Propagation of a spatial soliton in a system of tunneling-coupled optical waveguides with a variable coupling coefficient

F. Kh. Abdullaev

S. S. Starodubtsev Physicotechnical Institute; Scientific-Production Unit ''Fizika-Solntse,'' Uzbekistan Academy of Sciences, 700084 Tashkent, Uzbekistan
(Submitted October 24, 1996; resubmitted January 27, 1998)
Zh. Tekh. Fiz. **68**, 1–4 (June 1998)

The problem of the propagation of a spatial soliton in a system of tunneling-coupled optical waveguides is studied. The tunneling coupling coefficient is assumed to be modulated in a transverse direction. The adiabatic dynamics is studied for the case of periodic modulation of the tunneling coupling coefficient. The stationary points for the beam parameters are found. The effective potential is calculated for the center of the beam in the equivalent-particle model. The condition for resonance emission of waves by a soliton is obtained. © *1998 American Institute of Physics.* [S1063-7842(98)00106-8]

INTRODUCTION

Systems of tunneling-coupled nonlinear optical waveguides are attracting attention in connection with possible applications as optical decouplers, components in optical logic circuits, and so on.[1,2] Such configurations can be obtained, for example, by preparing flat waveguides on a $Ga_xAl_{1-x}As$ substrate.[3] This material possesses positive Kerr nonlinearity when the wave frequency is less than half the band gap. An important feature is that soliton propagation regimes of powerful optical beams—spatial optical solitons—are possible in such systems.[4,5] Such systems also have the advantage that it is possible to obtain strong nonlinear effects in them by using weak nonlinear materials in separate waveguides. For applications it is of interest to develop methods for controlling the parameters of spatial solitons in these systems. In Ref. 6 it is suggested that variation of the tunneling coupling coefficient between the waveguides be used for such purposes. Using a variational approach and the continuum approximation the authors were able to investigate the effect of linear and quadratic variation of the tunneling coupling coefficient. For their approach the choice of trial function is critical. Since the continuum approximation leads to a perturbed nonlinear Schrödinger equation, the method of the inverse problem in soliton theory provides a systematic approach.

In the present paper we shall study the effect of a periodic variation of the tunneling-coupling coefficient on the propagation of a spatial soliton in a stack of nonlinear planar waveguides. The investigation will employ perturbation theory based on the inverse-problem method. The problem exhibits special features compared with that studied in Ref. 6, specifically, in our case resonance emission of waves by a spatial soliton appears.

DESCRIPTION OF THE MODEL

The system of equations describing the propagation of a wave in an infinite system of tunneling-coupled nonlinear planar waveguides is[6,7]

$$-iu_{nz} = v_{n,n+1/2}u_{n+1} + v_{n-1/2,n}u_{n-1} + |u_n|^2 u_n. \tag{1}$$

Here $\bar{u}(z,n)$ is the dimensionless intensity of the electric field of the wave and $v_{n,m}$ is the coupling coefficient between the waveguides. An approximate expression for $v_{n,m}$ is[8]

$$v_{n,m} \sim \exp[-\gamma(n-m)].$$

We shall employ a continuum approximation to describe the evolution of a wave in the system of waveguides. This approximation can be used to study the variations of a wave on scales much larger than the distance h between the waveguide centers.[9] Then u_n and $v_{n,m}$ can be expanded in a Taylor series

$$u_{n\pm1} = u(x) \pm hu_x + \frac{h^2}{2}u_{xx} + \dots,$$

$$v_{n,n\pm1/2} = v(x) \pm \frac{h}{2}v_x + \frac{h^2}{8}v_{xx} + \dots.$$

Substituting these expressions in Eq. (1) we obtain for $u(z,x)$ the equation

$$-iu_z = 2vu + h^2\left(vu_{xx} + v_xu_x + \frac{1}{4}v_{xx}u\right) + |u|^2u. \tag{2}$$

We shall study the case of periodic modulation of the tunneling-coupling parameter along the direction $x, v = 1 + \varepsilon\sin(ax)$. Setting $h = 1$ and changing to the new variables $y = x/\sqrt{2}$, $z = 2t$, $\alpha = \sqrt{2}a$, we obtain finally the equation

$$iu_t + u_{yy} = 2|u|^2u = \varepsilon R(u,y)$$

$$= -\varepsilon\sin(\alpha_y)u_{yy} - \varepsilon\alpha\sin(\alpha_y)u_y$$

$$+ \frac{1}{4}\varepsilon\alpha^2\sin(\alpha_y)u - 4\varepsilon\sin(\alpha_y)u. \tag{3}$$

As a result, the problem reduces to investigation of the nonlinear Schrödinger equation with a periodic perturbation. As follows from Eq. (2), the perturbation is conservative—there exists a conserved energy integral

$$N = \int_{-\infty}^{\infty} |u|^2 dx = \text{const.}$$

We note that the influence of a local nonuniformity of the refractive index in a stack of waveguides on the dynamics of a spatial soliton was recently analyzed on the basis of a continuum approximation and the inverse-problem method for solitons.[10] The corresponding perturbation on the right-hand side of Eq. (3) has the form $R(u,x) = \varepsilon \delta(x) u$. Numerical modeling confirmed that such an approach gives a good description of beam dynamics.

ADIABATIC DYNAMICS OF A SPATIAL SOLITON IN A SYSTEM OF COUPLED WAVEGUIDES

We shall study the evolution of a single spatial soliton, propagating in a system of tunneling-coupled waveguides. When there is no modulation, i.e., $v = 0$, the soliton solution can be written in the form

$$u_s = 2i\eta \, \text{sech}[2\eta(y-\zeta)] \exp(-2i\xi y - 4i(\eta^2 - \xi^2)t).$$

Here η is the amplitude of the soliton and $\zeta = -4\xi t$, where $\xi = -v/4$ is the coordinate of the center and v is the velocity of the soliton. We note that in dimensional variables the soliton velocity is the angle $\psi = \sin^{-1}(\xi/2)$ of propagation of a beam in the stack of waveguides. In this section we shall study the effect of large- and small-scale modulations of the tunneling-coupling coefficient on the dynamics of a spatial soliton.

a) *Large-scale modulations of the tunneling-coupling coefficient.* Using the perturbation theory,[11,12] we find that the amplitude of the soliton is conserved $\eta = \text{const}$. This is a reflection of the previously noted fact that for Eqs. (2) and (3) there exists an integral of the motion, $N = \int_{-\infty}^{\infty} |u|^2 dx$. We obtain the following system of equations for the velocity and coordinate of the center of the soliton:

$$\frac{d\xi}{dt} = \frac{\pi\alpha^2\varepsilon}{2\eta \, \sin h(\pi\alpha/4\eta)} \cos(\alpha\zeta) \left(\frac{\eta^2}{3} + \xi^2 + \frac{\alpha^2}{48} - 1\right), \tag{4}$$

$$\frac{d\zeta}{dt} = -4\xi \left[1 + \frac{\pi\alpha\varepsilon \, \sin(\alpha\zeta)}{4\eta \, \sin h(\pi\alpha/4\eta)}\right]. \tag{5}$$

For low soliton velocities $\xi \ll 1$, $\xi^2 \ll \eta^2$ ("heavy" soliton) the system (4) and (5) is equivalent to the problem of the motion of a particle of unit mass in a periodic potential $u(\zeta)$

$$u(\zeta) = \frac{2\pi\alpha\varepsilon \, \sin(\alpha\zeta)}{\eta \, \sin h(\pi\alpha/4\eta)} \left(\frac{\eta^2}{3} + \frac{\alpha^2}{48} - 1\right). \tag{6}$$

In the general case the effective potential depends on the soliton velocity ξ. There are examples of velocity-dependent potentials in physics. For example, velocity-dependent potentials have been used to describe the nucleon interactions in nuclei.

