface. The thickness of this film depends on the intensity and duration of the deposition of Na and on the substrate temperature T. Figure 2 shows the difference between the absorption spectra $\Delta A(h\nu)$ of doped and pure gold (curve *1*), as well as Au preliminarily doped, but annealed and the pure metal (curve *2*). These curves reflect the absorption of light by the subsurface alloy. It can be seen that the saturation of gold with sodium increases the absorption of light and that raising the Na concentration leads to the appearance of an absorption band in the low-energy region of the spectrum at $h\nu < 2.8$ eV.

Since photostimulated diffusion is displayed when Au is doped with Na, K, and Cs and it is known that CsAu is a semiconductor, there is basis to theorize that NaAu also has a semiconductor structure. It is confirmed by the significant increase in the resistivity of the ribbon during formation of the alloy.[3] In this case we can use the approximation of the spectral dependence $\Delta A(h\nu)$ for a semiconductor with a band gap E_g, as was proposed in Ref. 10, by the expression

$$\Delta A(h\nu)\cdot(h\nu) = C(h\nu - E_g)^b, \tag{1}$$

where C is a coefficient and b is a parameter, which reflects the character of the optical transitions (for direct allowed transitions $b = 0.5$, for direct forbidden transitions $b = 1.5$, for indirect allowed transitions $b = 2.0$, and for indirect forbidden transitions $b = 3.0$).

After replotting the $\Delta A(h\nu)$ spectral dependence in $(\Delta A(h\nu))^{1/b}$ versus $h\nu$ coordinates and adjusting b so that this dependence would be linearized, we can find the value of E_g. In the range of $h\nu$ of interest to us, in the case of an annealed sample (curve *2* in Fig. 2) the plot of (1) is linearized in the range of photon energies $2.8 < h\nu < 3.5$ eV when $b = 2.0$ and $E_g = 2.8$ eV; in the case of a strongly doped substrate it is linearized for the photon energies $2.8 < h\nu < 3.6$ eV when $b = 1.5$ and $E_g = 2.7$ eV.

Since the rate of photoinduced diffusion depends linearly on the radiant flux density,[1] it can be assumed that

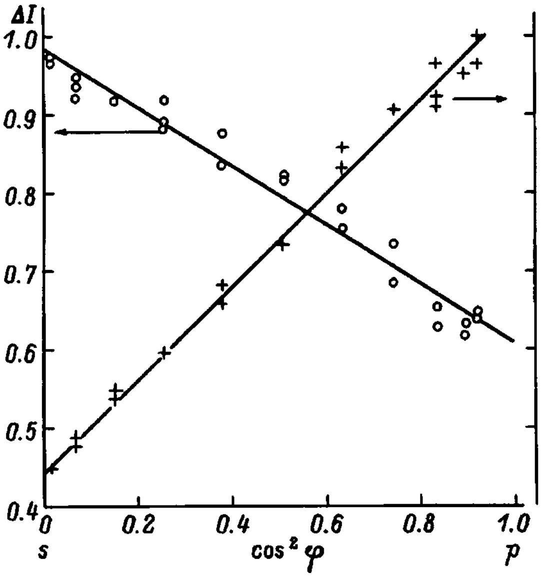

FIG. 3. Dependence of ΔI^{+} and A on $\cos^2(\varphi)$ for a photon energy $h\nu = 3.54$ eV.

stimulation of this process occurs as a result of the primary absorption of photons by the electronic structure of the solid. In this case the $\Delta I(h\nu)$ curve reflects the dependence of the absorption of light which causes photostimulated desorption, and the approximation considered above is capable of revealing the optical transitions responsible for stimulation of this process. In fact, an analysis for an annealed sample showed that $E_g = 2.8$ eV and that the values of the coefficients of the optical transitions amount to $b = 2$ in the range $2.9 < h\nu < 3.5$ eV. For a sample strongly doped by the prolonged deposition of Na we found $E_g = 2.7$ eV and $b = 1.5$, probably due to the alteration of the properties of the alloy in the presence of an excess quantity of Na. Thus, the optical absorption and photostimulated desorption spectra correspond to one another and show that photoexcitation of the bulk of the alloy NaAu, which has a semiconductor structure, is responsible for stimulating the diffusion process.

The fact that the light which stimulates the diffusion process is absorbed in the bulk of the alloy is confirmed by measurements of the dependence of the optical absorption coefficient and the efficiency of photostimulated diffusion on the polarization vector of the electric field of the light wave. As an example, Figure 3 shows plots of ΔI and A as functions of the orientation of the electric field vector of the light wave, which is specified by the angle φ between a normal to the ribbon surface and the plane of polarization of the electromagnetic radiation. The angle $\varphi = 0$ corresponds to p polarization of the light wave, and $\varphi = 90°$ corresponds to s polarization. Since light impinges on the surface at an angle $\alpha = 74°$ in the experimental setup, $\varphi = 0$ was not achieved in the experiments. It follows from Fig. 3 that in the case of gold doped with sodium and then annealed, s-polarized light most efficiently causes the emergence of Na on the surface from the bulk. The ratio between the efficiencies of the action of s- and p-polarized light $\Delta I_s/\Delta I_p = 1.5$. Conversely, the polarization dependence of the absorption coefficient A exhibits greater efficiency for the absorption of p-polarized light. For example, at $T = 300$ K we have $A_s/A_p = 0.40 - 0.45$ for the sample investigated. Raising the ribbon temperature to $T = 1150$ K gives $A_s/A_p = 0.30$. This value is equal to the ratio A_s/A_p calculated from the Fresnel formula for pure Au using the known value of the dielectric constant ε (from Ref. 9) for $\alpha = 74°$.

These results support the conclusion drawn in Ref. 11 on the basis of measurements of the kinetics of the diffusion Na in NaAu that a high-temperature anneal of Na-doped gold produces a narrow (1–2 lattice constants) region depleted of the alkali metal between the surface and the alloy layer, which conserves the alloy within the substrate. When T is lowered, Na is redistributed in the subsurface region with an increase in its concentration on the surface.

The lack of correspondence between the $\Delta I(\varphi)$ and $A(\varphi)$ curves can probably be explained in the following manner. Photostimulated diffusion is associated with the absorption of light in the bulk of the alloy, i.e., is determined by the superposition of the polarization dependences of the optical permeability of the subsurface region of the solid and the photoprocess. The $A(\varphi)$ curve is determined mainly by the reflection of light from the solid/vacuum interface. In the geometry used in this work for illuminating the object, in which light impinges on the plane of the surface at a glancing angle, s-polarized light is reflected from the surface to a greater degree than is p-polarized light, in good agreement with the measured values of the ratio A_s/A_p. It is possible that the light refracted in the substrate is subsequently reflected from the deep layers and that the value of ε of the medium varies with depth as a function of the Na concentration in the alloy. For example, it can be theorized that the ratio $\Delta I_s/\Delta I_p$ is determined by the features of the propagation of light from a medium with a higher optical density (gold depleted of sodium) to a medium with a lower density (NaAu) in the case of the glancing incidence of light onto the surface. The fact that the Na concentration on the surface significantly alters the optical parameters of the substrate is confirmed by the polarization dependences of the photostimulated diffusion of Na into the bulk of the alloy when Na is deposited on the Au surface. After prolonged holding of a ribbon at $T = 1150$ K in a flow with $i = 10^{12}$ cm^{-2}s^{-1}, we obtained $\Delta I_s/\Delta I_p = 0.80$. A comparison with the measured value of this ratio for an annealed sample reveals significant changes in the optical permeability of the alloy and the reflection coefficients for light of different polarization as the concentration of the alkali-metal atoms near the surface/vacuum interface varies.

Thus these investigations of the optical characteristics of the photostimulation of the diffusion of Na atoms in gold indicate that this process is caused by the growth of a NaAu film having semiconductor properties on the gold surface. The optical characteristics of the alloy vary with the concentration of Na in the gold.

This work was performed with the support of the Russian Fund for Fundamental Research (Project No. 96-02-16908) and the ''Surface Atomic Structures'' State Program of the Ministry of Science and Technology of the Russian Federation (Project No. 95-02.13).

[1] É. Ya. Zandberg, M. V. Knat'ko, V. I. Paleev *et al.*, Pis'ma Zh. Tekh. Fiz. **21**(19), 15 (1995) [Sov. Tech. Phys. Lett. **21**(10), 774 (1995)].
[2] M. V. Knat'ko, V. I. Paleev, and M. N. Lapushkin, Zh. Tekh. Fiz. **68**(10), 104 (1998) [Tech. Phys. **43**, 1231 (1998)].
[3] M. V. Knat'ko, V. I. Paleev, and E. Ya. Zandberg, Phys. Low-Dimens. Struct., Nos. 7–8, 27 (1996).
[4] J. V. Barth, R. J. Behm, and G. Ertl, Surf. Sci. **341**, 62 (1995).
[5] S. L. M. Schroeder, A. Neumann, T. Solomun *et al.*, Surf. Sci. **337**, 285 (1995).
[6] J. V. Barth, R. Schuster, J. Wintterlin *et al.*, Phys. Rev. B **51**, 4402 (1995).
[7] C. Koenig, N. E. Christensen, and J. Kollar, Phys. Rev. B **29**, 6481 (1984).
[8] W. E. Spicer, A. H. Sommer, and J. G. White, Phys. Rev. B **115**, 57 (1959).
[9] G. B. Irani, T. Huen, and F. Wooten, J. Opt. Soc. Am. **61**, 128 (1971); P. B. Johnson and R. W. Christy, Phys. Rev. B **6**, 4370 (1972).
[10] A. Hjelm, C. G. Granqvist, and J. M. Wills, Phys. Rev. B **54**, 2436 (1996).
[11] M. V. Knat'ko, V. I. Paleev, and M. N. Lapushkin, Phys. Low-Dimens. Struct., Nos. 5–6, 85 (1998).

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Determination of the dynamic characteristics of aerogels in the energy-release zone of a high-power electron beam

B. A. Demidov, V. P. Efremov, M. V. Ivkin, I. A. Ivonin, V. A. Petrov, and V. E. Fortov

Kurchatov Institute Russian Science Center, 123182 Moscow, Russia

(Submitted June 3, 1997)
Zh. Tekh. Fiz. **68**, 112–120 (October 1998)

The dynamics of the interaction of a high-current electron beam with SiO_2 aerogels of different density and porosity are investigated by optical methods. A model for describing highly porous materials is developed on the basis of the information obtained regarding the unloading of aerogels in the energy-release zone of an electron beam, as well as the measured energy-release profiles. A corresponding nonlinear self-consistent equation of state is obtained which reflects the fractal properties of aerogels and permits determination of the thermodynamic characteristics of aerogels as the porosity varies tens of times. The influence of electric space charge on the energy-absorption profile of a high-power electron beam in aerogels of different density is discussed. © *1998 American Institute of Physics.* [S1063-7842(98)02010-8]

INTRODUCTION

Porous condensed media are promising materials for damping short-lived single shock loads by converting the translational energy into heat. Therefore, the study of the action of intense energy fluxes on porous materials is of current interest. Such investigations would enable us to determine the thermodynamic properties of a porous medium and to predict the response of the material to intensive dynamic loads. Silica aerogels obtained by supercritical thermal extrusion of colloidal solutions of silicic acid are unique porous materials. They are transparent insulators, which have a cluster structure, a low density, and a small charge number Z. The interest in aerogels as a new class of materials has grown dramatically in the last few years.[1] This is due both to refinement of the technology for preparing them[2] and new possibilities for employing them.

Silica aerogels have been used to study the thermodynamics of a weakly ionized plasma obtained as a result of shock-wave compression[3] and the influence of electric space charge on the formation of the energy-release zone of a pulsed electron beam.[4] In optics aerogels have been used in spectroscopic investigations of large organic molecules embedded in an aerogel matrix.[5] The behavior of porous materials has previously been described using a non-self-consistent equation of state, which takes into account the internal structure of the material.[6]

The high transparency of aerogels permits the use of optical methods to experimentally investigate the penetration and bulk absorption of electromagnetic and corpuscular radiation. The experimental conditions are simplified if an electron beam, which has a fairly large penetration depth and accurately predictable ranges for electrons in materials, is used. In this case the energy-release profile of the electron beam can be determined, time-resolved information regarding the unloading of the aerogel in the energy-release zone can be obtained, and the equation of state of the aerogel can be investigated in portions of the phase diagram that are inaccessible in the case of ordinary shock loading.

The purpose of the present work is to perform experiments on the interaction of a pulsed high-current electron beam with aerogels, whose density varies over a broad range from 0.36 to 0.03 g/cm^3 and whose porosity increases from 6 to 80, and to devise a correct model of the equation of state of this class of materials, which are distinguished by an extremely low density and a complicated internal structure, on the basis of the experimental data obtained. The conventional models for describing porous media do not take into account the variation of the internal structure of aerogels in response to variation of the external conditions. A nonlinear regime appears under practically any load. Therefore, the development of a nonlinear self-consistent equation of state for aerogels, which reflects their fractal properties, is needed.

Another purpose of the present work is to investigate the induced high-energy conductivity of aerogels, which appears when they are irradiated by pulsed electron flows. The performance of such an investigation is important not only for solving practical problems, but also for understanding the nature of the relationship between the internal structure of aerogels and their dielectric properties. A theoretical investigation of the influence of electric space charge on the formation of the energy-release zone of electron beams was studied numerically in the constant-conductivity approximation in Refs. 7 and 8. The theory of high-energy induced conductivity developed in Ref. 9 does not describe prebreakdown regimes. The weak influence of electric space charge on the formation of the energy-release zone of a pulsed electron beam with a current density $J_0 > 10^3$ A/cm^2 in an aerogel with $\rho_0 = 0.36$ g/cm^3 was noted in Ref. 4. However, the situation can change when the density of the aerogel is diminished, and additional investigations are required.

DESCRIPTION OF THE EXPERIMENT

The experimental setup for investigating the dynamics of the interaction of a pulsed high-current electron beam (the

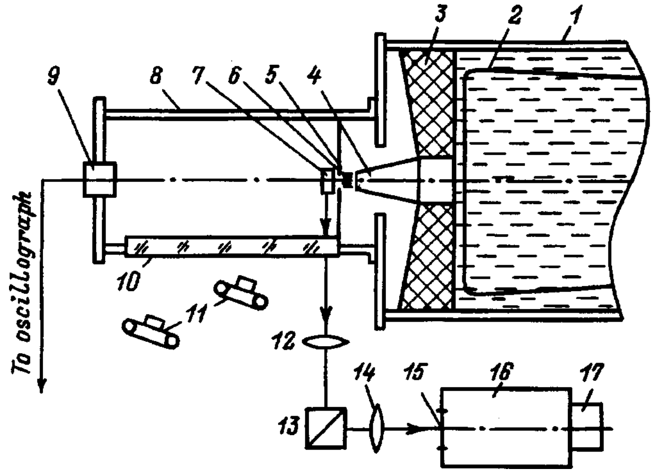

FIG. 1. Experimental setup.

electron energy $U_0 = 270$ keV, the current $I = 10$ kA, and the pulse duration at half maximum $\tau = 150$ ns) with SiO_2 aerogels having densities of 0.03, 0.14, and 0.36 g/cm^3 is shown in Fig. 1. The outlet portion of the Kal'mar accelerator,[10] which includes housing *1,* output transformer *2,* insulating separating diaphragm *3,* and cathode *4,* was connected to vacuum chamber *8* and control detector *9.* Electron beam *6* passed through aluminum foil *5* of thickness 10 μm and impinged on the aerogel sample under investigation *7.* Irradiation of the aerogel sample by the electron beam produced a flash of light, which passed through transparent acrylic (PMMA) plate *10* and lens *12* and impinged on prism system *13*, which rotated the image of the energy-release zone of the electron beam by 90° in the horizontal and vertical planes. The image was focused by lens *14* onto vertical inlet slit *15* of photochronometric unit *16, 17,* a FÉR-7 streak camera affording a high linearity of the transmission of the aerogel luminescence. For the most part a 0.25 μs/cm sweep with an inlet slit thickness of 0.2 mm was used in the experiments. A contact detector, which measures the time of flight of the aerogel fragments (particles) from the rear side of the target, was placed in the end wall of the vacuum chamber. Two cameras *11* with different angles of sight photographed the dispersal pattern of the aerogel fragments (particles) in vacuum chamber *8* with a length of 0.35 m in an integral regime. Aerogel samples measuring 20×20 mm with polished walls and a thickness which, for the most part, exceeded the presumed range of electrons with an energy of 270 keV in the aerogels studied were investigated. An electron beam with a diameter of about 10 mm struck on the aerogel sample, which was placed on a grounded metallic substrate. The vacuum gap between the aluminum foil *5* and the aerogel sample *7* measured 3 or 5 mm for aerogels of different density. The aerogel luminescence within the energy-release zone is characterized by a weak intensity. Therefore, in order to determine the depth of the energy-release zone more exactly, the aerogel samples were placed tightly against the aluminum foil and a 0.75 μs/cm sweep was used in separate experiments.

The system consisting of lenses *12* and *14* provided for magnification (reduction) in the optical system which would cause the image of the energy-release zone and the vacuum gap between the aerogel and the aluminum foil to completely fill the height of the FÉR-7 entrance slit, which measured 10 mm. The optical spatial resolution, which was verified by a marker placed in the region of the aerogel, permitted determination of the depth of the energy-release zone to within 5%. The photochronograms (streak photographs) obtained in the experiments were processed on an IFO-451 microphotometer. The dependence of the blackening of the photographic film on the exposure time was determined using an optical wedge. Because of the restricted number of aerogel samples investigated, PMMA targets of similar dimensions were used to adjust the optical system and to synchronize the triggering of the FÉR-7 unit.