Let us discuss the role of length scales in the problem. There are two characteristic scales: The modulation period $\lambda = 2\pi/\alpha$ of the tunneling-coupling parameter and the size

of the soliton $l = 1/2\eta$. The amplitude of the effective potential depends strongly on the ratio l/λ. For $l/\lambda \ll 1$ the effective potential equals

$$u(\zeta) \approx \frac{8}{3}\varepsilon \, \sin(\alpha\zeta)\eta^2 + 0(l^2/\lambda^2).$$

In this limit the effective potential is virtually independent of the ratio of the length scales and is proportional to the squared amplitude of the soliton.

For $l/\lambda \gg 1$ we have

$$u(\zeta) \approx \frac{\pi\varepsilon\alpha^3}{48\eta} \exp(-\pi\alpha/4\eta)\sin(\alpha\zeta).$$

One can see that the amplitude of the effective potential is exponentially small for solitons whose width is greater than the modulation period. There exists a corridor of parameter values near $\alpha \sim \varepsilon^{-1/3}$ where the amplitude of the potential $\lesssim 1$.

It is difficult to find the general solutions of the system (4) and (5). Let us analyze the behavior of the system in the phase plane. For this, let us find the stationary points for the system (4) and (5). Let us rewrite the system of equations in the form

$$\frac{d\xi}{dt} = A \, \cos(\alpha\zeta)(\xi^2 - B), \tag{7}$$

$$\frac{d\zeta}{dt} = -4\xi[1 + C \, \sin(\alpha\zeta)], \tag{8}$$

where

$$A = \frac{\varepsilon\pi\alpha^2}{2\eta \, \sin h(\pi\alpha/4\eta)},$$

$$B = -\left(\frac{1}{3}\eta^2 + \frac{\alpha^2}{48} - 1\right), \quad C = \frac{A}{2\alpha}.$$

We note that A, B, and C are decreasing functions of α. Then the conditions for the stationary points can be written in the form

$$\xi = 0, \quad \alpha\zeta = \pi/2 + n\pi; \quad \xi = \pm\sqrt{B},$$

$$\alpha\zeta = -\sin^{-1}(1/c)(-1)^n, \quad n = 1, 2, \dots. \tag{9}$$

It can be shown by analyzing the linearized system (4) and (5) that the points $\alpha\zeta = \pi/2 + 2n\pi$ are stable, while the points $\alpha\zeta = -\pi/2 + 2n\pi$ are unstable. The phase portrait of the system is presented in Fig. 1. Since the phase trajectories are periodic, we limited the figure to the values $-\pi < \alpha\zeta < 3\pi/2$. One can see regions of finite motion of an effective particle, which correspond to oscillations of a spatial soliton as a whole in the system of tunneling-coupled waveguides, and regions of unbounded motion, which correspond to a growing deflection of the beam. We note that two types of effects materialize here. Effects of the first type are due to the linear periodic potential, and effects of the second type are due to periodic modulation of the dispersion. The dispersion perturbation dominates for large ξ in $R(u) \rightarrow v(x)u_{xx} \sim \xi^2 v(x)u$, which one can see in the phase portrait—region 1. For low velocities ξ we have soliton oscillations around
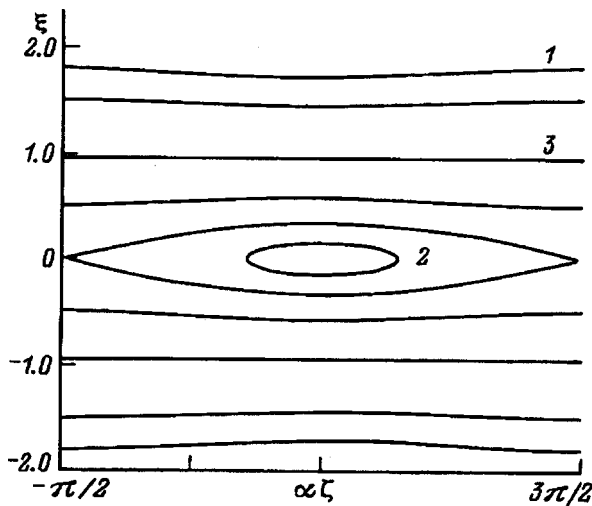
FIG. 1. Phase portrait of the system (4) and (5) with $\varepsilon = 0.1$, $2\eta = 1$, $\alpha = 1$.

the bottom of each well of the periodic potential—region 2. The line 3, where the soliton velocity is constant $\xi = \pm\sqrt{B}$, is of special interest. Here the effects of dispersion and potential disturbances on the motion of the spatial soliton are balanced. Since the soliton velocity ξ is related with the beam propagation angle in the waveguide stack, we conclude that a variation of the tunneling-coupling constant between the waveguides leads to periodic variations of the location where the soliton beam emerges from the waveguide stack. The latter effect can be used to develop optical devices based on spatial optical solitons.

b) *Small-scale modulations of the tunneling-coupling coefficient.* A different approach must be used to describe the limit $\alpha \gg \eta$, specifically, one or another variant of the method of averaging over fast variations of $\nu(y)$. Here we shall use the method suggested in Ref. 13. The aim is to obtain an equation for the slowly varying part of the wave field, using the expansion

$$u = U + A\,\cos(\alpha_y) + B\,\sin(\alpha_y)$$
$$+ C\,\cos(2\alpha_y) + D\,\sin(2\alpha_y) + \ldots, \tag{10}$$

where the functions U, A, B, C, and D are assumed to be slowly varying over distances $\sim 1/\alpha$.

A closed equation for $U(y,t)$ can be obtained by writing out the equations for these functions and using asymptotic expansions in $1/\alpha$. Assuming that $\varepsilon < 1/\alpha$ and $\alpha \gg 1$, we obtain

$$iu_t + u_{yy} + 2|u|^2 u = -\frac{\varepsilon^2\alpha^2}{16}u - \frac{\varepsilon^2}{16}$$
$$\times (2iu_t - 6u_{yy} + 24|u|^2 u). \tag{11}$$

Therefore the beam dynamics in a system of tunneling-coupled waveguides is described by a renormalized nonlinear Schrödinger equation. The first term on the right-hand side of the equation describes the variation of the soliton phase and the remaining terms describe the variation of the soliton width. It follows from Eq. (11) that the soliton width increases

$$d' = d_0\sqrt{\frac{1+\varepsilon^2/8}{1-3\varepsilon^2/8}}. \tag{12}$$

The adiabatic soliton dynamics studied in this section is valid when radiation effects can be neglected. We shall briefly discuss the effect of the emission of waves by a spatial soliton.

Analysis based on solving the equations for the Jost coefficient $b(\lambda)$ of an associated linear spectral problem, where λ is the spectral parameter[14] ($\lambda = 2k$ and k is the wave number), shows that wave emission by a soliton arises. The radiation is concentrated around two spectral points

$$\lambda_{1,2} = -\xi \pm \sqrt{\alpha\xi - \eta^2}. \tag{13}$$

The group velocity of the emitted waves is $v = -4\lambda_{1,2}$. Maximum emission is observed when

$$\alpha/2 - \sqrt{\alpha\xi - \eta^2} \ll 1. \tag{14}$$

For soliton amplitudes η small compared with the velocity ξ and modulation frequency α we find that the emission is maximum when

$$\alpha = v_s. \tag{15}$$

One can see that resonance radiation occurs when the modulation frequency equals the soliton velocity. On account of the strong soliton-radiation coupling the soliton decays as it propagates in the waveguide stack. In Ref. 15 a condition identical to Eq. (15) was obtained from qualitative considerations in a description of numerical experiments on the propagation of a nonlinear Schrödinger soliton in a periodically nonuniform medium. Since the soliton velocity is actually the propagation angle of a beam in a system of waveguides, we find that intense emission should occur for propagation angles

$$\psi_c = \sin^{-1}(\alpha/8). \tag{16}$$

Our results should be correct for these values of the parameters. Specifically, intense emission leads to the appearance of radiation damping and deceleration of the soliton. A complete analysis of the problem of wave emission by a spatial soliton in a system of waveguides and a calculation of the radiation decay length of a soliton will be performed separately.

CONCLUSIONS

In this paper the propagation of a spatial soliton in a system of tunneling-coupled waveguides with a periodically varying tunneling-coupling coefficient was investigated. The problem was studied in the long-wavelength approximation, where it reduces to investigation of the propagation of a nonlinear Schrödinger soliton with periodic dispersion and potential perturbation. The adiabatic dynamics of a soliton was studied. It was shown that the motion of the soliton center is described by the motion of a particle of unit mass in a velocity-dependent effective potential which is a periodic function of the coordinate of the soliton. The stationary points for the soliton were found. It was shown that the predicted effects can be used to control the parameters of a beam in a system of waveguides.

The effect of small-scale modulations of the tunneling-coupling coefficient on the propagation of a spatial soliton was also investigated. The method of averaging for nonlinear partial differential equations was used to derive a renormalized nonlinear Schrödinger equation for a beam and to find the change in the parameters of the soliton.

The effects due to wave emission by a spatial soliton were analyzed. The condition of resonance emission of waves by a soliton was found, and the propagation angle for which the spatial soliton decays resonantly was calculated.

[1] A. Maĭer, Kvant. Élektron. (Moscow) **9**, 2296 (1982) [Sov. J. Quantum Electron. **12**, 1490 (1982)].

[2] S. M. Jensen, IEEE J. Quantum Electron. **QE-18**, 1580 (1982).

[3] D. B. Mortimore and J. M. Arkwright, Appl. Opt. **29**, 1814 (1990).

[4] D. J. Christodoloudies and R. I. Joseph, Opt. Lett. **13**, 794 (1989).

[5] A. B. Aceves, C. de Angelis, S. Trillo, and S. Wabnitz, Opt. Lett. **19**, 332 (1994).

[6] R. Muschall, S. Schmidt-Hattenberger, and F. Lederer, Opt. Lett. **19**, 323 (1994).

[7] S. S. Abdullaev and F. Kh. Abdullaev, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **25**, 789 (1980).

[8] D. Marcuse, *Integrated Optics*, IEEE Press, New York, 1973; Mir, Moscow, 1974.

[9] S. Kawakami and H. A. Haus, IEEE J. Lightwave Technol. **LT-4**, 160 (1986).

[10] W. Krolikowski and Yu. S. Kivshar, Phys. Rev. E **54**, (1996).

[11] V. I. Karpman and E. P. Maslov, Zh. Éksp. Teor. Fiz. **73**, 537 (1977) [Sov. Phys. JETP **46**, 281 (1977)].

[12] D. J. Kaup and A. C. Newell, Proc. R. Soc. London, Ser. A **361**, 413 (1978).

[13] Yu. S. Kivshar and S. K. Turitsyn, Phys. Rev. E **49**, R2536 (1994).

[14] F. Kh. Abdullaev, J. G. Caputo, and N. Flytzanis, Phys. Rev. E **52**, 1552 (1994).

[15] R. Scharf and A. R. Bishop, Phys. Rev. A **43**, 6535 (1994).

Translated by M. E. Alferieff

Critical fields and currents in a weakly nonlinear medium near the percolation threshold

A. A. Snarskiĭ and S. I. Buda

National Technical University, 252056 Kiev, Ukraine
(Submitted June 30, 1997; resubmitted December 19, 1997)
Zh. Tekh. Fiz. **68**, 5–8 (June 1998)

For fields above a critical value the expansion of the conductivity in powers of the field ceases to be valid and the weak-nonlinearity approximation no longer works. The density behavior of the critical fields in strongly inhomogeneous media near the percolation threshold is found on the basis of two criteria—an average criterion and a local criterion. The parameter values of the medium for which crossover—a change of the critical behavior—occurs are determined. Similar calculations are performed for the critical currents. © *1998 American Institute of Physics.* [S1063-7842(98)00206-2]

Nonlinear phenomena play a special role in the study of the transport properties of strongly inhomogeneous media near the percolation threshold p_c. This is because near p_c there exist in a medium locations where the current density and voltage drop are substantial, making it necessary to take into account the deviations from the linear Ohm's law. To a first approximation there appears in Ohm's law, besides a term linear in the field, a nonlinear cubic term

$$\mathbf{j}(\mathbf{r}) = \sigma(\mathbf{r})\mathbf{E}(\mathbf{r}) + \chi(\mathbf{r})|\mathbf{E}(\mathbf{r})|^2\mathbf{E}(\mathbf{r}), \tag{1}$$

where $\mathbf{j}(\mathbf{r})$ is the current density, $\mathbf{E}(\mathbf{r})$ is the electric field, $\sigma(\mathbf{r})$ the ordinary conductivity, and $\chi(\mathbf{r})$ is a constant characterizing the cubic nonlinearity.