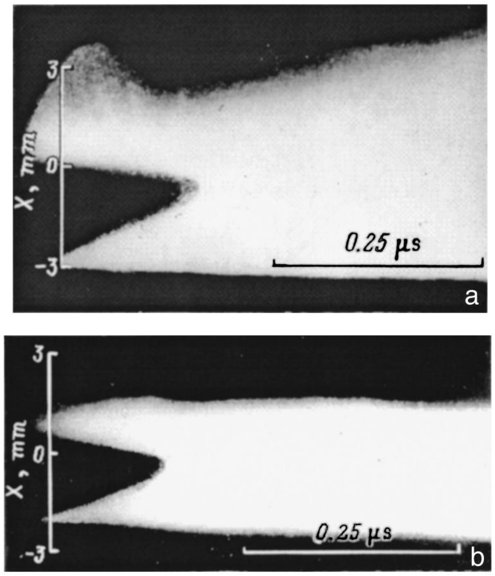

FIG. 2. Photochronograms (streak photographs) of the interaction of the electron beam with an aerogel (a) and with PMMA (b).

EXPERIMENTAL RESULTS

A typical photochronogram, which illustrates the dynamics of the interaction of a high-current electron beam with an aerogel is shown in Fig. 2a. In this case the current density $J_0 = 10$ kA/cm^2, the electron energy $U_0 = 270$ keV, and the aerogel density $\rho_0 = 0.14$ g/cm^3. It can be seen in Fig. 2a that the depth of the energy-release zone increases with time, reaching 3 mm in the region of the maximum accelerating voltage. The dispersal velocity of the aluminum foil is 9 km/s, and the dispersal velocity of the aerogel counter to the direction of the electron beam is 3.3 km/s. For comparison Fig. 2b presents the analogous photochronogram obtained under the same conditions but with a PMMA target. In this case the depth of the energy-release zone is less than 1 mm, and the dispersal velocity of PMMA equals 4.2 km/s.

The depth of the energy-release zone of the electron beam was determined more accurately in experiments in which the aerogel sample was placed on an aluminum foil, and slower sweeping by the FÉR-7 unit was used. Figure 3

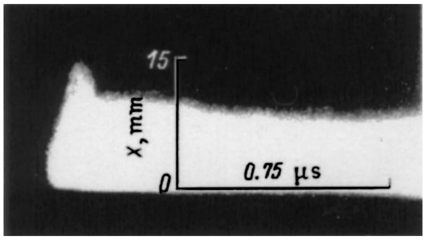

FIG. 3. Photochronogram of the interaction of the electron beam with an aerogel. The aerogel sample is right on the aluminum foil.

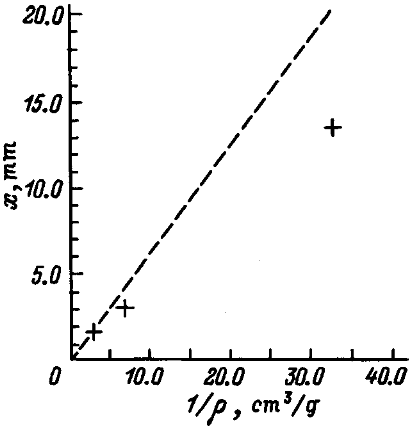

FIG. 5. Dependence of the depth of the energy-release zone of the electron beam on the density of the aerogel: crosses — experimental results.

shows the photochronogram of the interaction of an electron beam ($J_0 = 10$ kA/cm^2, $U_0 = 270$ keV) with the aerogel having a density $\rho_0 = 0.03$ g/cm^3.

Figure 4 presents the results of the microphotometric measurements of the photochronogram in Fig. 3, which show the dependence of the aerogel luminescence intensity on the distance into the interior of the target. It follows from the data presented that the depth of the energy-release zone approaches 15 mm in the aerogel with $\rho_0 = 0.03$ g/cm^3.

Figures 5 and 6 show the results of the processing of the photochronograms of the aerogels with densities of 0.03, 0.14, and 0.36 g/cm^3. Figure 5 illustrates the dependence of the depth of the energy-release zone of the electron beam on the density of the aerogel. In this case, as was shown in Ref. 4, it was assumed that the aerogel luminescence intensity in the interior of the target is proportional to the energy released by the electron beam. The dashed line is the result of a calculation without consideration of the space-charge effects.

In Fig. 6 the crosses show the dispersal velocities of aerogels of various density counter to the electron beam in a regime with $J_0 = 10$ kA/cm^2 and $U_0 = 270$ keV. A dispersal velocity of 7 km/s was previously obtained in Ref. 4 for an aerogel with $\rho_0 = 0.36$ g/cm^3 and an electron beam with the parameters $J_0 = 14$ kA/cm^2 and $U_0 = 290$ keV. The circles are the results of numerical simulation, and the details of the models are described below.

Figure 7 shows photographs of the dispersal of particles from the rear side of the aerogel with $\rho_0 = 0.03$ g/cm^3 (in order to observe the flight of particles a hole of diameter 10 mm was made in the metallic substrate: a — general dispersal pattern of aerogel fragments in the vacuum chamber, b — dispersal of particles in the immediate vicinity of the aerogel. Both photographs were taken during a single experiment. The trajectories of the particles and their reflection from the chamber walls are clearly seen. Some of the trajectories display modulation of the luminescence brightness along their course, and some trajectories have an erratic form. Figure 7b shows an aerogel target measuring $\simeq 20 \times 15$ mm (the thickness of the target is 15 mm).

The negatives of the photographs shown in Fig. 7 have a large depth of focus and good acutance, since they were obtained by objective lenses with relative apertures of 1:22. This permitted estimation of the dimensions of the dispersing particles from the photographs, which lie in the range $0.1 - 1.0$ mm. Attempts to directly measure the dispersal velocity of fragments were unsuccessful, since the particles did not pierce the outer layer of the time-of-flight detector, which consisted of a 10-μm aluminum foil and a 10-μm insulator, in these experiments.

DISCUSSION OF EXPERIMENTAL RESULTS

It was shown in Ref. 4 that, with allowance for the diffraction effects, the plot of the aerogel luminescence intensity as a function of the distance into the interior of the target coincides well with the calculated energy-release profile of the electron beam. Therefore, it can be assumed that the curve in Fig. 4 describes the energy-release profile of the

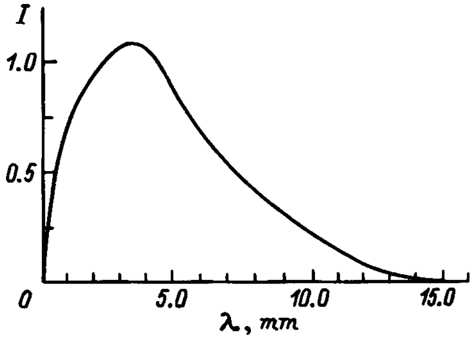

FIG. 4. Dependence of the luminescence intensity of an aerogel on the distance into the interior of the target.

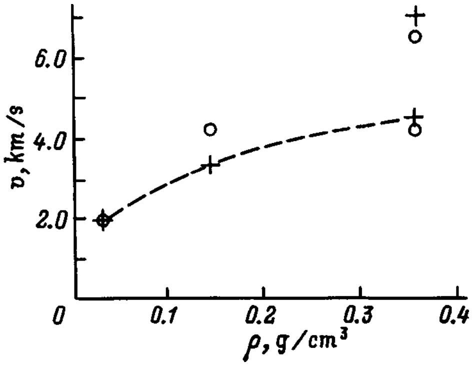

FIG. 6. Dependence on the dispersal velocity of an aerogel counter to the electron beam on the density of the aerogel.

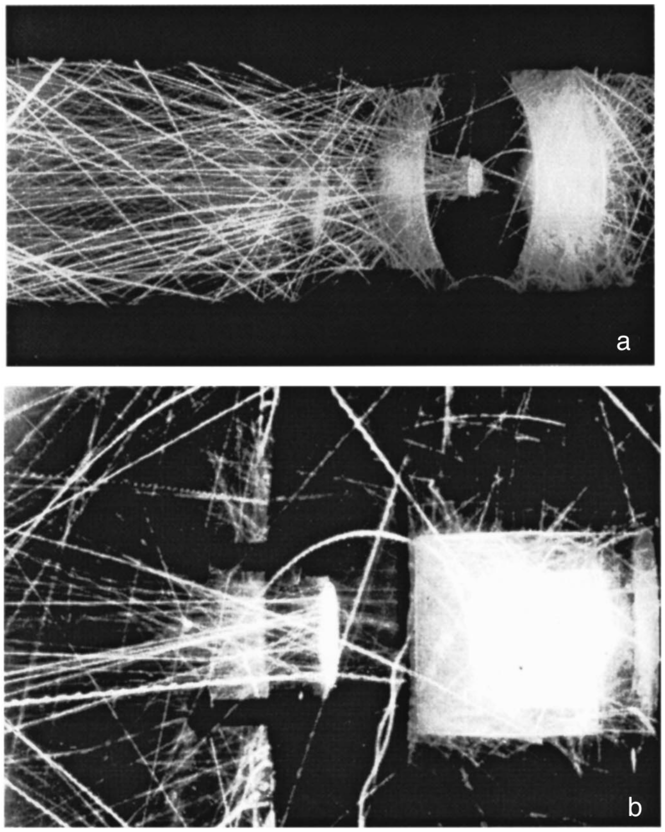

FIG. 7. Dispersal of aerogel particles from the rear side of the sample.

electron beam in the aerogel with a density of 0.03 g/cm^3. The smoother growth of the luminescence intensity (Fig. 4) in comparison to the calculated absorption profile of the energy of a beam of monoelectrons in matter is probably attributable to the bulk unloading of the aerogel during irradiation.

It should, however, be noted that the measured value of 14 mm for the depth of the energy-release zone differs significantly from the calculated value (which equals 20 mm). For the aerogel with $\rho_0 = 0.14$ g/cm^3 the difference between the measured value of 3 mm and the calculated value of 4.2 mm is also significant.

However, the difference for the aerogel with $\rho_0 = 0.36$ g/cm^3 is insignificant. Figure 5 presents calculated values of the depth of the energy-release zone without consideration of the influence of the electric space charge for aerogels of various density (dashed line). A comparison of the experimental and theoretical values reveals that the agreement is unsatisfactory for the aerogels with a low density $\rho_0 < 0.2$ g/cm^3. We attribute this to the space-charge effects. The potential of the electric field in the energy-release zone becomes comparable to the initial energy of the electrons in the beam. Consideration of this effect leads to agreement between the calculated and experimental estimates of the depth of the energy-release zone.

It can be seen from Fig. 6 that the measured dispersal velocities of the aerogels counter to the electron beam (crosses) increase nonmonotonically with the density. This is due to two counteractive effects. On the one hand, as the porosity of the aerogel increases, the effective Grüneisen coefficient and the dynamic stiffness decrease, causing a drop in the dispersal velocity. On the other hand, as the porosity increases, the depth of the energy-release zone increases, and unloading of the aerogel does not manage to occur during irradiation. This situation leads to a rapid rise in pressure and an increase in the dispersal velocity. This hypothesis is supported by the hydrodynamic calculations performed. The results of the calculations are indicated by circles in Fig. 6; satisfactory agreement between the experimental and theoretical results can be noted.

Let us examine the photographs of the dispersal of aerogel particles from the rear side of a target shown in Fig. 7. The trajectories of the particles are very bright and clear. Do the particles glow themselves or are they illuminated by the plasma in the cathode–anode gap of the diode chamber of the accelerator? It is known that the electron-acceleration phase in accelerators of high-current electron beams, which lasts $\tau \simeq 10^{-7}$ s, is followed by the onset of a so-called ''short-circuiting'' regime, which is characterized by an oscillating current of large amplitude and a very small accelerating voltage (the accelerator operates across an inductive load). In this regime the luminescence intensity of the diode-chamber plasma varies in time with twice the frequency of the current. In this case the modulation of the particle trajectories can be attributed to illumination by the plasma, and the velocity of the particles can be estimated using the stroboscopic effect. Under this assumption it follows from Fig. 7 that the velocity of individual aerogel particles reaches 10 km/s. However, such particles should break through the outer layer of the contact detector and activate it. Estimates of the voltages appearing in the aluminum layer show that the dispersal velocity of particles from the aerogel with $\rho_0 = 0.03$ g/cm^3 do not exceed 4 km/s. These estimates coincide with the energy estimates with consideration of the small thickness of the outer layer of the contact detector in comparison to the characteristic dimensions of the aerogel particles.

On the other hand, the time of flight of the particles along the vacuum chamber should exceed 35 μs. Additional experiments performed with the aid of a photoelectron multiplier and the FÉR-7 unit showed that the luminescence intensity of the plasma is weak, that the luminescence time is less than 30 μs, and that the short-circuit current decay time does not exceed 25 μs. This means that the illumination of the aerogel particles by the plasma is insignificant. Therefore, the particles themselves clearly glow due to the high temperature and their luminescence during their entire flight along the vacuum chamber. The modulation of the trajectory luminescence is attributed to rotation of the particles about their own axes during motion. Some particles may have the shape of a dumbbell, which leads to the characteristic form of the trajectories of such particles.

NUMERICAL SIMULATION

In order to determine the dynamic characteristics of the aerogels and to compare them with experiment, we performed numerical calculations of the unloading of the aerogels during irradiation by the electron beam of the Kal'mar accelerator. The specific energy input from the electron beam as a function of the Lagrangian coordinate of the absorption depth and the irradiation time was calculated in the diffusion approximation[11] with consideration of real current and voltage oscillograms of the Kal'mar accelerator and the dependence of the depth range of electrons on their energy from Ref. 12. To obtain additional information on the accelerator operating regime and determine the influence of the aluminum foil (10 μm) placed in front of the aerogel (Fig. 1) on the release of energy, we performed a numerical simulation of the unloading of PMMA with a definitely known equation of state.[13] The unloading was calculated using a second-order one-dimensional Lagrangian ''predictor-corrector'' scheme. The calculations for the irradiation of PMMA in a regime with a peak electron beam energy $U_0 = 270$ keV, a peak current density $J_0 = 10$ kA/cm^2, and a current pulse duration at half maximum $t_b = 150$ ns, showed that the dispersal velocities of the foil and PMMA were 7.2 and 4.6 km/s, respectively. These velocities are perfectly consistent with the experimentally measured values of 8–9 km/s for the dispersal velocity of the foil and 4.2 km/s for PMMA, attesting to the faithfulness of the description of the accelerator operating regime. In addition, it was confirmed numerically that a thin Al foil with a thickness equal to less than 5% of the mass energy-release depth of 270-keV electrons has virtually no influence on the unloading of PMMA. This allowed us to neglect the presence of the foil in the numerical simulation of the unloading of the aerogels.

It was shown in Ref. 14 that electric space charge does not have any influence on the formation of the energy-release zone of a pulsed electron beam in the aerogel with an initial density $\rho_0 = 0.36$ g/cm^3. This was substantiated, first, by the high mobility of high-energy electrons[9] and, second, by the low strength of the breakdown fields in highly porous materials. In fact, as was shown in Ref. 9, for energies E of high-energy electrons lying in the range $\{\hbar\omega_0, T\} \ll E \ll E_g$ ($\hbar\omega_0$, T, and E_g are the energy of the optical phonons, the temperature, and the gap width of the insulator, respectively), there is an energy-loss minimum, which is associated with the low probability of the interaction of an electron with optical phonons and the impossibility of the generation of electron–hole pairs. Just these electrons provide for high-energy breakdown and leakage of the excess charge from an insulator irradiated by a pulsed electron beam. Following Ref. 9, we can estimate the value of this breakdown field E_{br} as

$$\begin{aligned} E_{\mathrm{br}} &\simeq 2\pi \frac{\hbar\omega_0}{E_g} \frac{\omega_0}{e\mu} (m_e \hbar\omega_0)^{1/2} / (1+\Pi)^{1/3} \\ &\simeq 200/(1+\Pi)^{1/3}\,[\mathrm{kV/cm}], \end{aligned} \tag{1}$$

when $\hbar\omega_0 / E_g \simeq \mu^2$, $E_g \simeq 15$ eV, and $\mu \simeq 0.1$.

In this estimate, which was obtained on the basis of the condition that there is no cooling of high-energy electrons in the electric field E_{br}, μ is a small dimensionless parameter, and the multiplier $1/(1+\Pi)^{1/3}$ is a correction to the porosity Π of the aerogel. It can be seen from (1) that the value of E_{br} for any porosity Π exceeds $\simeq 200$ kV/cm. According to Ref. 12, the specific depth R_0 for the release of the energy of an accelerator electron beam with a peak energy $U_0 = 290$ keV in an aerogel is $R_0 \simeq 0.07$ g/cm^2. Therefore, the maximum change in the energy of an electron in the energy-release zone in the aerogel with $\rho_0 = 0.36$ g/cm^2 ($\Pi_0 \simeq 6$) does not exceed 20 keV, which is an order of magnitude smaller than U_0. For the aerogel with $\rho_0 = 0.14$ g/cm^3 ($\Pi_0 \simeq 16$) the energy change amounts to no more than 40 keV, which also weakly influences the formation of the energy-release zone of an electron beam with $U_0 = 270$ keV. The situation is different for the aerogel with $\rho_0 = 0.03$ g/cm^3 ($\Pi_0 \simeq 80$). In this case, despite the weakness of the breakdown field (1), $E_{\mathrm{br}} \simeq 50$ kV/cm, the electrostatic energy of an electron in the energy-release zone can be fairly high, $\simeq 100$ keV. This leads to considerable distortion of the dimensions and shape of the energy-release zone in experiments employing an electron beam with $U_0 = 270$ keV, $J_0 = 10$ kA/cm^2, and $t_b = 150$ ns in comparison to the values calculated using the formulas from Ref. 11. For comparison, Fig. 8 shows plots of the specific energy input from the electron beam as a function of the Lagrangian coordinate of the penetration depth with and without consideration of the internal electric fields. As can be seen, when electrification is taken into account, the penetration depth of the beam electrons decreases from 2 to 1.5 cm and coincides with the experimentally obtained depth of the aerogel luminescence zone (Fig. 4). When electrification is taken into account, the coordinates $x_{\max} \simeq 0.4$ cm of the maxima of the luminescence and energy-release zones also coincide. Such correspondence attests to the correctness of the estimates of the electrostatic fields induced in the energy-release zone of the aerogel and of their influence on the shaping of the energy-release zone in the calculations. The differences between the forms of the luminescence (Fig. 4) and energy-release (Fig. 2) curves at the beginning of the energy-release zone are due to the unloading of the aerogel during irradiation. For the further discussion it must also be noted that the mechanical pressures caused by electrification of the aerogel are extremely small ($\leqslant 0.1$ atm) and were disregarded in the present calculations.