Just as in the linear case ($\chi = 0$), to describe the effective properties of a randomly inhomogeneous medium one introduces effective transport coefficients, which by definition relate the volume averages of the field and current density

$$\langle \mathbf{j} \rangle = \sigma_e \langle \mathbf{E} \rangle + \chi_e |\langle \mathbf{E} \rangle|^2 \langle \mathbf{E} \rangle, \tag{2}$$

where $\langle \ldots \rangle = V^{-1}\int \ldots dV$ and the size $\sim V^{1/3}$ of the averaging region is assumed to be much larger than the characteristic self-averaging length—the correlation length ξ.

One of the main questions that must be answered in describing the effective properties of a medium taking into account the nonlinearity is the question of the values of the electric field and current density for which the relation (2) is still valid. It is assumed[1,2] that the relation (2) remains valid if

$$\langle j \rangle < \langle j \rangle_c, \quad \langle E \rangle < \langle E \rangle_c, \tag{3}$$

where $\langle E \rangle_c$ and $\langle j \rangle_c$ are the so-called critical field and critical current density determined from the condition that the first term in Eq. (2) equals the first term:

$$\langle E \rangle_c = \sqrt{\sigma_e/\chi_e}, \quad \langle j \rangle_c = \sqrt{\sigma_e^3/\chi_e}. \tag{4}$$

Many papers have been devoted to the calculation of the critical field and current density in percolation media (see, for example, Refs. 1–7). Some of these works are based on

the analogy between the critical behavior of χ_e and the relative spectral density C_e of $1/f$ noise, which were first established in Refs. 8 and 9.

As a rule, two limiting cases are studied. Above the percolation threshold, where the density of the high-conductivity phase $p > p_c$ and the medium contains an infinite cluster consisting of a phase with conductivity σ_1 ($\sigma_1 \gg \sigma_2$), the N/I (normal metal–insulator) case is studied. The low-conductivity phase is assumed to be an ideal insulator—$\sigma_2 = 0$. Below the percolation threshold, $p < p_c$, the S/N (superconductor–normal metal) case is studied. In this case the system does not contain an infinite cluster, and the current necessarily flows through segments consisting of the low-conductivity phase. It is assumed that the entire voltage drop occurs across these segments, i.e., the high-conductivity phase is an ideal conductor $\rho_1 = 1/\sigma_1 = 0$.

In Refs. 3–7 the following were obtained for these two cases:

$$\langle E \rangle_c \sim \tau^{-t+\nu(d-1)}, \quad \langle j \rangle_c \sim \tau^{\nu(d-1)}, \quad p > p_c, \quad N/I, \tag{5}$$

$$\langle E \rangle_c \sim |\tau|^\nu, \quad \langle j \rangle_c \sim |\tau|^{-q+\nu}, \quad p < p_c, \quad S/N, \tag{6}$$

where $\tau = (p - p_c)/p_c$ is the proximity to the percolation threshold, t and q are, respectively, the critical exponents of the effective linear conductivity above and below threshold, ν is the critical exponent of the correlation length $\xi \approx a_0|\tau|^{-\nu}$, and a_0 is the minimum length in the medium, which for a network problem is the bond length in the network. The analogy established in Refs. 8 and 9 between χ_e and the relative spectral density C_e of $1/f$ noise and the fact that taking the finite ratio $h = \sigma_1/\sigma_2$ into account may be essential for describing the critical behavior of C_e (in contrast to σ_e)[10–12] indicate that $h = \sigma_2/\sigma_1 \neq 0$ must also be taken into account when calculating the critical fields and currents.

Models of a percolation structure above and below p_c can be used to determine the fields and currents in a strongly inhomogeneous medium (Fig. 1).[13–15] The principal elements of this structure are a bridge consisting of a high-conductivity phase and a low-conductivity interlayer. Taking
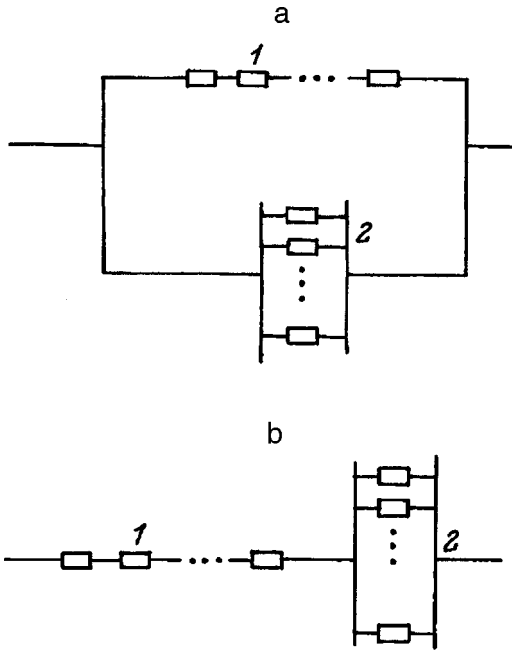
a

b

FIG. 1. Hierarchical model of a percolation structure with a finite ratio $h = \sigma_2/\sigma_1$. 1—Bridge, N_1 high-conductivity resistances connected in series; 2—interlayer, N_2 low-conductivity resistances connected in parallel. According to Refs. 13 and 14, $N_1 \sim |\tau|^{-t + \nu(d-2)}$, $N_2 \sim |\tau|^{-q - \nu(d-2)}$.

into account the analogy between the weak nonlinearity and $1/f$ noise (χ_e and C_e), we shall confine our attention to the two structural elements indicated above, since these elements give the main terms in C_e and χ_e for $h \neq 0$.[14] The N/I and S/N approximations correspond to taking into account only one element of the structure: above p_c only the bridge, and below p_c only the interlayers.

We now require that besides the condition (3) a similar condition hold locally, i.e., that at each point of each phase the local field and current density be less than the local critical values

$$E_i(r) < E_{ci}, \quad j_i(r) < j_{ci}, \quad i = 1, 2, \tag{7}$$

where for each phase the critical local field and current density are determined from a local law (1):

$$E_{ci} = \sqrt{\sigma_i/\chi_i}, \quad j_{ci} = \sqrt{\sigma_i^3/\chi_i}, \quad i = 1, 2. \tag{8}$$

In contrast to the conditions (3), which can be called an average criterion, the conditions (8) can be called a local criterion. Of course, there is no guarantee that the law (8) holds just because the condition (3) is satisfied. This means that in some cases the relations (5) and (6) should be replaced by different relations.