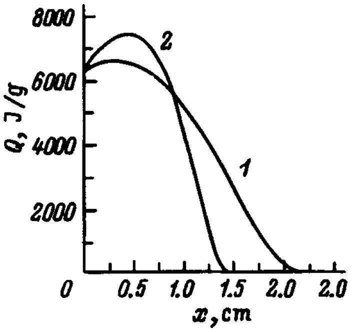

FIG. 8. Comparison of the integral specific energy input from the electron beam in the aerogel with an initial density $\rho_0 = 0.03$ g/cm^3: *1* — energy release without consideration of the internal electric fields; *2* — energy release with consideration of the internal electric fields.

The mechanical behavior of the aerogels was described using the plastic differential equation of state $P(\rho, Q, \beta)$ of a porous medium with self-consistent variation of the Grüneisen coefficient $\langle \Gamma \rangle$ and the bulk modulus $\langle K \rangle$ as functions of the porosity[4,6,14]

TABLE I.

Peak beam energer, keV	Peak current density, kA/cm^2	Initial density, g/cm^3	Aerogel dispersal velocity, km/s exp.	Aerogel dispersal velocity, km/s calc.	Displacement of aerogel surface at the end of the pulse, mm
270	10	0.03	2	2.0	0.15
270	10	0.14	3.3	4.2	0.36
270	10	0.36	4.5	4.2	0.2
290	14	0.36	7−8	6.5	0.31

$$dP=\langle K\rangle \frac{d\rho}{\rho}+\rho\langle\Gamma\rangle\, dQ,$$

$$dP_i=K_i\frac{d\rho_i}{\rho_i}+\rho_i\Gamma_i\, dQ,\quad \Gamma_i=\Gamma_i(\rho_i),\quad K_i=K_i(\rho_i),$$

$$dP=dP_i,$$

$$\langle K\rangle=K_i\beta^\gamma,\quad \langle\Gamma\rangle=\Gamma_i\beta^{\gamma-1},\quad \beta\equiv\frac{\rho}{\rho_i},\quad \Pi\equiv\frac{1-\beta}{\beta}. \qquad (2)$$

The first and second equations in this system are the Mie–Grüneisen differential equations of state of an aerogel and its SiO$_2$ skeleton (i). The third equation is the condition of thermodynamic equalization of pressure on scales of the order of the pore size. This equation reconciles the quasistationary variation of the porosity of the aerogel with the variation of its density and internal energy. The condition of such quasistationary equalization is perfectly natural for a finely porous material such as a transparent aerogel and holds during characteristic times $\leqslant 1$ ns. Finally, the last equations in system (2) are the percolation-cluster relationships between the characteristics of an aerogel and the characteristics of its skeleton. The power-law dependence of the elastic characteristics of porous materials on their porosity, which follows from the percolation-cluster model,[6,14] has been confirmed experimentally for many classes of highly porous materials. For example, numerous experimental data on the speed of sound in porous bodies (Al, Fe, Cu, W, polystyrene foam, and polyurethane foam) were assembled in Ref. 4. It was concluded as a result that the speed of sound has almost a power-law dependence on ρ/ρ_i and that the exponent is essentially identical for all the materials considered. In Refs. 15 and 16 it was shown, in addition, that the shock adiabat of porous bodies does not depend on the size of the structural element (the pore size varied by an order of magnitude). All this indicates the satisfactory applicability of the percolation model (2) to the description of porous bodies.

The percolation coefficient γ depends on many factors, but primarily on the type of interaction between the atoms.[6,14] For example, for an isotropic interaction manifested by a small shear modulus $\gamma=1.7$. For a central interaction with the maximum possible shear modulus $\langle G\rangle=3/4\langle K\rangle$ we have $\gamma=4.4$. The crossover from an isotropic to a central interaction was investigated in Ref. 6, and it was shown that the percolation coefficient does not exceed the range $1.7-4.4$.

It is noteworthy that the percolation coefficient γ is the only free coefficient in the model (2). All the other coefficients refer to the skeleton of the porous material and are generally unknown. In our case the following parameters of ordinary glass were used in the numerical experiments in the model (1):[17] $\rho_i^0=2.4$ g/cm^3, $\Gamma_i^0=2.5$, $C_{S_i}^0=2.0$ km/s. Since the glassy skeleton softens and the shear modulus tends to zero when the temperature is raised to only 400°C, it was natural to try taking the percolation coefficient for the isotropic interaction model, i.e., $\gamma=1.7$. In fact, the numerical experiments showed the best agreement with the experimental results for just this choice. The rapid softening of the aerogel skeleton also justifies the applicability of the hydrodynamic model used.

In our experiments the internal energy of the aerogels (i.e., the specific energy input minus the work of the pressure forces) did not exceed the heat of sublimation of its glassy skeleton (Fig. 8). Therefore, there was no need to ‘‘match’’ the system (2) to the equation of state of the gaseous phase in the calculations, although doing so does not present any fundamental difficulties.

A proportional dependence of the differential bulk modulus and an inversely proportional dependence of the Grüneisen coefficient on the density of the aerogel skeleton were chosen in the numerical experiments. The relation $K_i(\rho_i)\propto\rho_i$ ensures constancy of the speed of sound in the aerogel at large compressions. The relation $\Gamma_i(\rho_i)\propto 1/\rho_i$ usually holds for metals and an extensive class of polymer and crystalline materials. These relations are not fundamental for the model (2) under consideration and can easily be replaced by more correct relations.

A numerical calculation of the unloading of the aerogels with the initial densities $\rho_0=0.03$, 0.14, and 0.36 g/cm^3 was performed for an electron accelerator operating regime with the voltage and current density amplitudes $U_0=270$ kV and $J_0=10$ kA/cm^2, respectively. The irradiation time was $t_b=150$ ns at half maximum of the current pulse. The results of the numerical calculations of the dispersal velocities of the aerogels and a comparison of with the experimental results are presented in Table I. To complete the picture, the table also presents the values of the dispersal velocities of aerogels with an initial density $\rho_0=0.36$ g/cm^3 obtained in Ref. 4 for

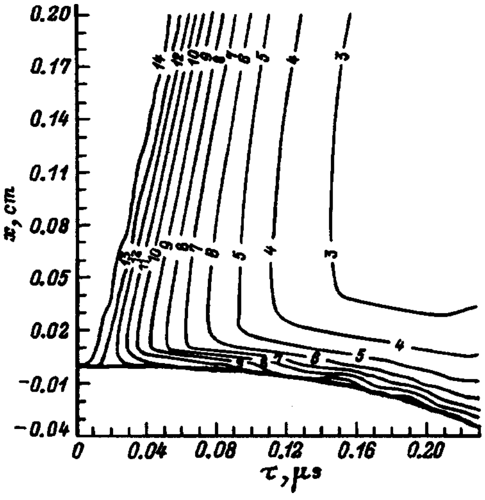

FIG. 9. Isolines of the porosity $\Pi(x,t)$ for the unloading of an aerogel with $\rho_0=0.14$ g/cm^3.

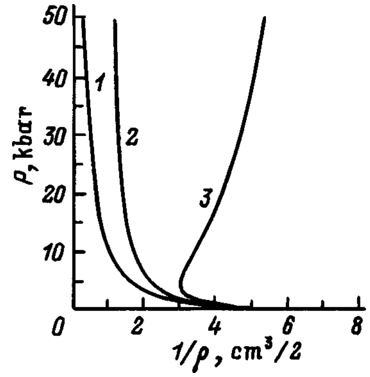

FIG. 10. Cold compression curve (*1*), adiabat (*2*), and Hugoniot shock adiabat (*3*).

an accelerator operating regime with the voltage and current density amplitudes $U_0=290$ kV and $J_0=10$ and 14 kA/cm^2, respectively. The results of a comparison of the calculated and experimental dispersal velocities of the aerogels in Table I attest to the satisfactory suitability of the percolation-cluster equation of state (2) for describing highly porous materials with initial porosities differing by more than an order of magnitude. The nonmonotonic dependence of the displacement of the surface (see Table I) of the irradiated aerogels on their initial porosity is noteworthy. In fact, in the samples with the maximum initial density ($\rho_0=0.36$ g/cm^3) the linear depth of the energy-release zone is minimal ($\simeq 1.7$ mm), and the central energy-release region begins to effectively unload long before the end of energy release. This ensures smaller peak compression of the samples with $\rho_0=0.36$ g/cm^3 in comparison to the second sample, which is approximately two times greater in size than the energy-release zone.

To illustrate the need for self-consistent allowance for the variation of the porosity during the loading of highly porous materials, Fig. 9 presents isolines of the porosity $\Pi(x,t)$ of the aerogel with $\rho_0=0.14$ g/cm^3, which correspond to the numerical experiment. It is seen that the porosity of the aerogel undergoes a nearly sevenfold decrease, which corresponds to an $\simeq 15$-fold increase in the bulk modulus, by the end of the irradiation period. Accounting for such strong nonlinear effects is difficult in the conventional models for describing porous materials, but is natural in the proposed model (2).

To conclude this section it should be noted that, despite the differential character of the equation of state (2) of a highly porous material, integral estimates can be obtained using (2). For example, if $K_i(\rho_i)\propto\text{const}$ and $\Gamma_i(\rho_i)\propto 1/\rho_i$, it is easy to obtain the integral Hugoniot shock adiabat. In fact, substituting the internal energy

$$dQ=-d\left\{\frac{P}{2}\left(\frac{1}{\rho}-\frac{1}{\rho_0}\right)\right\}$$

into (2), we obtain a system of two algebraic equations, which specify the variation of the density and the porosity of a material on a shock-wave front of amplitude P_H:

$$\begin{gathered}P_H=K_i/\gamma\ln\frac{1-\beta_0^\gamma}{1-\beta^\gamma},\quad \{\rho_0,\Pi_0\}\to\{\rho,\Pi\},\\ \frac{1}{\gamma}\ln\left\{\frac{\beta^\gamma(1-\beta_0^\gamma)}{\beta^\gamma(1-\beta^\gamma)}\right\}=\ln\left\{\frac{\rho}{\rho_0}\right\}-\frac{P_H}{2}\left\{\frac{1}{\rho}-\frac{1}{\rho_0}\right\}\frac{\rho_i\Gamma_i}{K_i},\\ \Pi\equiv\frac{1-\beta}{\beta}.\end{gathered}\tag{3}$$

Substitution of the adiabatic increment of the internal energy $dQ=-Pd(1/\rho)$ into (2) leads to a differential equation, which relates the density and porosity increments on the P_a adiabat:

$$\begin{gathered}P_a=K_i/\gamma\ln\frac{1-\beta_0^\gamma}{1-\beta^\gamma},\\ \frac{d\rho}{d\beta}=\frac{\rho^2}{\beta(1-\beta^\gamma)}\Bigg/\left\{\rho+\frac{\rho_i\Gamma_i}{K_i}P_a(\beta)\right\},\\ \Pi\equiv\frac{1-\beta}{\beta}.\end{gathered}\tag{4}$$

Finally, a zero increment of the internal energy ($dQ=0$), which corresponds to the curve for cold compression P_x, also permits explicit integration of the system (2):

$$\begin{gathered}P_x=K_i/\gamma\ln\frac{1-\beta_0^\gamma}{1-\beta^\gamma},\\ \frac{\rho^\gamma}{\rho_0^\gamma}=\frac{\beta^\gamma(1-\beta_0^\gamma)}{\beta_0^\gamma(1-\beta^\gamma)},\quad \Pi\equiv\frac{1-\beta}{\beta}.\end{gathered}\tag{5}$$

In Fig. 10 curves *1*–*3* correspond to the cold compression curve (5), the adiabat (4), and the Hugoniot shock adiabat (3) in the coordinates $\{1/\rho,P\}$ for the initial aerogel density $\rho_0=0.14$ g/cm^3.

It can be seen that, as expected, these adiabats coincide at low pressures. However, at $P\simeq 5$ kbar the Hugoniot adiabat becomes anomalous due to the high porosity. It is noteworthy that the anomalous part of the adiabat asymptotically emerges from the point with a specific volume equal to 2.5 cm^3/g. This value is $\simeq 6$ times greater than the specific volume of the SiO_2 skeleton, which is used for these purposes in the ordinarily employed model of porous materials.[18]

CONCLUSION

We have investigated by optical methods the dynamics of the interaction of a high-current electron beam with SiO_2 aerogels of various density from 0.03 to 0.36 g/cm^3. We have studied the evolution of the luminescence of the energy-release zone, as well as the dispersal of the aerogel counter to the electron beam. On the basis of the experimental data we have developed a model for describing highly porous materials. We obtained a corresponding nonlinear self-consistent equation of state that reflects the properties of aerogels. This equation of state permits the correct determination of the thermodynamic characteristics of aerogels as the porosity is varied by tens of times. The percolation coefficient of the aerogels $\gamma = 1.7$, which indicates that the aerogels belong to the class of materials with a predominantly isotropic interaction, has been determined. We have shown that the influence of electric space charge on the formation of the energy-release zone of a pulsed high-current electron beam cannot be neglected at low aerogel densities ($\rho_0 \leqslant 0.2$ g/cm^3). This charge can be calculated correctly using the model of the high-energy pulsed conductivity of insulators in Ref. 9 supplemented with allowance for the nonlinear influence of the electric fields. Allowance for their influence is especially important for electrons with energies of the order of the band gap, which provide for the high-energy breakdown of an insulator. A comparison of the calculated and experimental depths of the energy-release zone in the aerogel with $\rho_0 = 0.03$ g/cm^3 has enabled us to estimate the high-energy breakdown electric field $E_{br} \simeq 50$ kV/cm, which is two orders of magnitude less than the usual value.

We express our thanks to L. I. Rudakov for some useful discussions and valuable remarks, as well as the staff of the Institute of Catalysis of the Russian Academy of Sciences for preparing the samples.

This work was supported by the Russian Fund for Fundamental Research (Project No. 97-02-16729).

[1] M. Burchel and R. Thomson, Bull. Am. Phys. Soc. **40**, 1409 (1995).
[2] V. E. Fortov, A. S. Filimonov, V. K. Gryaznov *et al.*, in *Proceedings of the International Conference on Physics of Strong Coupled Plasmas*, W. D. Kraeft and M. Langes [Eds.], World Scientific, Singapore (1996), pp. 317–321.
[3] V. Ya. Ternovoĭ, V. K. Gryaznov, D. N. Nikolaev *et al.*, in *Abstracts of the 9th International Conference "Equations of State of Matter"* [in Russian], Nal'chik, (1996), p. 5.
[4] B. A. Demidov, V. P. Efremov, I. A. Ivonin *et al.*, Zh. Tekh. Fiz. **67**(11), 19 (1997) [Tech. Phys. **42**, 1258 (1997)].
[5] S. A. Bagnich, V. N. Bogomolov, D. A. Kurdyukov *et al.*, Fiz. Tverd. Tela (St. Petersburg) **37**, 2979 (1995) [Phys. Solid State **37**, 1642 (1995)].
[6] S. Feng and P. N. Sen, Phys. Rev. Lett. **52**, 216 (1984).
[7] A. F. Akkerman, Yu. I. Nikitushev, and V. A. Botvin, *Monte Carlo Solution of Problems Involving the Transport of Fast Electrons in Matter* [in Russian], Nauka, Alma-Ata (1972), 163 pp.
[8] V. V. Milyavskiĭ and V. A. Skvortsov, Teplofiz. Vys. Temp. **33**, 795 (1995).
[9] D. I. Vaĭsburd [Ed.], *High-Energy Solid-State Electronics* [in Russian], Nauka, Novosibirsk (1982), 227 pp.
[10] B. A. Demidov, M. V. Ivkin, V. A. Petrov *et al.*, J. Appl. Crystallogr. **46**, 100 (1979).
[11] K. Kanaya and S. Okadama, J. Phys. D **5**, 43 (1972).
[12] V. F. Baranov, *Dosimetry of Electron Radiation* [in Russian], Atomizdat, Moscow (1974), 229 pp.
[13] *Investigation of the Properties of Matter under Extreme Conditions* [in Russian], Moscow (1990), 217 pp.
[14] I. M. Sokolov, Usp. Fiz. Nauk **150**, 221 (1986) [Sov. Phys. Usp. **29**, 903 (1986)].
[15] R. R. Boade, J. Appl. Phys. **41**, 4542 (1970).
[16] Yu. A. Krysonov and S. A. Novikov, Prikl. Mekh. Tekh. Fiz., No. 6, 57 (1988).
[17] *Physical Parameters. A Handbook* [in Russian], Énergoatomizdat, Moscow (1991), 1232 pp.
[18] Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, Academic Press, New York (1966–1967) [Russian original, Nauka, Moscow (1966), 687 pp.].

Translated by P. Shelnitz

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

BRIEF COMMUNICATIONS

Application of a double insulated probe for the investigation of charged-particle beams in plasmas

V. Ya. Martens

Stavropol State Technical University, 355038 Stavropol, Russia

(Submitted May 27, 1997)
Zh. Tekh. Fiz. **68**, 121–123 (October 1998)

It is shown that the existence of a current in the external circuit of a double insulated probe without electrical bias between its electrodes is evidence of the presence of a stream (beam) of charged particles in a plasma. The species of particles forming a beam in a plasma, the direction and divergence of the beam, and information on the distribution of the current density over the beam cross section in the plasma can be obtained by determining the direction and strength of the current in the external circuit of the probe as it is positioned at various sites in the investigated system. © *1998 American Institute of Physics.* [S1063-7842(98)02110-2]

The application of a double insulated probe to determine electron temperature and plasma concentration is well known.[1] Here we reveal the capabilities of such a probe for the investigation of charged-particle beams in a plasma.

We have previously investigated the behavior of an insulated electrode immersed in a quasineutral plasma/electron-beam system. If this electrode is divided into two sections in such a way that the beam is incident on one section only, a current will flow between the sections when they are interconnected through an external electrical circuit and the total (net) current is zero. We inquire what information about the charged-particle beam can be obtained from measurements of the current in the external circuit.