As one can see from Fig. 1a, the voltage across the bridge equals the external voltage. Assuming that the electric field on the bridge equals the critical field E_{c1} and expressing the external voltage drop in terms of the average field, we find the average field corresponding to the local criterion (7) of weak nonlinearity to be $\langle E \rangle_{c1} = (N_1 a_0/\xi) E_{c1}$. Similar arguments for the interlayer give $\langle E \rangle_{c2} = (a_0/\xi) E_{c2}$ or

$$\langle E \rangle_{c1} = \sqrt{\sigma_1/\chi_1}\,\tau^{-t+\nu(d-1)}, \quad \langle E \rangle_{c2} = \sqrt{\sigma_2/\chi_2}\,\tau^\nu. \tag{9}$$

The critical current densities for the fields (9) are

$$\langle j \rangle_{c1} = \sigma_1 \sqrt{\sigma_1/\chi_1}\,\tau^{\nu(d-1)}, \quad \langle j \rangle_{c2} = \sigma_2 \sqrt{\sigma_2/\chi_2}\,\tau^{\,t+\nu}. \tag{10}$$

For $p > p_c$ the average fields and current densities for which relations (7) still hold and Eq. (2) can be used are the smaller of the two possible values:

$$\{E\}_c = \min\{\langle E \rangle_{c1}; \langle E \rangle_{c2}\}, \quad \{j\}_c = \min\{\langle j \rangle_{c1}; \langle j \rangle_{c2}\}. \tag{11}$$

Below the percolation threshold the entire current flows mainly through the bridge and interlayer (Fig. 1b). As the voltage increases, the local criteria for weak nonlinearity (7) are violated here first. Further calculations are most conveniently performed in terms of the resistivities. Then, to within a cubic term, the relation between the current density and electric field will have the form

$$\mathbf{E}(\mathbf{r}) = \rho(\mathbf{r})\mathbf{j}(\mathbf{r}) + \mu(\mathbf{r})|\mathbf{j}(\mathbf{r})|^2 \mathbf{j}(\mathbf{r}), \tag{12}$$

where to this accuracy $\rho_i = 1/\sigma_i$ and $\mu_i = -\chi_i/\sigma_i^4$.

The local critical current densities are determined by equating the terms in Eq. (2), i.e., $j_{ci} = \sqrt{\rho_i/\mu_i}$. Knowing the cross-sectional area of the bridge a_0^{d-1} and the interlayer $a_0^{d-1}|\tau|^{-q-\nu(d-2)}$, we find the average critical current densities from the condition that the current $\langle j \rangle \xi^{d-1}$ through the entire medium equals the critical currents for the bridge and interlayer ($j_{c1} a_0^{d-1}$, $j_{c2} a_0^{d-1}|\tau|^{-q-\nu(d-2)}$):

$$\langle j \rangle_{c1} = \sigma_1 \sqrt{\sigma_1/\chi_1}\,|\tau|^{\nu(d-1)},$$

$$\langle j \rangle_{c2} = \sigma_2 \sqrt{\sigma_2/\chi_2}\,|\tau|^{-q+\nu}. \tag{13}$$

From the condition that the total voltage drop across the correlation length equals the sum of the voltages across the bridge and interlayer we obtain the critical average fields

$$\langle E \rangle_{c1} = \frac{\sigma_1}{\sigma_2}\sqrt{\frac{\sigma_1}{\chi_1}}\,|\tau|^{q+\nu(d-1)}, \quad \langle E \rangle_{c2} = \sqrt{\frac{\sigma_2}{\chi_2}}\,|\tau|^\nu. \tag{14}$$

Both above and below the percolation threshold the critical current density and field are the smaller of the expressions in Eqs. (13) and (14) (see Eq. (11)).

Depending on the values of the parameters τ, $h = \sigma_2/\sigma_1$, and $H = \chi_2/\chi_1$ the minimum values in Eq. (11) can be the first or the second values of the field and current density. Equating the expressions for the fields in (9) and current densities in Eq. (10), for the case $p > p_c$ we obtain $\sqrt{(\sigma_2\chi_1)/(\sigma_1\chi_2)}\,\tau^{t-\nu(d-2)} = 1$, or, equivalently,

$$\tau_0 = (H/h)^{1/2(t-\nu(d-2))}, \quad h = \sigma_2/\sigma_1, \quad H = \chi_2/\chi_1. \tag{15}$$

Crossover—a change in the critical behavior of the critical field current density—occurs when τ passes through the value τ_0. The critical current density depends on τ as $\sim \tau^{t+\nu}$ for $\tau < \tau_0$ and $\sim \tau^{\nu(d-1)}$ for $\tau > \tau_0$. The critical field also undergoes crossover: For $\tau < \tau_0$ it is $\sim \tau^\nu$ and for $\tau > \tau_0$ it is proportional to $\tau^{-t+\nu(d-2)}$.

The region 1 indicated in Fig. 2a corresponds to the case when $\{E\}_c = \langle E \rangle_{c2}$ from Eq. (9) and $\{j\}_c = \langle j \rangle_c$ from Eq. (10). The region 2 corresponds to the situations when $\{E\}_c = \langle E \rangle_{c1}$, $\{j\}_c = \langle j \rangle_{c1}$, i.e., the critical exponents of the current density and electric field are identical to Eq. (5).
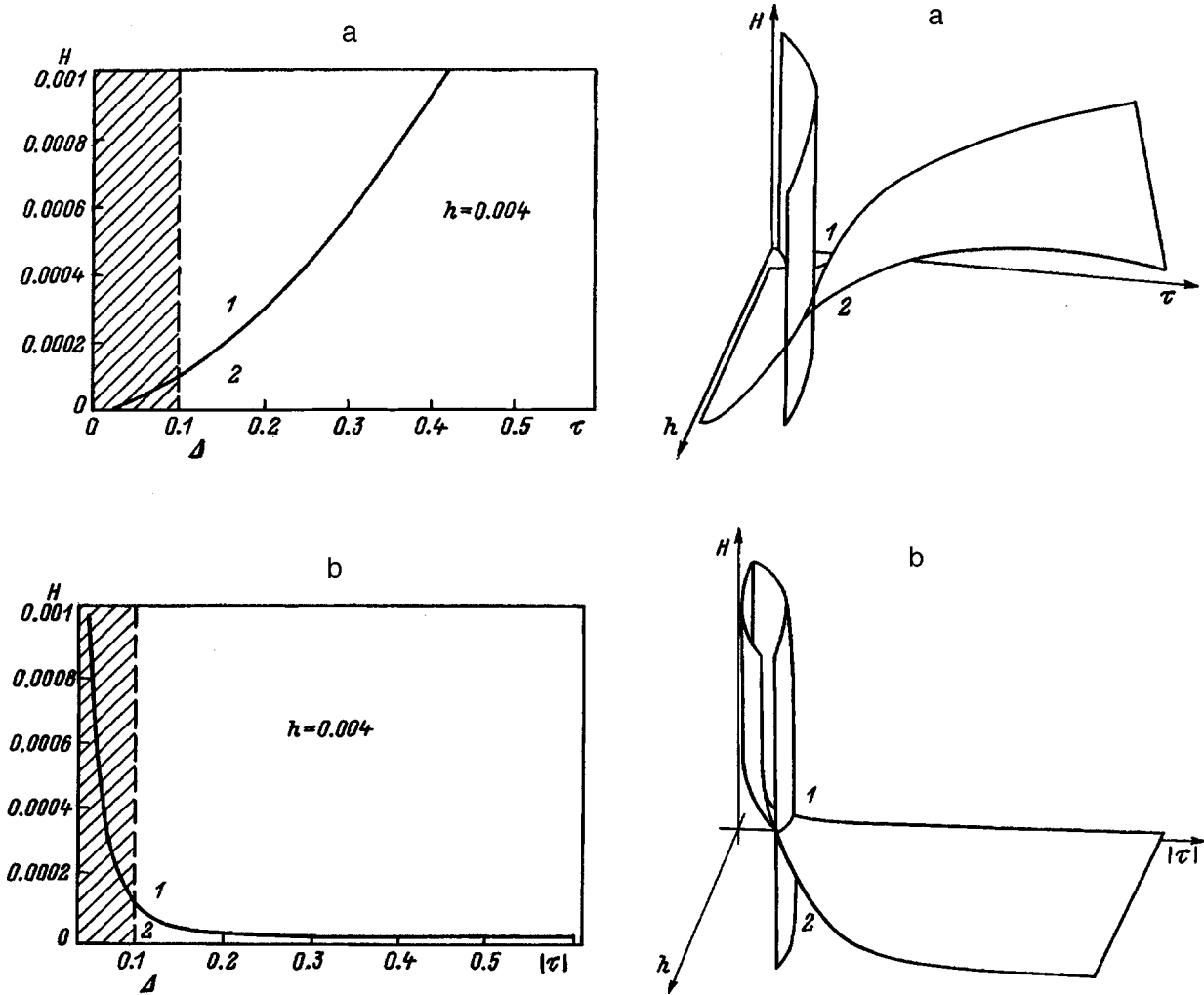
FIG. 2. Surfaces bounding the regions of the constants ratios $h = \sigma_2/\sigma_1$ and $H = \chi_2/\chi_1$ for a weakly nonlinear medium where the results of the standard approach[3-7] are valid (N/I and S/N cases (5) and (6)) and the critical current densities and fields obtained on the basis of a local criterion. Region 2—standard approach, 1—results of this work. The hatched region of values of τ corresponds to the region of broadening Δ. The quantity Δ is determined on the basis of the analogy between the effective properties of the weakly nonlinear medium and the effective $1/f$ noise,[10] $\Delta = h^{1/(t+q)}$.