We conditionally designate the first electrode as the section of the electrode section on which the electron beam is incident, and we call the other section the second electrode. The balance equations for the currents onto the first and second electrodes can be written in the form

$$i_{ip1}-i_{ep1}-i_{eb}(1-\sigma)+i=0, \tag{1}$$

$$i_{ip2}-i_{ep2}-i=0, \tag{2}$$

where i_{ip1} and i_{ep1} are the currents of plasma (thermal) ions and electrons incident onto the first electrode, i_{ip2} and i_{ep2} are the analogous currents onto the second electrode, i_{eb} is the current of beam electrons hitting the first electrode, σ is the secondary emission coefficient, and i is the current in the external circuit.

Taking Eqs. (1) and (2) into account, we can write an expression for the current in the external circuit in the form

$$i=\frac{1}{2}[i_{ip2}-i_{ep2}-i_{ip1}+i_{ep1}+i_{eb}(1-\sigma)]. \tag{3}$$

Applying similar reasoning to a plasma/ion-beam system, we obtain

$$i=\frac{1}{2}[i_{ip2}-i_{ep2}-i_{ip1}+i_{ep1}-i_{ib}(1+\gamma)], \tag{4}$$

where i_{ib} is the current of beam ions hitting the first electrode, and γ is the ion–electron emission coefficient.

The current in the external circuit therefore depends on the currents onto both electrodes and, as shown in Ref. 2, they are governed by the parameters of the plasma, the beam, and the material of the electrodes. Without the beam ($i_{eb}=0$ or $i_{ib}=0$) the sum of the remaining components of the current in the external circuit is also equal to zero, because for electrodes situated in a quasineutral plasma under a floating potential we have $i_{ip1}=i_{ep1}$ and $i_{ip2}=i_{ep2}$. It follows from this result that the very existence of a current in the external circuit is evidence of a stream of charged particles flowing in the plasma.

If the plasma concentration near equal-size electrodes were the same in the presence of a beam, the current in the external circuit would be equal to

$$i=\frac{1}{2}i_{eb}(1-\sigma) \tag{5}$$

or

$$i=-\frac{1}{2}i_{ib}(1+\gamma), \tag{6}$$

since $i_{ip1}=i_{ip2}$ and $i_{ep1}=i_{ep2}$.

In reality, owing to ionization of the gas by the beam, the plasma concentration near the first electrode can exceed the concentration near the second electrode, and it therefore follows from Eqs. (3) and (4) that the current in the external circuit can be lower in absolute value than the values given by Eq. (5) or (6). However, the direction of the current in the external circuit remains the same, and this fact can be exploited to determine the sign of the charge of the particles forming the beam. Figure 1 shows the direction of the cur-

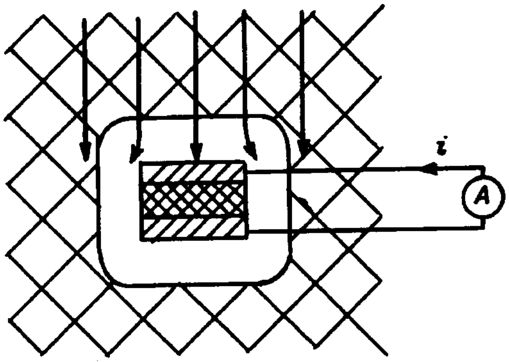

FIG. 1. Diagram of the double insulated probe.

rent in the external circuit for an electron beam in a plasma. The current *i* is in the opposite direction for a beam of positive ions in a plasma.

If the species and energy of the particles forming the beam are not known *a priori*, the electrodes must be made of a material for which the maximum secondary emission coefficient is smaller than unity. Otherwise the current *i* for an electron beam in a plasma can reverse direction as a result of the secondary-emission current exceeding the current i_{eb} [the sign of the coefficient $(1-\sigma)$ in Eq. (5) changes], an event that can be misinterpreted as the existence of an ion beam in the plasma.

It is important to note that an increase in the beam current density causes the current to increase in the external circuit; however, the maximum of the current density still corresponds to the maximum of the current in the external circuit. This fact can be used to investigate the distribution of the current density over the cross section of a beam in a plasma. Moreover, the current in the external circuit depends on the position of the probe relative to the direction of propagation of the beam. The current *i* is a maximum when the beam is incident only on one electrode (Fig. 1, and it is equal to zero when the beam particles incident on the probe are distributed uniformly between both electrodes. To achieve the latter situation, the double probe shown in Fig. 1 must be rotated 90° about an axis perpendicular to the direction of propagation of the beam. This fact can be used to determine the direction and divergence of a charged-particle beam in a plasma.

For experimental verification we used a double probe comprising two 3×3-mm silvered plates separated by a ceramic insulator. The probe was placed in the beam former (a metal tube of inner diameter $2r_0=110\,\mathrm{mm}$) of an electron gun.[3] The plasma in the beam former was generated by a divergent beam of electrons injected through an orifice of diameter 4 mm from a reflex discharge with a hollow cathode along the beam-former axis. The probe was set a distance of 5 cm from the injection orifice and could be moved in the radial direction, the planes of its electrodes perpendicular to the beam-former axis. The air pressure in the beam former was $\sim10^{-2}$ Pa, the plasma concentration on the axis was $\sim10^{8}\,\mathrm{cm}^{-3}$, the energy of the injected electrons was ~100 eV, and the beam current was 0.4 times the discharge current I_d, which varied from 100 mA to 200 mA in our experiments. Here the current *i* in the external circuit of the double probe attained several tens of microamperes on the beam-former axis.

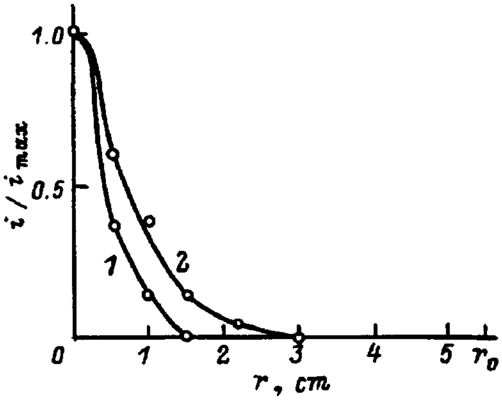

FIG. 2. Current in the external circuit of a double insulated probe versus radial position of the probe in the beam former. *1)* $I_d=100$ mA; *2)* $I_d=200$ mA.

Figure 2 shows graphs of the normalized current in the external circuit of the probe as a function of the radial position of the probe in the beam former. They indicate a nonuniform distribution of the current density over the beam current, with a maximum on the beam-former axis. An increase in the discharge current causes the beam diameter to increase form 3 cm to 6 cm in the plane of the measurements as a result of an increase in the divergence of the beam. The information obtained in these experiments about an electron beam in a plasma concur with the results of earlier investigations in the same system by other methods.[4,5]

The current $i=0$ in the region between the electron beam and the walls of the beam former (Fig. 2). To prove that a plasma exists in this region, but not an electron beam, we recorded the current–voltage (I–V) curves of the individual probes, which in this case were the first and second electrodes of the double probe. The probe I–V curves of the first and second electrodes included the electronic and ionic branches and coincided exactly for $I_d=100\,\mathrm{mA}$ in the case $r>1.5\,\mathrm{cm}$ and for $I_d=200\,\mathrm{mA}$ in the case $r>3\,\mathrm{cm}$. For shorter distances from the axis the electron and ion currents onto the upper electrode, i.e., the electrode facing injection of the beam, exceeded the corresponding currents onto the second electrode.

In summary, the species of particles forming a beam in a plasma, the direction and divergence of the beam, and information on the distribution of the current density over the beam cross section in the plasma can be obtained by determining the direction and strength of the current in the external circuit of a double probe positioned at various sites in the investigated system. In addition, the energy of the beam can be estimated from the probe potential.[2] The proposed investigations have all the inherent advantages of an insulated probe: minimal perturbation of the investigated system, thanks to the smallness of the tapped currents, and the ability to function in a nonstationary plasma.

The probe can be used to detect a stream of charged particles in a plasma by measuring its parameters and investigating the dynamics of the beam as it interacts with the gas and the plasma.

This work has been performed in cooperation with the International Soros Science Education Program (ISSEP), Grant No. 904d, 1997.

[1] A. M. Khovatson, *Introduction to the Theory of Gaseous Discharge* [in Russian] (Atomizdat, Moscow, 1980), p. 131.

[2] V. Ya. Martens, Zh. Tekh. Fiz. **66**(6), 70 (1996) [Tech. Phys. **41**, 559 (1996)].

[3] Yu. E. Kreĭndel', V. Ya. Martens, V. Ya. S'edin *et al.*, Prib. Tekh. É́ksp. No. 4, 18 (1982).

[4] V. A. Gruzdev, Yu. E. Kreĭndel', and V. Ya. Martens, Izv. Akad. Nauk Ser. Fiz. **43**, 1883 (1979).

[5] V. Ya. Martens and N. G. Rempe, *Abstracts of the Fifth All-Union Symposium on High-Current Electronics* [in Russian], Part 2 (Tomsk, 1984), pp. 195–197.

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Approximate relation for the distribution of the surface current densities on two ideally diamagnetic cylinders

A. I. Spitsyn

Kharkov State Technical University of Radio Electronics, 310726 Kharkov, Ukraine

(Submitted June 3, 1997; resubmitted April 20, 1998)
Zh. Tekh. Fiz. **68**, 124–125 (October 1998)

An approximate relation is obtained for the distribution of the surface current density on two ideally diamagnetic cylinders in which currents of arbitrary strengths are flowing. The relation is highly accurate and valid over a wide range of variation of the gap between cylinders. © *1998 American Institute of Physics.* [S1063-7842(98)02210-7]

An important factor in the investigation of separate parts of radiators or several radiators in close proximity,[1] in diamagnetic suspension,[2] and in connection with superconducting solenoids is the redistribution of the current density on the individual elements under the influence of other nearby parts.[3] In this paper we investigate the same influence in a system of two parallel, ideally diamagnetic cylinders of equal radius r_0, which carry currents I_1 and I_2 and are separated by a center-to-center distance $2H$ (see Fig. 1).

An exact relation for the distribution of the surface current density $J_i(\varphi_i)$ on one of the cylinders can be written in the form[4]

$$J_i(\varphi_i)=\frac{1}{2\pi r_0}\,\frac{h-\cos\theta_i}{\sqrt{h^2-1}}\left\{I_i+(I_1+I_2)\left[\frac{K}{\pi}\mathrm{dn}\left(\frac{K\theta_i}{\pi}\middle|m\right)-\frac{1}{2}\right]\right\}, \tag{1}$$

$$h=\cosh\left(\frac{\pi K'}{K}\right),\quad h=\frac{H}{r_0},$$

$$\theta_i=\arccos\frac{1-h\cos\varphi_i}{h-\cos\varphi_i},\quad i=1,2, \tag{2}$$

where $0<\varphi_i<\pi$ $(\pi>\theta_i>0)$ is the polar angle of one of the cylinders relative to the line joining the centers of the cylinders, K and K' are complete elliptic integrals of the first kind with direct and supplementary parameters m and $m_1=1-m$, and dn is a Jacobian elliptic function.[5]

The value of the parameter m for elliptic integrals is given by the first equation (2).

For calculations it is more convenient to use elementary functions rather than the special functions appearing in Eq. (1). From Eq. (1) we obtain an approximate expression for $J_i(\varphi_i)$, which is valid for closely spaced cylinders. In the limit $h\to 1$ the elliptic integral K tends to ∞, and $m\to 1$ $(m_1\to 0)$. Expanding the first equation (2) in a series up to and including terms to the fourth power, invoking the representation $m_1\sim 0$, $K'=\pi/2$, $K=\ln(4/m_1^{1/2})$, and solving the resulting equation for m_1 by means of the formula for $\mathrm{dn}(y|m)$ $(y\leqslant K)$ (formula 16.15.3 in Ref. 5), after suitable manipulations we can write the density distribution on, for example, the first cylinder in the form

$$\frac{J_1(\varphi_1)}{J_{01}}=\frac{h-\cos\theta_1}{\sqrt{h^2-1}}\left\{\frac{1}{2}\left(1-\frac{I_2}{I_1}\right)+\left(1+\frac{I_2}{I_1}\right)\frac{\alpha}{\pi}\times\left[\frac{1}{\cosh(\alpha\theta_1/\pi)}+4e^{-2\alpha}\sinh\left(\frac{\alpha\theta_1}{\pi}\right)\right]\right\},$$

$$\alpha=\frac{\pi^2}{2\sqrt{6}\left[\left[1+\frac{2}{3}(h-1)\right]^{1/2}-1\right]^{1/2}},\quad J_{01}=\frac{I_1}{2\pi r_0}, \tag{3}$$

and then on the second cylinder by interchanging the indices $1\leftrightarrow 2$. The solid curves in Fig. 1 represent the functions $J_1(\varphi_i)/J_{01}$ (curve *1*) and $J_2(\varphi_2)/J_{02}$ (curve *2*) (J_{0i} denotes the surface current density when it is distributed uniformly over the surface of the *i*th cylinder), calculated from Eq. (1) for a current ratio $I_2/I_1=2$; also shown in the figure is the

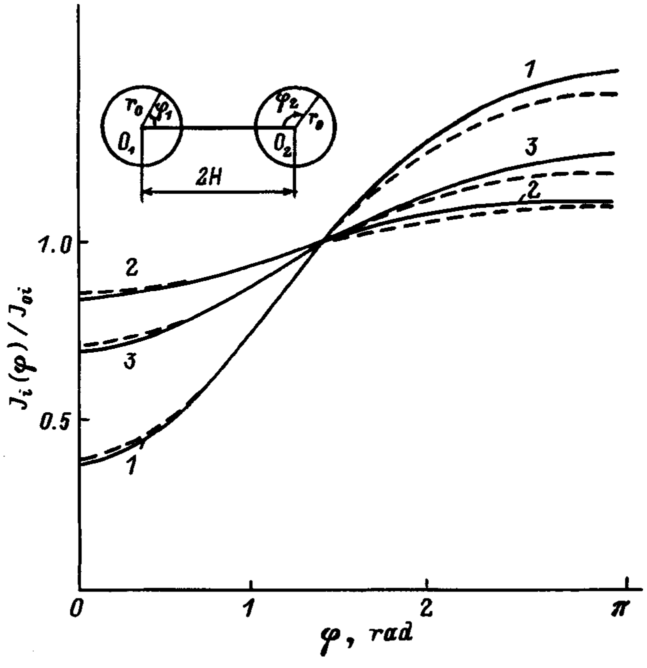

FIG. 1. Normalized current densities on the surfaces of the first (*1*) and second (*2*) cylinders for currents $I_2/I_1=2$ and on one of the cylinders for $I_2/I_1=1$ (*3*).

function $J(\varphi)/J_0$ (curve *3*) on one of the cylinders for $I_2/I_1=1$. The relative gap is $\xi=2(h-1)=5.14$. The dashed curves represent the corresponding functions calculated from Eq. (3). For a gap $5r_0$ the maximum error of deviation of the second term in Eq. (3) from the corresponding term in (1) does not exceed 4%. This error is even smaller for smaller gaps. Equation (3) is therefore highly accurate and valid over a wide range of variation of the gap between the cylinders. It can be used fairly easily in calculating the distribution of current densities for two closely spaced cylinders.

[1] E. A. Ivanov, *Diffraction of Electromagnetic Waves by Two Bodies* [in Russian] (Nauka i Tekhnika, Minsk, 1968), 584 pp.

[2] B. V. Jayawant, Rep. Prog. Phys. **44**, 411 (1981).

[3] P. L. Kalantarov and L. A. Tseĭmlin, *Calculation of Inductances* [in Russian] (Énergoatomizdat, Leningrad, 1986), 488 pp.

[4] A. I. Spitsyn, Zh. Tekh. Fiz. **63**(12), 1 (1993) [Tech. Phys. **38**, 1037 (1993)].

[5] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (U. S. Govt. Printing Office, Washington, D.C., 1964; reprint, Dover, New York, 1965) [Nauka, Moscow, 1979], 832 pp.

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Invariant-focusing two-dimensional lens for a finite-emittance beam

E. V. Shpak

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

(Submitted June 16, 1997)
Zh. Tekh. Fiz. **68**, 126–128 (October 1998)

A two-dimensional lens consisting of three pairs of plates is investigated. Expressions are given for the magnification of the lens in crossover-to-crossover conversion. Invariant-focusing regimes that preserve a constant crossover position against variation of the beam energy are determined for this lens. The influence of the angular and linear beam dimensions on the crossover position is analyzed in detail in the example of a particular lens configuration. © *1998 American Institute of Physics.* [S1063-7842(98)02310-1]

An invariant-focusing, axisymmetric lens consisting of three diaphragms or three cylinders has been investigated in detail;[1–3] the lens is designed to produce a finite-emittance beam. Two-dimensional lenses are widely used in mass spectrometers, in velocity analyzers, and in a number of other instruments. The invariant-focusing, two-dimensional, three-electrode lens maintains the resulting crossover in a fixed position as the beam energy varies. The results of similar calculations for zero-emittance beams have been published in a book,[4] but none have been carried out for finite-emittance beams.

In this paper we investigate a two-dimensional lens consisting of three pairs of plates extending along the y axis. The lens is shown schematically in Fig. 1. The distance between each pair of plates is H, and the distances between pairs is equal to $0.1H$. The length of the middle electrode is $A=0.5H$. The particle energy in the beam is varied by varying the ratio of the potentials on the third and first pairs of electrodes V_3/V_1. Operating regimes in which a variation of V_3/V_1 does not cause displacement of the lens-formed crossover are selected by varying the ratio of the potentials on the second and first pairs of electrodes V_2/V_1. The reference plane from which is measured the distances to the crossover at the input and output coincides with the middle of the lens ($z=0$, Fig. 1).

In practice the most commonly encountered boundary phase contours at the input to the system are well approximated by the expressions

$$\left(\frac{x_0}{X_0}\right)^k+\left(\frac{x_0'}{X_0'}\right)^k=1, \quad \left(\frac{y_0}{Y_0}\right)^k+\left(\frac{y_0'}{Y_0'}\right)^k=1, \tag{1}$$

where $k=2m/(2n-1)$, m and n are integers, and X_0, Y_0, X_0', and Y_0' are the maximum values of x_0, y_0, x_0', and y_0', respectively; for $k=2$ the boundary phase contours are ellipses.

In an electrostatic two-dimensional lens there is no force component acting on the charged particles along the y axis. It therefore makes sense to find the system-generated crossover only in the direction of the x axis. For contours of the form (1) the crossover of a paraxial beam coincides with the minimum of the projection of the envelope onto the xz plane. An expression describing crossover-to-crossover conversion has been derived in Ref. 5:

$$\frac{f_{2x}}{q_x}+\frac{f_{1x}}{p_x+B}=1, \tag{2}$$

where $B=w^k(p_x-f_{1x})^{(1-k)}$, $w=X_0/X_0'$, f_{1x} and f_{2x} are the focal lengths of the object and image spaces, respectively, p_x is the distance from the input crossover to the principal plane in the object space, and q_x is the distance from the output crossover to the principal plane in the image space.

Substituting $p_x=P_x-F_{1x}+f_{1x}$ and $q_x=Q_x-F_{2x}+f_{2x}$ into Eq. (2), we obtain

$$f_{1x}f_{2x}=(P_x+B-F_{1x})(Q_x-F_{2x}), \tag{3}$$

where $B=w^k(P_x-F_{1x})^{1-k}$, $P_x>F_{1x}$, $Q_x>F_{2x}$, F_{1x} and F_{2x} are the distances from the reference plane to the foci of the object and image spaces, respectively, P_x is the distance from the input crossover to the middle of the lens, and Q_x is the distance from the output crossover to the same plane.

Using previously published[4] values of the cardinal elements of a three-electrode, two-dimensional lens, we calculate the required interrelationships of the potentials on the lens electrodes to ensure the satisfaction of Eq. (3) for given P_x, Q_x, and w. Figure 2 shows the potential ratios V_2/V_1 as functions of the ratios V_3/V_1 for various values of the parameter w, $P_x=Q_x=6H$, and $k=2$. Curve *1* coincides with the zero-emittance curve. As the parameter w is increased, i.e., as the width of the beam in the input crossover X_0 is increased or the initial angular spread X_0' is decreased, the upper branch of the curve shifts downward, and the lower branch shifts upward, so that the curves lie inside curve *1*. This effect becomes more pronounced as w is increased.

Figure 3 shows the potential ratios V_2/V_1 as functions of V_3/V_1 for distances from the input and output crossovers to the reference plane $P_x=4H$ and $Q_x=6H$, $k=2$. This case corresponds to high lens magnifications. A departure of the curves characterizing crossover-to-crossover conversion from the zero-emittance curves begins to show up for smaller values of the parameter w. Curve *1*, calculated for $w=0.1$, essentially coincides with the corresponding curve for a zero-emittance beam. Curve *2* corresponds to $w=0.6$ and already

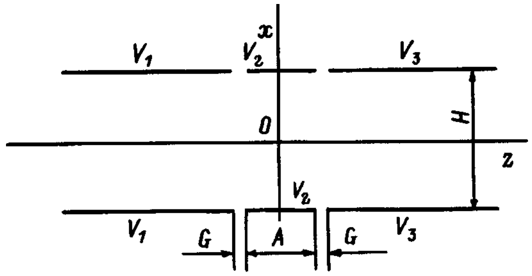

FIG. 1. Two-dimensional, three-electrode lens.

exhibits an appreciable departure from curve *1*. It is closed and lies inside the region bounded by curve *1*. With a further increase in w only the lower part of the curve remains intact. Curve *3* is calculated for $w=1$. The upper part of curve *1* differs from the lower part in that for each given ratio V_3/V_1 and large values of V_2/V_1 the principal plane and the focal plane of the object space are shifted closer to the object. As w is increased, the resulting crossover moves from the image position for $w=0$ to the focal plane of the lens. To maintain it in a fixed position, the force of the lens must be increased. For $F_{1x} \geqslant P_x$ the lens does not form a crossover. With an increase in $w \leqslant 1$ for a given value of $P_x=4H$, it is impossible to form crossover for a distance $Q_x=6H$ from the middle of the lens.

An expression for the lens magnification required for crossover-to-crossover conversion has been derived previously:[5]

$$M_x=\left\{\left[X_0' p_x\left(1-\frac{q_x}{g_x}\right)^k\right]+\left[X_0\left(1-\frac{q_x}{f_{2x}}\right)\right]^k\right\}^{1/k} X_0'^{-1}, \tag{4}$$

where q_x and g_x are the distances from the principal plane of the image space to the resulting crossover and to the image, respectively, and p_x is the distance from the input crossover to the principal plane of the object space.

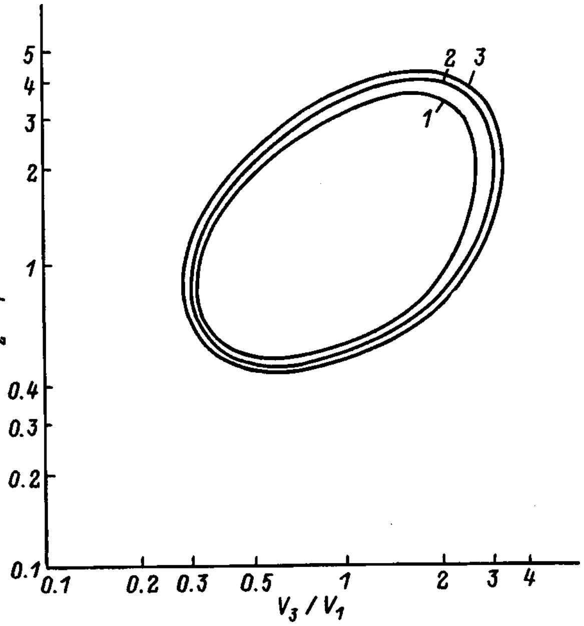

FIG. 2. Interdependence of the ratios of the potentials on the lens electrodes for $P_x=Q_x=6H$. *1*) $w=0.1$; *2*) $w=1$; *3*) $w=2$.

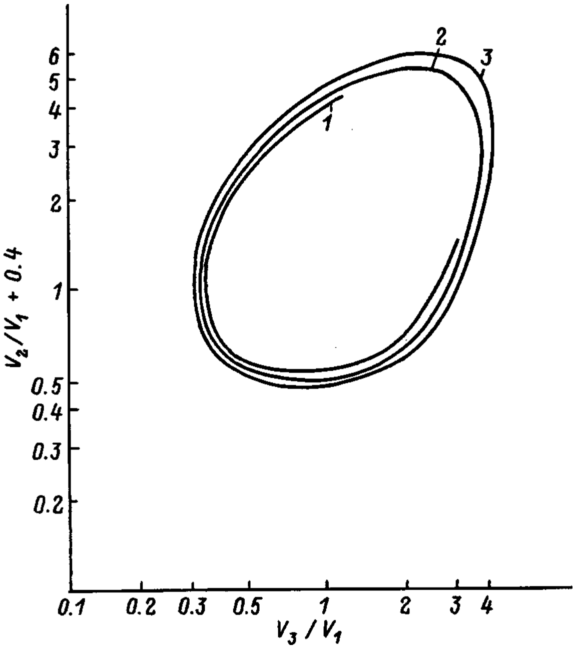

FIG. 3. Interdependence of the ratios of the potentials on the lens electrodes for $P_x=4H$ and $Q_x=6H$. *1*) $w=0.1$; *2*) $w=0.6$; *3*) $w=1$.

Equation (4) can be somewhat simplified:

$$M_x=\left[p_x^k\left(1-\frac{q_x}{g_x}\right)^k+w^k\left(1-\frac{q_x}{f_{2x}}\right)^k\right]^{1/k}. \tag{5}$$

We now write expressions for the magnification as a function of the parameters P_x, Q_x, F_{1x}, and F_{2x} relative to the reference plane. Substituting the quantities $p_x=P_x-F_{1x}+f_{1x}$ and $q_x=Q_x-F_{1x}+f_{2x}$ and making use of Eq. (2) together with Newton's equation relating the focal lengths f_{1x} and f_{2x} to the positions of the source and the image, we obtain

$$M_x=(Q_x-F_{2x})(B^k+w^k)^{1/k}/f_{2x}. \tag{6}$$

We have assumed here that the input crossover is in the same position as the object.

To compare with elliptical phase contours in one regime, we calculate the potentials on the electrodes for $k=4$ for the lower branch of the curve. For equal arms $P_x=Q_x=6H$, $w=1$, and $V_3/V_1=2$ the quantities V_2/V_1 and M_x change insignificantly and have the values $V_2/V_1=0.312$ and $M_x=0.938$ (for $k=2$, $V_2/V_1=0.361$ and $M_x=1.14$).

The following conclusions can be drawn from the calculations and a comparison with the results of Refs. 1–3. For a two-dimensional lens and also for a three-electrode axisymmetric lens the interdependence of the electrode potentials exhibits a complex behavior as the parameter $w=X_0/X_0'$ is varied. As w is decreased, the variations of the potentials approach the curves for a zero-emittance beam. As w is increased, the curves shift in such a way as to fall within the same curves for smaller values of w. With a further increase in w the curves spread out more, the upper part vanishing at first. An increase in the emittance by virtue of an increase in the angular spread and an increase in the dimensions of the beam is manifested by a reversal in the behavior of the curves. The influence of an increase in the beam width (or a decrease in its angular dimensions) becomes stronger as the output arm G_x increases relative to the input arm P_x.

It is instructive to calculate how the electrode potentials of a two-dimensional lens should vary not only to maintain the generated crossover in a fixed position, but also to preserve a constant magnification. At least four electrodes are needed to meet these conditions. Data are extremely scarce for such systems, even for a zero-emittance beam, and calculations of this kind are nonexistent for a finite-emittance beam.

[1] E. V. Shpak, Zh. Tekh. Fiz. **63**(1), 162 (1993) [Tech. Phys. **38**, 35 (1993)].
[2] E. V. Shpak and A. A. Smirnova, Zh. Tekh. Fiz. **65**(3), 109 (1995) [Tech. Phys. **40**, 285 (1995)].
[3] E. V. Shpak, Nucl. Instrum. Methods Phys. Res. A **363**, 64 (1995).
[4] E. Harting and F. H. Read, *Electrostatic Lenses* (Elsevier, Amsterdam–New York, 1976), 322 pp.
[5] E. V. Shpak and S. Ya. Yavor, Nucl. Instrum. Methods Phys. Res. A **239**, 288 (1985).

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Short-circuit current limiter utilizing a high-T_c superconductor

M. I. Petrov and S. G. Ovchinnikov

L. V. Kirenskiĭ Institute of Physics, Siberian Branch of the Russian Academy of Sciences,
660036 Krasnoyarsk, Russia; Physicotechnical Scientific-Research Institute at Krasnoyarsk State University,
660036 Krasnoyarsk, Russia

D. A. Balaev

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

V. I. Kirko

Physicotechnical Scientific-Research Institute at Krasnoyarsk State University, 660036 Krasnoyarsk, Russia

(Submitted June 24, 1997)
Zh. Tekh. Fiz. **68**, 129–130 (October 1998)

Experimental data are given on a prototype short-circuit current limiter utilizing a polycrystalline high-T_c superconductor of composition $Y_1Ba_2Cu_3O_{7-\delta}$. The limiter comprises a series-connected dc circuit element immersed in liquid nitrogen. To improve the efficiency, a polycrystalline high-T_c superconductor having an *S*-shaped current–voltage characteristic is used as the current limiter. © *1998 American Institute of Physics.* [S1063-7842(98)02410-6]

The protection of electrical circuits against catastrophic extremes in short-circuit regimes poses a timely problem in view of the nonexistence of 100% reliable protection elements. The conceptual possibility of employing superconducting in shutdown devices (cryotrons) has been known for some time.[1] The advent of high-temperature (high-T_c) superconductors has led to reports of the design feasibility of current limiters operating at liquid-nitrogen temperature.[2,3] It has been proposed[4] that a single-crystal high-T_c superconductor be used as the active element of a cryotron, thereby affording the possibility of exploiting anisotropy to separate the control current and the working current. The development of a 2.2-kA cryotron with a control winding made from a bismuth high-T_c superconductor has been reported.[5]

Here we give the results of measurements of the parameters of a polycrystalline (ceramic) high-T_c superconducting device, which is not connected into a cryotron configuration, but functions as a series circuit component. In the ideal scenario this type of ballast resistance is equal to zero for a below-critical (J_c) transport current, but in the event of an emergency excess with the current exceeding J_c, it connects into the circuit and limits the current at a safe level. In the practical implementation of the limiter it is necessary to solve a number of technical problems, foremost of which, in our opinion, are the problem of contacts in the high-T_c superconductor, the removal of heat from the high-T_c superconducting protective element in the short-circuit-protection regime (when the high-T_c superconductor remains the sole user), and the corollary problem of matching the voltage drop across the high-T_c superconducting element in the resistive state with the voltage of the protected circuit; the matching problem dictates the geometry of the element and its power release per unit volume.

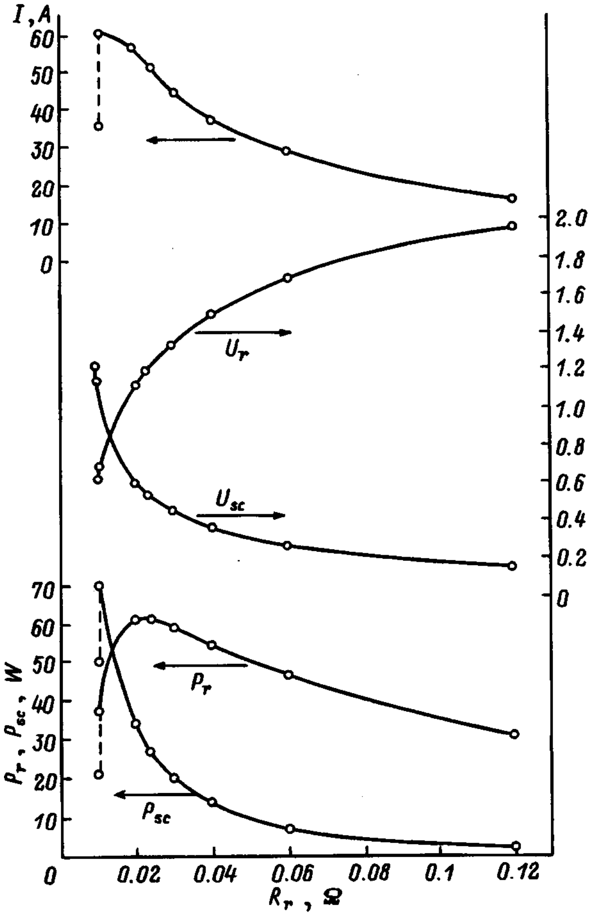

FIG. 1. Circuit variables (the current I in the circuit formed by the series-connected battery, superconductor, and load resistance R_r; the voltage drop U_r across R_r; the voltage drop U_{sc} across the superconductor; the power P_r released in R_r, and the power P_{sc} released in the superconductor) versus load resistance R_r.

The samples used for the measurements were made by pressing from previously prepared high-T_c superconducting $Y_1Ba_2Cu_3O_{7-\delta}$ powder. Silver contact patches for the current leads were formed on the end faces of the sample. The samples had typical dimensions of $2\times7\times44$ mm. To maximize the contact area, the current contacts were coated with an In–Ga eutectic. The assembled device was placed in a liquid-nitrogen tank. Owing to the smallness of the contact resistance, the difference in the voltage drops measured by two-contact and four-contact techniques was significant; two-probe data are given here. The current source was a 6 ST-132 battery with its six elements connected in parallel. In Fig. 1 all the curves for one series of measured samples are plotted as functions of the load resistance R_r, in this case a step rheostat.

It is evident from the figure that as R_r is decreased, the current in the circuit for $I>20$ A begins to deviate from the hyperbolic law $I=U/R_r$; at $R_r=0.01\,\Omega$ the current attains its maximum value ~60 A and then spontaneously decreases to ~35 A. We attribute this behavior to the transition of the current–voltage (I–V) curve into a branch having a higher differential resistance, a phenomenon that we have observed previously.[6] The figure shows how the voltage drops across the superconductor U_{sc} and across the load U_r are redistributed.

Clearly, a protection effect is observed at the minimum attained load resistance, where specifically the voltage drop across the load is lower than that across the superconductor. The variations of the power releases in the load P_r and in the superconductor P_{sc} are also shown in the figure. It is evident that when the load is a minimum, the power released in it is only ~20 W, as opposed to ~400 W without the high-T_c superconductor (in the first approximation $P_r=U_2/R_r$, and the indicated value is obtained for $U=2$ V and $R_r=0.01\,\Omega$). The power in the superconductor does not exceed 70 W in this case and is effectively removed by the liquid nitrogen.

Consequently, a high-T_c superconducting sample having an *S*-shaped I–V curve can be used to achieve a switching effect and to construct an alternative to the cryotron current limiter.

Although these results imply that the parameters of the current limiter utilizing a high-T_c superconductor are far from perfect, we are hopeful that a detailed study of the hysteresis feature of the I–V curves of polycrystalline high-T_c superconductors with a view toward the practical exploitation of this hysteresis will set the stage for the design of a protection element with parameters closely approaching practical requirements.

[1] I. N. Glazkov, in *Research on Superconducting Electrical Power Equipment* [in Russian] (ITF SO RAN, Novosibirsk, 1980), pp. 22–29.

[2] G. P. Amelin, in *Proceedings of the Conference on Problems in High-Temperature Superconductivity* [in Russian], Vol. 2 (Sverdlovsk, 1987), pp. 240–241.

[3] J. Chen and Z. Chen, in *Proceedings of the Fifth International Conference on Materials and Mechanisms of Superconductivity and High-Temperature Superconductors* (Beijing, China, 1997), pp. 305–306.

[4] A. Yu. Volkov, Pis'ma Zh. Tekh. Fiz. **16**(4), 11 (1990) [Sov. Tech. Phys. Lett. **16**(2), 128 (1990)].