For $p < p_c$ the crossover condition for the critical current density and electric field can be found similarly to Eq. (15) and has the form

$$|\tau_0| = (h^3/H)^{1/2(q + \nu(d-2))}. \tag{16}$$

The surface bounding the regions of σ_i and χ_i where the results of the average criterion (6) are valid is shown in Fig. 2b. For the region 1 the results of the standard approach are valid, i.e., Eq. (6). The region 2 corresponds to the case when $\{j\}_c = \langle j \rangle_{c1}$ from Eq. (13) and $\{E\}_c = \langle E \rangle_{c1}$ from Eq. (14).

We thank É. M. Baskin for a discussion of the questions touched upon here.

[1] R. Blumenfeld and D. J. Bergman, Phys. Rev. B **43**, 13 682 (1991).

[2] P. M. Hui, Phys. Rev. B **49**, 15 344 (1994).

[3] K. W. Yu and P. M. Hui, Phys. Rev. B **50**, 13 327 (1994).

[4] P. M. Hui, Phys. Rev. B **49**, 15 344 (1994).

[5] K. H. Chung and P. M. Hui, J. Phys. Soc. Jpn. **63**, 2002 (1994).

[6] J. J. Lin, J. Phys. Soc. Jpn. **61**, 4125 (1994).

[7] G. M. Zhang, J. Phys.: Condens. Matter **8**, 6933 (1996).

[8] D. Stroud and P. M. Hui, Phys. Rev. B **37**, 8719 (1988).

[9] A. Aharony, Phys. Rev. Lett. **58**, 2726 (1987).

[10] A. A. Snarskii, A. E. Morozovsky, A. Kolek, and A. Kusy, Phys. Rev. E **53**, 5596 (1996).

[11] A. E. Morozovskiĭ and A. A. Snarskiĭ, Zh. Éksp. Teor. Fiz. **102**, 683 (1992) [Sov. Phys. JETP **75**, 366 (1992)].

[12] A. E. Morozovskiĭ and A. A. Snarskiĭ, Zh. Éksp. Teor. Fiz. **95**, 1844 (1992) [Sov. Phys. JETP **68**, 1066 (1989)].

[13] A.-M. S. Tremblay, B. Fourcade, and P. Brenton, Physica A **157**, 89 (1989).

[14] A. Kolek, Int. J. Elec. **73**, 1095 (1992).

[15] A. Kolek, Phys. Rev. B **45**, 205 (1992).

Translated by M. E. Alferieff

Multipole solutions of the wave equation

V. V. Zashkvara[†] and N. N. Tyndyk

Physicotechnical Institute, Kazakhstan Academy of Sciences, 480082 Alma-Ata, Kazakhstan

The wave equation is solved by the operator separation method proposed in V. V. Zashkvara and N. N. Tyndyk, Zh. Tekh. Fiz. **61**(4), 148 (1991) [Sov. Phys. Tech. Phys. **36**, 456 (1991)]. Solutions describing the evolution of circular-multipole fields are obtained in a cylindrical coordinate system. © *1998 American Institute of Physics.* [S1063-7842(98)00306-7]

A new approach to solving Laplace's equation by separation of variables has been proposed and substantiated in Refs. 1–3. In this approach a solution is constructed in the form of a sum of paired products of functions of one variable that satisfy chains of second-order ordinary differential equations and zero boundary conditions on a circle. The differential operators of these equations are coordinate-separated parts of the Laplacian. A class of circular multipoles in cylindrical and spherical coordinate systems has been obtained by the operator separation method.[4]

In Ref. 5 a new approach is used to solve Poisson's equation. It is shown that if the right-hand side of Poisson's equation in a cylindrical system with dimensionless coordinates R/r_0, $\xi = z/r_0$ (r_0 is the radius of an axial circle) is represented by the function

$$\mathcal{F}(R,\xi) = \xi^N \omega(R), \tag{1}$$

where N is an integer and $\omega(R)$ is a smooth function of R, then a solution of Poisson's equation is given by the sum

$$V_n(R,\xi) = \sum_{m=0}^{n} \varphi_m(\xi) \cdot F_{n-m}(R). \tag{2}$$

In this sum $n = N/2$ if N is even and $n = (N-1)/2$ if N is odd. The set of radial functions $F_i(R)$ satisfies a chain of n differential equations

$$TF_0 = N!\,\omega(R),$$

$$TF_1 = -F_0,$$

$$\ldots\ldots,$$

$$TF_n = -F_{n-1}, \tag{3}$$

where

$$T = \frac{1}{R}\frac{d}{dR}\left[R\frac{d}{dR}\right]$$

is a second-order differential operator—the radial part of the Laplacian.