[5] G. B. Lubkin, Phys. Today **49**, No. 3, 48 (1996).

[6] M. I. Petrov, S. N. Krivomazov, B. P. Khrustalev, and K. S. Aleksandrov, Solid State Commun. **82**, 453 (1992).

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Silicon diffused diodes with nearly ideal current–voltage characteristics

N. S. Boltovets, K. A. Ismaĭlov, R. V. Konakova, and M. B. Tagaev

Institute of Semiconductor Physics, National Academy of Sciences of Ukraine, 252650 Kiev, Ukraine

(Submitted July 9, 1997)
Zh. Tekh. Fiz. **68**, 131–132 (October 1998)

It is shown that nearly ideal current–voltage characteristics ($I-V$ curves) can be obtained for small-area, shallow silicon diffused p–n junctions irradiated with ^{60}Co gamma rays in the dose range $10^3-5\times10^5$ Gy. In the irradiated diodes the current transfer mechanism is observed to shift from generation–recombination to diffusion. The nonideality factor on the forward branch of the $I-V$ curve decreases from 1.68 in the unirradiated diode to 1.17 in a diode irradiated to a dose of 5×10^5 Gy. A saturation current is observed on the reverse branch of the $I-V$ curves of irradiated diodes at room temperature. © *1998 American Institute of Physics.* [S1063-7842(98)02510-0]

It is common knowledge that generation and recombination constitute the typical current transfer mechanism in silicon diffused p–n junctions over a wide range of temperatures, including room temperature. This mechanism, which was proposed by Sah, Noyse, and Schockley[1] and which was later confirmed by many authors and incorporated into a number of papers,[2–6] remained essentially unchallenged by anyone prior to 1980. During 1980–84 V. V. Tuchkevich *et al.* succeeded in showing that a diffusion mechanism of current transfer can be realized at room temperature in silicon p–n junctions that have been subjected to special heat-treatment regimes,[7] in p–n junctions configured so as to eliminate the influence of the surface,[8] and in p–n junctions between polycrystalline and single-crystal silicon.[9,10] The cited studies employed nontrivial techniques for fabrication and fabrication and processing of the p–n junction before the contacts were applied to the device structure.

Here we demonstrate the feasibility of obtaining a nearly ideal $I-V$ curve for small-area silicon diffused p–n junctions of the kind used in the fabrication of IMPATT (impact avalanche and transit-time) diodes. In contrast with Refs. 7–10, we achieve this end by exposing encapsulated diodes to ^{60}Co gamma rays in the dose range $10^3-5\times10^5$ Gy (irradiation intensity 3 Gy/s; the temperature in the irradiated zone did not exceed $+50$ °C).

Silicon diffused p–n junctions were prepared by diffusing boron from the gaseous phase at a temperature of 1050 °C into a silicon epitaxial $n-n^+$ structure for 30–45 min. The depth of the p–n junction was 0.6 μm. The dopant concentrations were $10^{20}\,\mathrm{cm}^{-3}$ (boron) in the p^+ layer, $3\times10^{16}\,\mathrm{cm}^{-3}$ (phosphorus) in the n layer, and $10^{19}\,\mathrm{cm}^{-3}$ in the n^+ substrate, and the thicknesses of n and n^+ regions were 1.5 μm and 300 μm, respectively. Diodes of diameters of 30 μm were fabricated by an integrated heat-sinking technology and were mounted in an IMPATT diode casing.[11]

The forward and reverse branches of the $I-V$ curves of the diodes were measured in the temperature range 300 $-$370 K before and after ^{60}Co gamma irradiation in the dose range $10^3-5\times10^5$ Gy. Irradiation of the samples in the interval 10^3-10^4 Gy scarcely altered the $I-V$ curves at all.

Figure 1a shows the forward branches of the $I-V$ curves of one of the investigated diodes before (curve *1*) and after irradiation to doses of 10^5 Gy, 2×10^5 Gy, and 5×10^5 Gy (curves *2–4*, respectively). The initial $I-V$ curve is described by the equation

$$I=I_0[\exp(eV/nkT)-1],$$

where I_0 is the saturation current, e is the elementary charge, n is a nonideality factor, k is Boltzmann's constant, T is the temperature, and V is the applied voltage.

In the unirradiated sample the nonideality factor is equal to 1.68, and the initial part of the $I-V$ curve has an excess leakage current. The nonideality factor drops to 1.5 after ^{60}Co gamma irradiation to a dose of 10^5 Gy and to 1.17 after a irradiation to 5×10^5 Gy. In the latter case the leakage current essentially vanishes, and the saturation current decreases by two orders of magnitude relative to the unirradiated sample. The lowering of the nonideality factor to 1.17 indicates a nearly total absence of a recombination component of the forward current.

Typical reverse branches of the $I-V$ curves before and after ^{60}Co gamma irradiation are shown in Fig. 1b. The initial $I-V$ curves of the unirradiated diode have two intervals: a thermal generation interval with $I_{\mathrm{op}}\approx\sqrt{V_{\mathrm{op}}}$ and a breakdown interval. As the gamma irradiation dose is increased from 10^5 Gy to 5×10^5 Gy, the reverse current decreases, and the reverse current is independent of the voltage over an interval of increasing length. The temperature dependence of the current in the latter interval maintains a single slope with activation energy $E_a\sim1.1$ eV, indicating that the diffusion current is predominant over the thermal generation current (Fig. 2, curves *1–4*). A significant attribute of the temperature dependence of the reverse current of the unirradiated sample is its segregation into two parts: a diffusion part in the high-temperature range and a generation–recombination part in the temperature interval $300-370$ K. For irradiation to 10^5 Gy the length of the generation–recombination interval decreases, and when the dose is further increased to 5

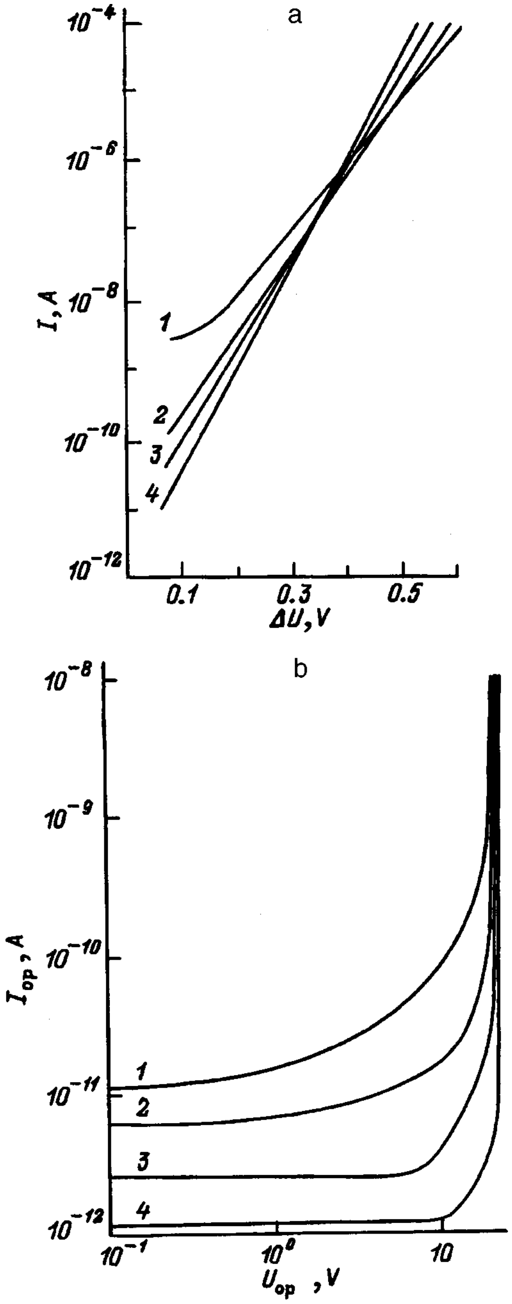

FIG. 1. a) Forward branches of the $I-V$ curves of a silicon diffused $p-n$ junction of diameter 30 μm: *1*) unirradiated diode; *2–4*) after ^{60}Co γ-irradiation to doses of 10^5 Gy, 2×10^5 Gy, and 5×10^5 Gy, respectively. b) Reverse branches of the $I-V$ curves of a silicon diffused $p-n$ junction of diameter 30 μm: *1*) unirradiated diode; *2–4*) after ^{60}Co γ-irradiation to doses of 10^5 Gy, 2×10^5 Gy, and 5×10^5 Gy, respectively.

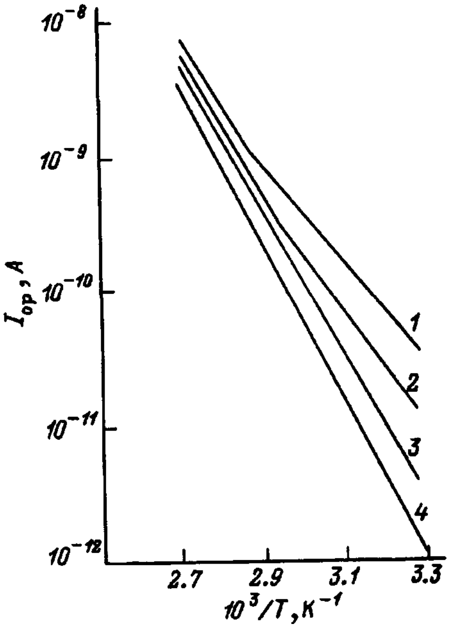

FIG. 2. Temperature dependence of the reverse current of a silicon $p-n$ junction of diameter 30 μm, measured at $V_{op}=5$ V: *1*) unirradiated diode; *2–4*) after ^{60}Co γ irradiation to doses of 10^5 Gy, 2×10^5 Gy, and 5×10^5 Gy, respectively.

$\times10^5$ Gy, the temperature dependence of the reverse current is governed by a single predominant current transfer mechanism: diffusion.

Our investigations have thus shown that diodes with almost ideal room-temperature $I-V$ curves can be obtained by irradiating encapsulated small-area silicon diffused diodes with gamma rays in the dose range $10^5-5\times10^5$ Gy.

This work has been supported by the Ukrainian Science and Technology Center, Project No. 464.

[1] S. T. Sah, R. N. Noyse, and W. Schockley, Proc. IRE **45**, 1228 (1957).

[2] S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. (Wiley–Interscience, New York, 1981; Mir, Moscow, 1984), 456 pp.

[3] K. V. Ravi, *Imperfections and Impurities in Semiconductor Silicon* (Wiley, New York, 1981; Mir, Moscow, 1984), 475 pp.

[4] W. C. Till and J. T. Luxon, *Integrated Circuits: Materials, Devices, and Fabrication* (Prentice-Hall, Englewood Cliffs, N.J., 1982; Mir, Moscow, 1985), 501 pp.

[5] V. P. Grigorenko, P. G. Dermenzhi *et al.*, *Modeling and Automation of the Design of Power Semiconductor Devices* [in Russian] (Énergoatomizdat, Moscow, 1988), 280 pp.

[6] S. V. Bulyarskiĭ and N. S. Grushko, *Physical Principles of the Functional Diagnostics of p–n Junctions Containing Defects* [in Russian] (Shtiintsa, Kishinev, 1992), 236 pp.

[7] E. V. Ivanov, L. B. Lopatina *et al.*, Fiz. Tekh. Poluprovodn. **16**, 207 (1982) [Sov. Phys. Semicond. **16**, 129 (1982)].

[8] E. V. Ivanov, L. B. Lopatina *et al.*, Pis'ma Zh. Tekh. Fiz. **6**, 874 (1980) [Sov. Tech. Phys. Lett. **6**, 377 (1980)].

[9] L. B. Lopatina, V. L. Sukhanov *et al.*, Pis'ma Zh. Tekh. Fiz. **5**, 11 (1979) [Sov. Tech. Phys. Lett. **5**, 4 (1979)].

[10] L. E. Klyachkin, L. B. Lopatina *et al.*, Fiz. Tekh. Poluprovodn. **17**, 1648 (1983) [Sov. Phys. Semicond. **17**, 1049 (1983)].

[11] R. V. Konakova, M. Ya. Skorokhod, and L. V. Shcherbina, Élektron. Promst., No. 6, 68 (1990).

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Method and device for the nonperturbative diagnostics of a high-current beam of low-energy neutral particles using photoionization electrons

A. S. Artemov

Joint Institute for Nuclear Research, 141980 Dubna, Moscow District, Russia

(Submitted September 10, 1997)
Zh. Tekh. Fiz. **68**, 133–135 (October 1998)

A diagnostic method is proposed wherein the momentum distribution of electrons emanating essentially in a single plane from a photoionization zone of small geometrical dimensions is measured at two different angles relative to the beam axis. To implement the method for a high-current beam of low-energy neutral particles (tens to hundreds of kiloelectron-volts for hydrogen or deuterium atoms), a device is proposed which is capable of real-time, nonperturbative measurement of the distributions of the particles with respect to momentum, over the beam cross section, and in transverse phase space. © *1998 American Institute of Physics.* [S1063-7842(98)02610-5]

High-current beams of low-energy (tens or hundreds of kiloelectron-volts) hydrogen or deuterium atoms are widely used for heating a plasma to thermonuclear temperatures and for noninductive current maintenance in existing and projected tokamaks.[1] A beam of this type carrying a current in the tens of amperes and having a pulse duration of several seconds is generated from individual elementary beams of positive or negative ions by means of charge transfer on a gas with subsequent separation in a magnetic field. The energy flux density of the beam at the tokamak input can be varied over a wide range (from a state of maximum concentration to a uniform distribution over a large surface) by regulating the angles of divergence of the elementary beams and profiling the emission surfaces. To monitor the formation of such high-current neutral beams, it is necessary to develop nonperturbative methods utilizing secondary particles or photons. A possible technique is proposed below, based on the detection of electrons generated in the photoionization of a negligibly small fraction of the atoms in the beam.

In the low-energy range the correspondence between the distribution of hydrogen (or deuterium) atoms and the distribution of ionization electrons on a diagnostic corpuscular or photon target is very imprecise.[2] Nonetheless, the highly monochromatic state of the laser beam used to form the photon target and the two-particle character of the particle decay during the production of electrons on the photon target opens up additional diagnostic possibilities in this case. Figure 1a shows a kinematic diagram of the generation of an electron in the YZ plane after the absorption of a photon by a nonrelativistic particle A^0 with momentum $\mathbf{P}_0$. It is assumed here that the axis of the beam coincides with the axis of the transport channel (Z' axis). The electron momentum (P_e) in the rest frame of the decaying particle depends on the energy ($\hbar\omega$) of the absorbed photon and the photoionization threshold (ε_n) of the active quantum state: $P_e=\sqrt{2m_e(\hbar\omega-\varepsilon_n)}$. It is readily shown that if two analyzers aimed at the photon target O have sufficient angular resolution ($\delta\theta$) and momentum resolution ($\delta P_e/P_e$) and if $\theta_0 \ll \Delta\theta_e/2$, the characteristics $\Delta P_0/P_0$ and $\Delta\theta_0$ of the beam particles in the vicinity of the diagnostic target are determined from the values (P_{e1} and P_{e2}) and spreads ($\Delta P_{e1}>0$ and $\Delta P_{e2}>0$) of the electron momenta at the corresponding angles of measurements θ_1 and θ_2 relative to the Z axis ($\theta_0 \ll \theta_2 < \theta_1 < \Delta\theta_e/2$):

$$\Delta P_0/P_0=\left|[\tan\theta_1\cdot\Delta f_e(\theta_2) -\tan\theta_2\cdot\Delta f_e(\theta_1)]/(\tan\theta_1-\tan\theta_2)\right|, \qquad (1)$$

$$\Delta\theta_0=\left|[\Delta f_e(\theta_2)-\Delta f_e(\theta_1)]/(\tan\theta_1-\tan\theta_2)\right|, \qquad (2)$$

where $\Delta f_e(\theta)=[\Delta P_{e1}(\theta)+\Delta P_{e2}(\theta)]/[P_{e1}(\theta)+P_{e2}(\theta)]$; $P_0=M_0\cdot[P_{e1}(\theta)+P_{e2}(\theta)]/(m_e 2\cos\theta)$ is the average momentum of the incident particles, $\theta=\theta_1$ or θ_2, and M_0 and m_e are the neutral-particle and electron masses, respectively.

A schematic of the device used to measure the particle-beam distributions with respect to the momentum P_0, in the XY cross section, and in the $Y'Y$ phase plane is shown in Fig. 1b. The directions and acceptances of the analyzers $A1$ and $A2$ are chosen so as to detect photoelectrons emanating only from the focused region of the target 0 and in a near-YZ median plane ($\Delta\varphi_e \ll 1$). The beam is scanned over the cross section by shifting the lens E with the analyzers along the X axis with the mirror (M) fixed in space and also by displacing all these elements as a unit whole along the Y axis. If the power of the photon target (which governs the photoionization probability) is monitored, these beam characteristics can be reproduced by means of Eqs. (1) and (2) from the measured electron momentum distribution as a function of the position of the target focus in the XY plane. Unobstructed beam transmission and nonperturbative beam diagnostics are achieved in this arrangement at neutral-particle energies (E_0) given by the expression

$$\sqrt{\frac{M_0}{m_e}\,\frac{(\hbar\omega-\varepsilon_n)}{E_0}}\,L_e \geqslant d_b, \qquad (3)$$

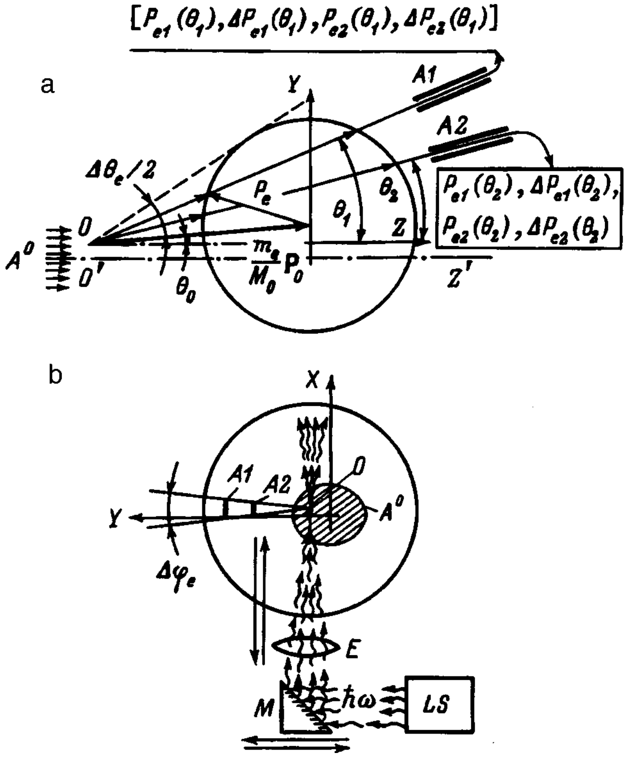

FIG. 1. a) Kinematic diagram of the generation of detected photoelectrons; b) schematic of the device for nonperturbative diagnostics of a neutral-particle beam A^0.

where d_b and L_e are the transverse width of the beam in the Y direction and the transit base line of electrons from the target to the analyzer, respectively.