The set of axial functions $\varphi_i(\xi)$ is

$$\varphi_n(\xi) = \frac{1}{(2n)!}\,\xi^{2n},$$

$$n = 0, 1, 2, \ldots, N/2, \quad \text{if } N \text{ is even,}$$

$$\varphi_n(\xi) = \frac{1}{(2n+1)!}\,\xi^{2n+1},$$

$$n = 0, 1, 2, \ldots, (N-1)/2, \quad \text{if } N \text{ is odd.} \tag{4}$$

The functions $F_i(R)$ and $\varphi_i(\xi)$ and their derivatives at the point $R = 1$, $\xi = 0$ satisfy the boundary conditions

$$F_i(1) = \left.\frac{dF_i}{dR}\right|_{R=1} = 0, \quad i = 0, 1, 2, \ldots$$

$$\varphi_i(0) = \left.\frac{d\varphi_i}{d\xi}\right|_{\xi=0} = 0, \quad i = 1, 2, \ldots\,. \tag{5}$$

In Ref. 5 it was shown that for different values of n the solutions $V_n(R,\xi)$ of Poisson's equation possess the structure of modified circular multipoles, which were termed non-Laplacian. In what follows we shall call $f_i(R)$ and $F_i(R)$, respectively, the radial functions of Laplacian and non-Laplacian circular multipoles.

Our objective in the present paper is to use the operator separation method to solve the wave equation in a cylindrical coordinate system, to obtain particular solutions on the basis of the multipole approach, and to determine the character of the temporal processes described by these solutions.

The wave equation in dimensionless cylindrical coordinates R, ξ and dimensionless time $t = ct_1/r_0$ (t_1 is the dimensional time and c is the propagation velocity of a disturbance in a uniform medium) is

$$(\Delta - \chi)U(R,\xi,t) = 0, \tag{6}$$

where Δ is the Laplacian and $\chi = \partial^2/\partial t^2$ is the temporal operator.

According to the operator separation method,[1,2] the solution of Eq. (6) will be a sum of paired products of the coordinates and temporal functions $\nu_i(R,\xi)$ and $\Phi_i(t)$

$$U_n(R,\xi,t) = \sum_{s=0}^{n} \nu_s(R,\xi) \cdot \Phi_{n-s}(t). \tag{7}$$

If these functions satisfy the chains of equations

$$\Delta\nu_0 = 0, \quad \Delta\nu_1 = \nu_0,$$

$$\Delta\nu_2 = \nu_1, \quad \ldots\ldots,$$

$$\Delta\nu_n = \nu_{n-1}, \tag{8}$$

$$\chi\Phi_0 = 0,$$

$$\chi\Phi_1 = \Phi_0,$$

$$\chi\Phi_2 = \Phi_1,$$

$$\ldots\ldots\ldots,$$

$$\chi\Phi_n = \Phi_{n-1}. \tag{9}$$

Indeed, operating on the sum (7) with the operator $[\Delta - \chi]$, we obtain

$$(\Delta - \chi)\sum_{s=0}^{n} \nu_s(R, \xi)\Phi_{n-s}(t)$$

$$= \sum_{s=0}^{n} (\Delta\nu_s)\Phi_{n-s} - \sum_{s=0}^{n} \nu_s(\chi\Phi_{n-s})$$

$$= \sum_{s=1}^{n} \nu_{s-1}\Phi_{n-s} - \sum_{s=0}^{n-1} \nu_s\Phi_{n-s-1}. \tag{10}$$

Let us now shift by 1 the summation index in the first sum $s-1 = k$, $\nu_{s-1} = \nu_k$, $\Phi_{n-s} = \Phi_{n-k-1}$. Then

$$\sum_{s=1}^{n} \nu_{s-1}\Phi_{n-s} = \sum_{k=0}^{n-1} \nu_k\Phi_{n-k-1}.$$

Thus, the sums in Eq. (10) are equal and Eq. (6) is satisfied. The solution of the system of equations (9) is given by the functions

$$\Phi_n(t) = \frac{1}{(2n)!} t^{2n}, \quad n = 0, 1, 2, \ldots, \tag{11}$$

if $\Phi_0(0) = 1$, $\Phi_n(0) = 0$ ($n \neq 0$), and $\partial\Phi_n/\partial t|_{t=0} = 0$, or the functions

$$\Phi_n = \frac{1}{(2n+1)!} t^{2n+1}, \quad n = 0, 1, 2, \ldots, \tag{12}$$

if $\Phi_n(0) = 0$, $\partial\Phi_0/\partial t|_{t=0} = 1$, and $\partial\Phi_n/\partial t|_{t=0} = 0$ ($n \neq 0$).

According to Eq. (8), to find the coordinate functions $\nu_i(R, \xi)$ it is necessary to solve a chain of partial differential equations, consisting of Laplace's equation for the functions ν_0 and Poisson's equations for the remaining functions ν_i. Before attempting to solve this problem we shall supplement the results of Ref. 5 and find the solution of Poisson's equation in the case when the right-hand side of this equation is a polynomial in ξ with arbitrary R-dependent coefficients. Let this be an even polynomial of degree $2N$

$$\mathcal{F}(R, \xi) = \sum_{n=0}^{N} \omega^{(n)}(R)\xi^{2n}. \tag{13}$$

Then the solution of Poisson's equation will be given by the sum

$$V(R, \xi) = \sum_{n=0}^{N} V^{(n)}(R, \xi). \tag{14}$$

In accordance with Eq. (2)

$$V^{(n)}(R, \xi) = \sum_{m=0}^{n} \varphi_m F_{n-m}^{(n)}. \tag{15}$$

Let us substitute expression (15) into Eq. (14). Performing some manipulations, we arrive at a solution in a form that facilitates further analysis of our problem:

$$V(R,\xi) = \sum_{n=0}^{N} \sum_{m=0}^{n} \varphi_m F_{n-m}^{(n)} = \sum_{m=0}^{N} \varphi_m \sum_{n=m}^{N} F_{n-m}^{(n)}$$

$$= \sum_{m=0}^{N} \varphi_m \sum_{s=0}^{N-m} F_s^{(m+s)}. \tag{16}$$

Two successive operations lead to Eq. (16): switching the order of m and n summations and then replacing the summation over n by summation over $s = n - m$. The radial functions $F_s^{(m)}$ form a triangular matrix

$s \rightarrow$	0	1	2	3	4	5	.	N
$m \downarrow$								
0	$F_0^{(0)},$							
1	$F_0^{(1)}$	$F_1^{(1)},$						
2	$F_0^{(2)}$	$F_1^{(2)}$	$F_2^{(2)},$					
3	$F_0^{(3)}$	$F_1^{(3)}$	$F_2^{(3)}$	$F_3^{(3)},$				
4	$F_0^{(4)}$	$F_1^{(4)}$	$F_2^{(4)}$	$F_3^{(4)}$	$F_4^{(4)},$			
5	$F_0^{(5)}$	$F_1^{(5)}$	$F_2^{(5)}$	$F_3^{(5)}$	$F_4^{(5)}$	$F_5^{(5)},$		
..	..	..	..	..	..	..		
N	$F_0^{(N)}$	$F_1^{(N)}$	$F_2^{(N)}$	$F_3^{(N)}$	..	..		$F_N^{(N)}.$