The neutral-particle quantum state to be used for the measurements is determined from the condition for its selective photoionization. The best choice for $H^0(D^0)$ atoms is their metastable $2s$ state. In contrast with the diagnostic method discussed in Ref. 3, the optimum for our case is the maximum possible value of $\hbar\omega-\varepsilon_n$. The required average and pulse energy flux densities of the laser (LS) for performing measurements are determined by the condition for the extraction of information from the total flux of electrons generated on the diagnostic target and in the residual gas and by the required duty cycle of the device.

We now examine the capabilities of the diagnostic method in the example of a high-current beam of hydrogen atoms of energy $E_0=400$ keV, $d_b\approx 4$ cm, and particle current density $j_0\approx 1$ A/cm^2. The percent content of the metastable $2s$ quantum state in the generation of such a beam is estimated as $\delta_{02}\approx 5\%$ (Ref. 4). The selective photoionization of this state can be achieved, for example, by means of a fourth-harmonic beam from a Nd^{+3}:YAG laser with $\hbar\omega=4.66$ eV (Ref. 3). The corresponding cross section of interaction with H($2s$) atoms is $\sigma_{i|2s}\approx 7\times 10^{-18}\,\mathrm{cm}^2$, and a photoionization coefficient close to unity is attained in the vicinity of the focused target at a target power

$$P_\gamma \approx 10^7\sqrt{E_0}\cdot d_{\gamma f}, \tag{4}$$

where P_γ and E_0 are expressed in W and keV, respectively, $d_{\gamma f}\approx f\cdot\alpha_\gamma$ is the minimum transverse width of the focal region of the target in the XY plane, f is the focal length of the lens (cm), and α_γ is the angular divergence of the laser beam (rad).

For $f=30$ cm and $\alpha_\gamma\approx 10^{-3}$ rad we obtain $P_\gamma\approx 6$ MW. This quantity falls well within the standard range of pulsed lasing power for the given type of laser. If we use electron analyzers with $\delta\theta\approx 5\times 10^{-4}$ rad and $\delta P_e/P_e\approx 10^{-4}$ (e.g., of the Hughes–Rojansky type), canted in the YZ plane to satisfy the relation $\theta_1-\theta_2\approx\theta_2\approx\Delta\theta_e/6$, from Eqs. (1)–(3) we obtain $L_e\approx 60$ cm and the following estimates of the error of measurement of the beam characteristics:

$$\frac{\delta P_0}{P_0}\approx 3\frac{\delta P_e}{P_e}\approx 3\times 10^{-4};$$

$$\delta Y'\equiv\delta\theta_0\approx 6\sqrt{\frac{m_eE_0}{M_0(\hbar\omega-\varepsilon_n)}}\,\frac{\delta P_e}{P_e}\approx 8\times 10^{-3}\,\mathrm{rad}. \tag{5}$$

The prospects for improving these accuracy ratings rest primarily on the possibility of decreasing $\delta P_e/P_e$ with allowance for the spreading of the bunch of electrons to be detected en route to and inside the analyzer on account of its space charge.

The current of photoelectrons detected, for example, by analyzer $A2$ from the focal region of the target over the duration of the laser pulse for $\Delta\varphi_e\approx 0.1$ rad is estimated as

$$I_{e2}\approx 0.2\cdot j_o\cdot\delta_{02}\cdot d_{\gamma f}\cdot L_{\gamma f}\cdot(\delta\theta\times\Delta\varphi_e/\Delta\theta_e)\approx 10^{-7}\,\mathrm{A}, \tag{6}$$

where $L_{\gamma f}\approx\alpha_\gamma\cdot f^2/D_\gamma$ is the depth of the waist of the laser beam in the focal region at the level of a 10% increase in its transverse size as compared to $d_{\gamma f}$, and D_γ is the diameter of the photon beam before it enters the lens.

In Eq. (6) it is assumed that $D_\gamma\approx 1$ cm and that the dimensions of the target zone viewed by the analyzer along the X and Y coordinates coincide with $L_{\gamma f}\approx 10$ mm and $d_{\gamma f}\approx 0.3$ mm, respectively. The value obtained for I_{e2} is more than two orders of magnitude greater than the background electron current at a residual gas pressure $P_b\approx 10^{-4}$ torr. Here the average photoelectron energy is $E_e\approx 220$ eV, and the interval between the maxima of the corresponding distribution is $|E_{e|\max 1}(\theta_2)-E_{e|\max 2}(\theta_2)|\approx 2E_e\times\Delta\theta_e\approx 65$ eV.

En route to the analyzer the ribbonlike stream of electrons to be detected spreads out in the transverse direction under the influence of its own space charge. This effect produces losses, increases the angular spread of electrons at the analyzer input slit, and leads to a loss of resolution. The degree to which these effects are manifested is characterized by the quantity $\Delta d_{\gamma f}/d_{\gamma f}\approx 5\times 10^4 I_e[\mathrm{A}]\cdot(E_e[\mathrm{eV}])^{-1.5}L_e^2/(L_{\gamma f}\cdot d_{\gamma f})\approx 0.2$, and they can be disregarded in the given situation. The influence of the space charge of the electrons on the dispersion properties of the given analyzer is characterized by the quantity $Q_a=I_e/(d_{\gamma f}\times L_{\gamma f}\times E_e^{1.5})$ and is perceptible for values $10^{-7}-10^{-6}\,\mathrm{A}\cdot\mathrm{cm}^{-2}\mathrm{eV}^{-1.5}$ (Ref. 5). For the case analyzed here we have $Q_a\approx 10^{-9}\,\mathrm{A}\cdot\mathrm{cm}^{-2}\mathrm{eV}^{-1.5}$, which is more than two orders of magnitude smaller than the above-indicated threshold.

The duration of the essentially constant pulse amplitude of the given laser can be equal to $\tau_\gamma\approx 10$ ns. For the generated photoelectron current this time is sufficient for performing the necessary measurements with the analyzers operating in the electron-momentum scanning regime on a selected part of the beam cross section. The time required for obtaining complete information depends on the speed of mechanical movement of the corresponding elements of the device along the *X* and *Y* coordinates within the limits of the beam of hydrogen atoms. With a laser firing frequency greater than or of the order of 10 Hz, detailed information about the characteristics of the investigated beam can be obtained over the duration of the quasicontinuous pulse (a few seconds).

[1] N. V. Pleshivtsev and N. N. Semashko, in *Progress in Science and Technology, Series on the Physical Foundations of Laser and Beam Technology*, Vol. 5 (*Ion-Beam Technology*) [in Russian] (VINITI, Moscow, 1989), pp. 55–112.

[2] A. S. Artëmov, in *Proceedings of the 14th Conference on Charged-Particle Accelerators* [in Russian], Vol. 2 (1995), pp. 36–40.

[3] A. S. Artiomov, JINR Preprint No. E9-92-501, Joint Institute for Nuclear Research, Dubna (1992).

[4] G. H. Gillespie, Nucl. Instrum. Methods Phys. Res. B **10/11**, pp. 22–25 (1985).

[5] Yu. G. Gamayumov and I. G. Kozlov, Zh. Tekh. Fiz. **38**, 531 (1968) [Sov. Phys. Tech. Phys. **13**, 392 (1968)].

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Influence of the doping profile on the collector junction breakdown voltages in planar *n–p–n* transistors

N. A. Samoĭlov, A. N. Frolov, and S. V. Shutov

Kherson Industrial Institute, 325008 Kherson, Ukraine

(Submitted October 13, 1997)
Zh. Tekh. Fiz. **68**, 136–137 (October 1998)

Planar *n–p–n* transistors are fabricated with various profiles of the base impurity distribution, and their electrical parameters are investigated. An equation for calculating the breakdown voltage of the collector–base junction is proposed on the basis of an experimental data analysis. It is shown that the use of a planar *p–n* junction raises the collector junction breakdown voltages of planar *n–p–n* transistors. *© 1998 American Institute of Physics.* [S1063-7842(98)02710-X]

The operational reliability of bipolar transistors, including planar configurations, depends on the maximum allowed (breakdown) voltages $U_{\mathrm{em.b.0}}$, $U_{\mathrm{c.b.0}}$, and $U_{\mathrm{c.em.0}}$. The interrelationship of $U_{\mathrm{c.em.0}}$ and the breakdown voltage $U_{\mathrm{c.b.0}}$ of the collector junction by the cascade breakdown mechanism is given by the equation[1]

$$U_{\mathrm{c.em.0}}=\frac{U_{\mathrm{c.b.0}}}{\sqrt[m]{B_N+1}}, \qquad (1)$$

where B_N is the current gain of the transistor connected into a circuit with a common emitter, and $m=4$ for *n–p–n* transistors.

In transistor design the radius of curvature of the diffusion region is taken into account in calculating the breakdown voltages.[2] If this radius is set equal to the depth of the collector *p–n* junction, the equation for calculating the breakdown voltage of a planar collector–base junction of cylindrical shape assumes the form

$$U_{\mathrm{c.b.0}}=\frac{U_{nn}\cdot x_{j\sigma}}{Z_{pn}}\left[\sqrt{2Z_{pn}/x_{j\sigma}+1}-1\right]. \qquad (2)$$

Here U_{nn} is the breakdown voltage of the plane part of the *p–n* junction due to the resistivity of the high-resistivity part of the collector at $U=U_{nn}$.

The voltage $U_{\mathrm{c.b.0}}$ can be raised by any of several industrial design techniques: the application of a guard ring,[3] an extended base contact, mesa structures, divider rings,[4] etc. In particular, the character of the dopant distribution during the formation of the *p–n* junction has a significant influence on the value of the breakdown voltage.

To establish the dependence of the breakdown voltage $U_{\mathrm{c.b.0}}$ on the method of fabrication of the collector *p–n* junction, we have synthesized transistor structures from three nominal types of n–n^+ epitaxial structures having different thicknesses and resistivities of the epitaxial layers ($\rho=2.4\ \Omega\cdot\mathrm{cm}$, $d=12\ \mu\mathrm{m}$; $\rho=3.5\ \Omega\cdot\mathrm{cm}$, $d=16\ \mu\mathrm{m}$; $\rho=4.5\ \Omega\cdot\mathrm{cm}$, $d=20\ \mu\mathrm{m}$). The structures were grown by vapor-phase epitaxy on (111)-oriented 350 ÉKÉS 0.01 substrates using monosilane ($\mathrm{SiH_4}$) as the silicon-containing reactant (hydride process). The dopant source was phosphine ($\mathrm{PH_3}$). In contrast with the conventional hydride process described in detail in Ref. 5, in our process a high-resistivity layer of thickness $0.5-0.6\ \mu\mathrm{m}$, grown at $T=1180\ ^\circ\mathrm{C}$, was used to block self-diffusion from the substrate. The temperature was then lowered, and a doped layer of prescribed thickness was overgrown at $T=1040-1060\ ^\circ\mathrm{C}$. This approach enabled us to achieve a uniform distribution of the impurity along the thickness of the epitaxial layers of the final representative structures.

The base region was formed by boron implantation at a dose of $450\ \mu\mathrm{C/cm^2}$ for the creation of a sharp junction, and an implantation dose of $4\ \mu\mathrm{C/cm^2}$ was used to create a graded *p–n* junction.With allowance for the dependence of $U_{\mathrm{c.b.0}}$ on the depth of the collector junction [Eq. (2)], the technological parameters of the dispersion process have been calculated with a view toward establishing an identical value of $x_{j\sigma}$ for both the sharp and the graded *p–n* junctions. The emitter regions in every case were formed by phosphorus diffusion at a temperature of 1040 °C. The gains B_N and the breakdown voltages $U_{\mathrm{c.em.0}}$ were determined on an L2-56 instrument, which is designed to measure the characteristics of semiconductor devices. The voltage $U_{\mathrm{c.b.0}}$ was calculated from experimentally determined values of B_N and $U_{\mathrm{c.em.0}}$ according to Eq. (1). This approach is based on from the fact thatthe space-charge region of the collector for graded *p–n* junctions is much broader than for sharp junctions, and the directly measured value of $U_{\mathrm{c.b.0}}$ can be attributed, not to avalanche breakdown but to breakdown associated with the collector space-charge region coming into contact with the low-resistance substrate.[2] The experimental data and the calculated [from Eq. (2)] values are given in Table I.

Not only do the tabulated data reveal good agreement between the calculated and experimental results for the case of sharp *p–n* junctions , but they also indicate an increase in the experimentally observed values of $U_{\mathrm{c.b.0}}$ above the values calculated from Eq. (2) for graded *p–n* junctions. The discrepancy of the experimentally determined values of $U_{\mathrm{c.b.0}}$ relative to the values calculated from Eq. (2) in the case of graded *p–n* junctions is attributable to the fact that Eq. (2), generally speaking, does not contain any parameters describ-

TABLE I. Experimental parameters and calculated values of $U_{\text{c.b.0}}$ for the collector–base junctions of planar *n*–*p*–*n* transistors.

Epitaxial layer resistivity, $\Omega\cdot$cm	Implantation dose, μC/cm^2	Depth of *p*–*n* junction $x_{j\sigma}$	Values of $U_{\text{c.b.0}}$, calculated	experimental	Factor $\varkappa$
4.5	450	4.25	117.5	123−126	1.029
	4	4.25	117.5	226−231	1.92
3.5	450	4.1	98.7	100−104	1.013
	4	4.1	98.7	190−194	1.93
2.4	450	4.05	89.9	93−95	1.03
	4	3.95	89.8	173−178	1.93

ing the graded junction, i.e., the nature of the distribution of the base impurity. Consequently, Eq. (2) is valid for the calculations of $U_{\text{c.b.0}}$ for sharp junctions.We introduce a correction factor $\varkappa$ and rewrite (2) in the form

$$U_{\text{c.b.0}}=\frac{\varkappa\cdot U_{nn}\cdot x_{j\sigma}}{Z_{pn}}\left[\sqrt{2Z_{pn}/x_{j\sigma}+1}-1\right], \qquad (3)$$

where $\varkappa$ is the breakdown voltage enhancement factor, which is equal to the ratio of the experimentally determined value of $U_{\text{c.b.0}}$ to the value calculated from Eq. (2).

The values of $\varkappa$ are also given in Table I. It is obvious that $\varkappa\approx 1$ in the case of sharp *p*–*n* junctions, and Eq., (3) coincides with (2). The factor $\varkappa$ remains constant in the range of resistivities of the epitaxial layer $2.4-4.5\,\Omega\cdot$cm and is independent of the thickness of the particular epitaxial layer in the interval of thicknesses $12-20$ μm.

These results show that graded *p*–*n* junctions can be used to raise the breakdown voltages $U_{\text{c.b.0}}$ of bipolar planar transistors.Additional investigations will be needed to determine the interrelationship between the breakdown voltage enhancement factor $\varkappa$ and other parameters of *n*–*p*–*n* transistors.

[1] A. V. Greben, *Design of Analog Integrated Circuits* [in Russian] (Énergiya, Moscow, 1976), 435 pp.

[2] S. M. Sze and G. Gibbons, Solid-State Electron. **9**, 831 (1966).

[3] A. Goetzberger, B. McDonald, R. H. Haitz, and R. M. Scarlett, J. Appl. Phys. **34**, 1591 (1963).

[4] L. C. Kao and E. D. Wolley, Proc. IEEE **55**, 1409 (1967).

[5] V. V. Kharchenko, *Production of Epitaxial Silicon Structures with a Controlled Impurity Profile* [in Russian] (FAN, Tashkent, 1989), 168 pp.

Translated by James S. Wood

TECHNICAL PHYSICS VOLUME 43, NUMBER 10 OCTOBER 1998

Focusing of charged-particle beams with energy–angular correlation in a plane-parallel capacitor

S. Ya. Yavor

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

(Submitted January 13, 1998)
Zh. Tekh. Fiz. **68**, 138–139 (October 1998)

The focusing condition is determined for beams having energy–angular correlation with the source and the detector situated on the lower plate of a plane-parallel capacitor. Expressions are given for the second-order spherical aberration and the dispersion. © *1998 American Institute of Physics.* [S1063-7842(98)02810-4]

It has been noted previously[1] that the use of crossed beams in ion–ion collisions gives rise to charged-particle fluxes that manifest an ordered energy distribution across the beam cross section. In Ref. 1 we discussed the distinctive character of the focusing of beams having a linear energy distribution over a cross section in various types of electrostatic and magnetic deflectors. We now examine this problem in the case of a plane-parallel capacitor.

We establish the z axis in the lower plate of a plane-parallel capacitor and orient the y axis perpendicular to it (see Fig. 1). We direct the trajectory of the beam at an angle θ relative to the lower plate. The trajectory equation is then written in the form

$$y=-\frac{eV}{4Eg\cos^2\theta}z^2+z\tan\theta. \tag{1}$$

Here e is the particle charge, E is its initial energy, which has the value $E=0.5mv_0^2$, V is the potential difference on the capacitor plates, and g is the distance between them. The coordinates at which the trajectory reaches its highest point are

$$y_{\max}=\frac{Eg}{eV}\sin^2\theta,\quad z=\frac{Eg}{eV}\sin 2\theta. \tag{2}$$

We denote by z_m the coordinate of the point where the trajectory crosses the lower plate of the capacitor. It is equal to

$$z_m=\frac{2Eg}{eV}\sin 2\theta. \tag{3}$$

We denote by θ_0 the angle corresponding to an axial beam trajectory, $\theta_0\pm\alpha$ are the angles corresponding to edge beam trajectories, E_0 is the energy of an axial particle, and $E=E_0(1\pm k\alpha)$ are the energies of edge particles, where k is the energy–angular correlation coefficient. Here, up to and including α^2, the coordinates z_m of the edge trajectories can have the form

$$z_m=\frac{2E_0g}{eV}[\sin 2\theta_0\pm\alpha(2\cos 2\theta_0+k\sin 2\theta_0)$$
$$+2\alpha^2(-\sin 2\theta_0+k\cos 2\theta_0)]. \tag{4}$$

Here the first term corresponds to the point at which an axial beam trajectory turns back at the lower plate; the focusing condition stipulates that the coefficient of the first power of α be equal to zero. From this condition we deduce the initial angle of inclination of an axial trajectory θ_{0f} at which a particle beam with linear energy–angular correlation is focused onto the lower plate:

$$\tan 2\theta_{0f}=-2/k. \tag{5}$$

It follows from Eq. (5) that such focusing is feasible and is governed by the correlation coefficient.

Let us consider a numerical example. We set the initial angle of opening of the beam equal to $2\alpha=0.04$ and set the relative energy increment over the cross section equal to 2%. The coefficient k then has the value $k=\pm0.5$, the plus or minus sign signifying that the energy increases or decreases, respectively, as the angle increases. Substituting these values of k into Eq., (5), we find that the focusing angle is equal to $\theta_{0f}\approx52°$ in the first case and $\theta_{0f}\approx38°$ in the second case. Hence it follows that z_m is smaller than in the focusing of a monochromatic beam, and in both cases it is given by the expression $z_m\approx0.97eV/(2E_0g)$. The value of $y_{\max}$ increases

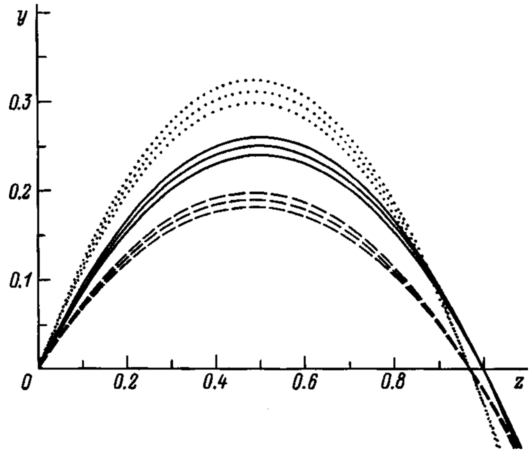

FIG. 1. Trajectories of charged particles in a plane-parallel capacitor for a monochromatic beam ($k=0$, solid curves), for a beam whose energy increases as the angle of opening increases ($k=0.5$ dashed curve), and for a beam whose energy decreases as the angle increases ($k=0.5$ dot-dash curves). The curves are plotted on a scale of $(eV)/(2E_0g)$.

in the first case, $y_{\max} \approx 0.621 eV/(E_0 g)$, and decreases in the second case, $y_{\max} \approx 0.379 eV/(E_0 g)$ (see Fig. 1).

The coefficient of α^2 in Eq., (5) characterizes the second-order spherical aberration C_2 of a beam with energy–angular correlation. Taking (5) into account, we obtain

$$C_2 = -\frac{4E_0 g}{eV} \sin 2\theta_{0f} (1 + 2 \cot^2 2\theta_{0f}). \tag{6}$$

The first term in parentheses corresponds to the second-order spherical aberration of a monochromatic beam. Moreover, the aberration coefficient has acquired an additional term having the same sign. It is evident, therefore, that the spherical aberration increases for a beam with energy–angular correlation. In both cases of our numerical example we have $C_2 = -1.09 eV/(4E_0 g)$, i.e., the absolute value of the coefficient C_2 increases by 9% regardless of whether the energy increases or decreases with increasing angle.

The dispersion equation for a plane-parallel capacitor has the form

$$D = E \frac{\partial z_m}{\partial E} = \frac{2Eg}{eV} \sin 2\theta_{0f}. \tag{7}$$

The dispersion is lower than in the usual situation; the amount of the decrease in our numerical example is 3%. The specific dispersion, equal to $\delta = D/C_2 \alpha^2$, is also lowered, by virtue of a decrease in D and an 11% increase in the coefficient C_2.

We note that the parameters of a plane-parallel capacitor can be calculated similarly for beams having a linear energy distribution over a cross section when the source and the detector lie outside the boundaries of the lower plate.

[1] S. Ya. Yavor, Zh. Tekh. Fiz. **67**(12), 50 (1997) [Tech. Phys. **42**, 1417 (1997)].

Translated by James S. Wood

Influence of defects on the temperature dependence of the density of electrons in the two-dimensional and doped channels of selectively doped $Al_xGa_{1-x}As/GaAs$ heterostructures

S. G. Dmitriev and K. I. Spiridonov

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 141120 Fryazino, Moscow District, Russia

(Submitted February 2, 1998)
Zh. Tekh. Fiz. **68**, 140–142 (October 1998)

The densities of electrons in the two-dimensional and doped channels of selectively doped AlGaAs/GaAs heterostructures are calculated. It is shown that traps and surface states in the AlGaAs layer can change the sign of the temperature dependence of the electron density in the two-dimensional channel. © *1998 American Institute of Physics.* [S1063-7842(98)02910-9]

The densities (and mobilities) of electrons in the two-dimensional (2D) (n_s) and doped (AlGaAs layer) (n) channels of selectively doped heterostructures are traditional characteristics of their quality. It is difficult to measure these parameters directly in the individual channel, and more indirect methods have been proposed in the literature,[1] along with calculations. Here we intend to show that the presence of traps and surface states in the AlGaAs layer influence the temperature dependence of the electron densities in the channels and can cause the temperature dependence of n_s to change sign.

In calculations the doping impurity in the AlGaAs layer (or so-called DX centers[2]) is treated on the basis of the U^--center model (a doubly charged center with negative electron correlation energy) (Ref. 3); the two-dimensional electron gas is treated in a model with two size-quantization levels;[1] the method of calculation is described in Ref. 4. The parameters of the heterostructure are as follows: thicknesses of the spacer (d) and the doped region (l) in $Al_xGa_{1-x}As$: 30 Å and 500 Å (or 350 Å), respectively; $x=0.3$; the dopant (silicon) concentration corresponds to that found in structures used in technology: $N_d=10^{18}\,cm^{-3}$.

The temperature dependence of the static dielectric constant[5,6] is approximated by quadratic splines; the discontinuity of the conduction band at the boundary of the heterojunction is $\Delta E_c=0.60\Delta E_g$ (Ref. 7), where ΔE_g is the difference in the temperature-dependent[6] band gaps of $Al_xGa_{1-x}As$ and GaAs; the distances from the minima of the L and X valleys to the minimum of the Γ valley are $\Delta E_{L\Gamma}=0.284-0.605x$ and $\Delta E_{X\Gamma}=0.476-1.122x+0.143x^2$ (Ref. 8); the effective masses of the density of states for the Γ, L, and X valleys are $m_\Gamma=0.0665+0.0835x$, $m_L=0.56+0.22x$, and $m_X=0.85-0.06x$,[6] respectively; the depths of the shallow (neutral) and deep (negatively charged) levels of the dopant (DX center) are $\varepsilon_0=10\,meV$ (Ref. 3) and $\varepsilon_1=120\,meV$ (Ref. 9), respectively; the degeneracies are $g^+=1$, $g^0=2$, and $g^-=4$. These values of the parameters are consistent with experimental data in the cited papers.

Figure 1 shows the results of calculations of the temperature dependences of the two-dimensional electron densities in the 2D channel $n_s(T)$ (curve *1*), in the Γ valley (curve *2*), and in the Γ, L, and X valleys (curve *3*) of the AlGaAs layer, along with the total electron density in the AlGaAs conduction band and in the dopant shallow level $n(T)$ for the case of a free (not metallized) AlGaAs surface with an effective fixed surface-state charge density $N_s=3\times10^{12}cm^{-2}$ (in units of the electron charge); at this density of surface states the height of the barrier at the outer surface of the AlGaAs layer relative to the Fermi level is of the order of 0.8 eV. The situation here is one in which the exchange of electrons between surface states (all states of the GaAs cap layer on the AlGaAs surface) and the AlGaAs conduction band is impeded by the presence of the barrier, and the surface states are ''frozen out.'' Here (and in other figures) the dashed curve describes the ''freezing out'' at $T=120$ K of charges at the deep level of a DX center.[2,3,9] For comparison the inset shows a plot of $n_s(T)$ for an ordinary shallow doping impurity.

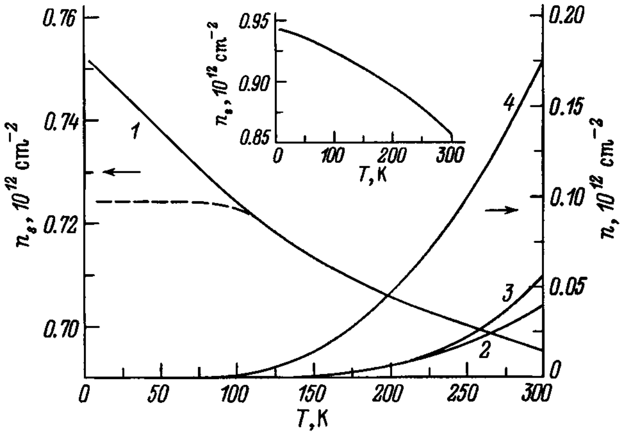

FIG. 1. Temperature dependence of the density of electrons in the 2D $n_s(T)$ and doped $n(T)$ channels of a selectively doped heterostructure, $d=30$ Å, $x=0.3$, $l=500$ Å, $N_d=10^{18}\,cm^{-3}$, $N_s=3\times10^{12}\,cm^{-2}$. *1*) $n_s(T)$; *2*) density of electrons in the AlGaAs Γ valley; *3*) density of electrons in the AlGaAs Γ, L, and X valleys; *4*) total density of electrons in the conduction band and in the shallow dopant level in the AlGaAs layer; the dashed curve corresponds to the ''freezing out'' of charges at the deep level of a DX center at $T=120$ K. Inset: $n_s(T)$ for an ordinary shallow impurity.

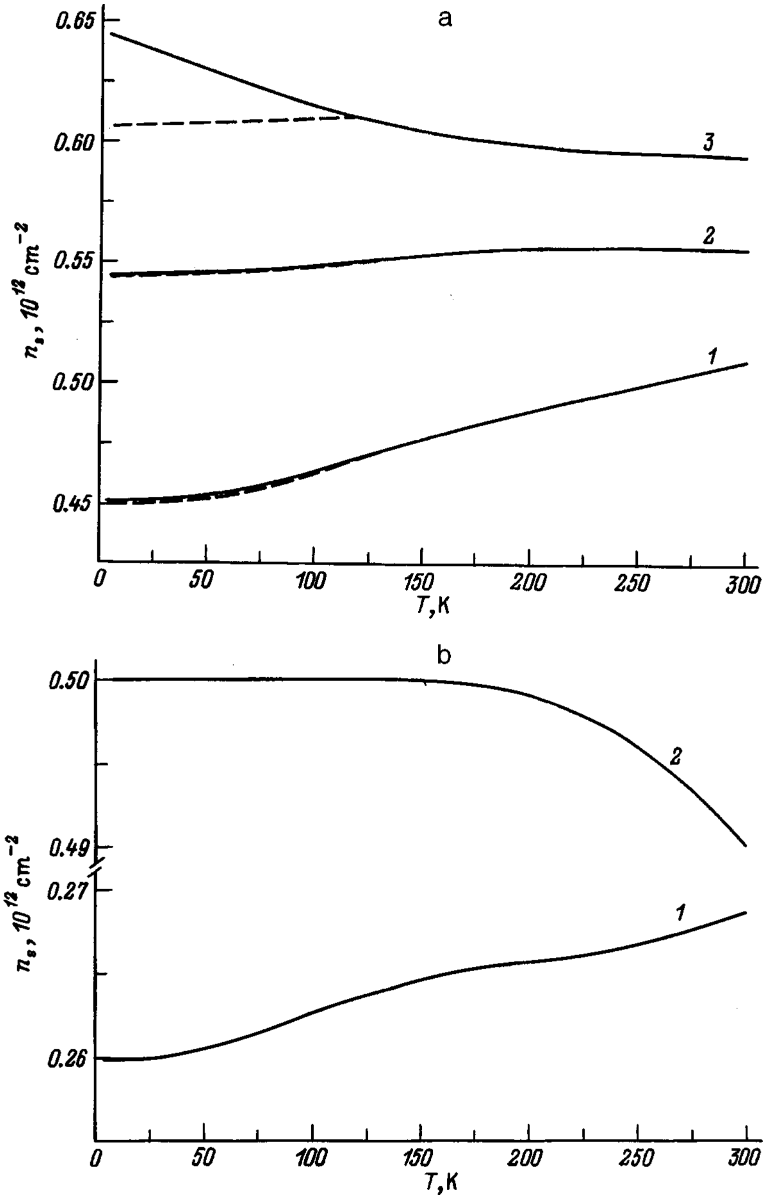

FIG. 2. Influence of traps and surface states in the AlGaAs layer on the temperature dependence $n_s(T)$ and the density of electrons in the 2D channel of a selectively doped heterostructure, $d=30$ Å, $x=0.3$, $N_d=10^{18}\,\mathrm{cm}^{-3}$, $\varepsilon_t=200$ meV. a) Influence of traps on $n_s(T)$, $l=500$ Å: *1)* $N_s=3\times10^{12}\,\mathrm{cm}^{-2}$, $N_t=5\times10^{17}\,\mathrm{cm}^{-3}$; *2)* $N_s=4\times10^{12}\,\mathrm{cm}^{-2}$, $N_t=2\times10^{17}\,\mathrm{cm}^{-3}$; *3)* $N_s=3\times10^{12}\,\mathrm{cm}^{-2}$, $N_t=3\times10^{17}\,\mathrm{cm}^{-3}$. b) Influence of surface states on $n_s(T)$, $l=350$ Å, $N_t=0$: *1)* fixed barrier height $V=1$ eV at the outer surface of the AlGaAs layer; *2)* $N_s=3\times10^{12}\,\mathrm{cm}^{-2}$; the dashed curve corresponds to the ''freezing out'' of charges at the deep level of a *DX* center at $T=120$ K.

The figure illustrates the stability of the sign of the function $n_s(T)$ against various factors: the presence of a deep donor level, the ''freezing out'' of charges there (due to the high, ~300 meV, barrier for electron trapping[2]), etc. This property is associated with the stronger (at a constant density of electrons in the layer) temperature dependence of the Fermi level in the AlGaAs layer than in the 2D channel. Here electrons are transported both from the conduction band and from impurity levels.

The situation changes with the onset of additional transfer channels: traps (density N_t) and surface state levels. Figure 2a illustrates the influence of a trap ($\varepsilon_t=200$ meV) on the behavior of $n_s(T)$ for ''frozen-out'' charges at a surface state. In these structures traps are often associated with native defects (or defect clusters).[6,10] It is seen that the sign of $n_s(T)$ changes and that the magnitude of the effect depends on N_t and N_s (along with other factors).

The influence of surface states on the function $n_s(T)$ for a shorter structure, $l=350$ Å, for which this influence is especially pronounced, is shown in Fig. 2b. The barrier height (relative to the Fermi level) on the outer surface of the AlGaAs layer is equal to $V=1$ eV and is fixed (corresponding to the case of contact with a metal). A similar phenomenon is obviously observed when the exchange of charges with surface states is ''turned on.'' Also shown here for comparison is a graph with ''frozen-out'' surface states, $N_s=3\times10^{12}\,\mathrm{cm}^{-2}$.

We note that for realistic dopant densities the conduction band ''merges'' with the impurity band, and a Mott transition takes place[6] (within the framework of the given model the total electron density in the conduction band and in a shallow level is represented by curve *4* in Fig. 1). In this case, owing to the presence of an electron mobility threshold in the conduction band, an activation-type process can occur in the AlGaAs layer.

The existence of such a threshold can be observed in the relaxation processes following the pumping of electrons into the AlGaAs channel,[11] while masking effects associated with impurity charge transfer in the active GaAs layer are usually slower.[12]

In closing, we note that the temperature variation of the density of electrons in the 2D channel of a selectively doped heterostructure is relatively small — of the order of a few percent. The density normally increases as the temperature is lowered, but the presence of defects and surface states in the AlGaAs layer can cause this temperature dependence to change sign.

[1] M. Shur, *GaAs Devices and Circuits* (Plenum Press, New York, 1987; Mir, Moscow, 1991), 632 pp.
[2] P. M. Mooney, J. Appl. Phys. **67**, R1 (1990).
[3] T. N. Theis *et al.*, J. Electron. Mater. **20**, 35 (1991).
[4] S. G. Dmitriev *et al.*, Pis'ma Zh. Tekh. Fiz. **22**(11), 13 (1996) [Tech. Phys. Lett. **22**(6), 436 (1996)].
[5] G. A. Samara, Phys. Rev. B **27**, 3494 (1983).
[6] L. Pavesi and M. Guzzi, J. Appl. Phys. **75**, 4779 (1994).
[7] S. R. Smith *et al.*, J. Appl. Phys. **75**, 1010 (1994).
[8] S. Adachi, J. Appl. Phys. **58**, R1 (1985).
[9] G. Oelgart *et al.*, Semicond. Sci. Technol. No. 5, 894 (1990).
[10] P. Krispin *et al.*, J. Appl. Phys. **77**, 5773 (1995).
[11] V. I. Borisov, S. G. Dmitriev, V. E. Lyubchenko *et al.*, Fiz. Tekh. Poluprovodn. **28**, 1199 (1994) [Semiconductors **28**, 683 (1994)].
[12] A. L. Powell *et al.*, Phys. Rev. Lett. **67**, 3010 (1991).

Translated by James S. Wood