The functions $F_g^{(m)}$, which together form the solution (15) and correspond to individual terms of the polynomial (13), stand in the rows of the matrix ($m = $ const). The solution of Poisson's equation with the right-hand side in the form of the polynomial (13), according to Eq. (16), consists of paired products of φ_m times the sum of the functions F_s^{m+s}. These functions lie along the diagonals of the matrix. The end point of the diagonal along which the index $s = 0, 1, 2, \ldots, N - m$ varies is determined by specifying the value of m. The set of functions forming a diagonal is

$$F_0^{(m)}, F_1^{(m+1)}, F_2^{(m+2)}, \ldots F_{N-m}^{(N)}. \tag{17}$$

Remaining within the multipole approach, we shall solve the system of equations (8), choosing as $\nu_0(R, \xi)$ a circular multipole of order $2N$, which is a harmonic polynomial in ξ (Refs. 1 and 2). Then

$$\omega^{(n)}(R) = \frac{1}{(2n)!} f_{N-n}(R), \quad n = 0, 1, \ldots, N, \tag{18}$$

where f_{N-n} are radial functions of the first or second kind, studied in Ref. 4.

We shall show that in this case the functions $F_s^{(m+s)}$ standing on the same diagonal of the matrix are equal, and the solution of the chain of equations (8) can be easily found. Let us operate with the operator T on the function (17), taking into account the differential equation (3) for F_i and the system of differential equations satisfied by the radial functions f_i (Refs. 1 and 2):

$$Tf_{i+1} = -f_i, \quad i = 0, 1, 2, \ldots, n. \tag{19}$$

As an example, let us find $\nu_1(R, \xi)$. According to Eqs. (3) and (18), $TF_0^{(m)} = \omega^{(m)}(R) = f_{N-m}$. Using Eq. (19), we obtain from this relation

$$F_0^{(m)} = -f_{N-m+1}. \tag{20}$$

According to Eqs. (3) and (20), $TF_1^{(m+1)} = -F_0^{(m+1)} = f_{N-m}$, and we have on the basis of (19)

$$F_1^{(m+1)} = -f_{N-m+1}. \tag{21}$$

According to Eqs. (3) and (21), $TF_2^{(m+2)} = -F_0^{(m+2)} = f_{N-m}$. On the basis of Eq. (19) we have

$$F_2^{(m+2)} = -f_{N-m+1} \tag{22}$$

and so on. Indeed, all $N-m+1$ diagonal elements $F_s^{(m+s)}$ are equal and have the value $-f_{N-m+1}$; there are $N-m+1$ of them. Returning to formula (16), we conclude that if the right-hand side of Poisson's equation is represented by a circular multipole $\nu_0(R, \xi)$ of order $2N$, then the solution of this equation $\nu_1(R, \xi)$ is a polynomial of the same order in ξ and has the form

$$\nu_1(R, \xi) = -\sum_{m=0}^{N} (N-m+1)\varphi_m(\xi)f_{N-m+1}(R). \tag{23}$$

Generalizing Eq. (23), one can show that the solution of an arbitrary link $\Delta\nu_s = \nu_{s-1}$ in the chain of differential equations (8) is

$$\nu_s(R,\xi) = \frac{(-1)^s}{s!} \sum_{m=0}^{N} \frac{(N-m+s)!}{(N-m)!} \varphi_m(\xi)f_{N-m+s}(R). \tag{24}$$

Substituting expression (24) into Eq. (7), we write the solution of the wave equation (6) in the form

$$U_n(R, \xi, t) = \sum_{s=0}^{n} \frac{(-1)^s}{s!} \Phi_{n-s}(t)$$
$$\times \sum_{m=0}^{N} \frac{(N-m+s)!}{(N-m)!} \varphi_m(\xi)f_{N-m+s}(R). \tag{25}$$

The spatiotemporal structure of the solution (7) or (25) near the axial circle $R=1$, $\xi=0$ is determined by the leading terms of the functions (24). To find them it is necessary to separate out the leading terms of the radial functions f_i appearing in Eq. (24). It turns out that in consequence of the chain structure of the system of differential equations,[1,2] which f_i satisfy with zero boundary conditions, we have

$$f_{i+1}/f_i \sim \rho^2 \quad \text{as} \quad \rho \to 0 \quad (\rho = R-1). \tag{26}$$

It follows from this relation that the leading terms f_i equal

$$f_i \sim (-1)^i \rho^{2i+1}$$

for the radial functions of the first kind and

$$f_i \sim (-1)^i \rho^{2i}$$

for the radial functions for the second kind,

$$i = 0, 1, \dots . \tag{27}$$

We shall confine our attention to the radial functions of the second kind. Then the leading terms in $\nu_s(R, \xi)$ (24) can be represented by the formula

$$\nu_s^*(R,\xi) \sim \frac{1}{s!}\rho^{2s} \sum_{m=0}^{N} (-1)^{N-m}$$
$$\times \frac{(N-m+s)!}{(N-m)!(2m)!}[\xi^m\rho^{N-m}]^2. \tag{28}$$

A characteristic feature here is that the sum is a homogeneous polynomial of degree $2N$ in ρ and ξ containing sign-alternating coefficients. This polynomial by itself describes a multipole structure with nodal point $R=1$, $\xi=0$, where the $2N$ zero equipotential lines converge. The factor ρ^{2s} in front of the sum contributes to the structure of the field an additional zero equipotential lying on the ξ axis, which separates the single-potential region into two symmetric parts, thereby increasing the multipole order by 2. In Ref. 5 such structures were called incomplete non-Laplacian circular multipoles. We shall also determine the spatiotemporal structure of the field (25) on a surface of section by the plane $\xi=0$ near $R=1$. In accordance with Eq. (4), we have for even functions $\varphi_n(\xi)$

$$\varphi_m(0) = \begin{cases} 1, & \text{if } m=0, \\ 0, & \text{if } m=1, 2, \dots . \end{cases} \tag{29}$$

According to Eq. (25) we have

$$U_n(R, 0, t) = \sum_{s=0}^{n} (-1)^s f_s(R)\Phi_{n-s}(t). \tag{30}$$

As an example, let us consider even functions for the choice of $\Phi_n(t)$. Then

$$\Phi_{n-s}(t) = \frac{1}{[2(n-s)]!}t^{2(n-s)}. \tag{31}$$

Substituting expressions (27) and (31) into Eq. (30), we arrive at the following expression for estimating the spatiotemporal structure of the field in the section $\xi=0$

$$U_n(R, 0, t) = \sum_{s}^{n} \frac{1}{[2(n-s)]!}\rho^{2s}t^{2(n-s)}. \tag{32}$$

It follows from Eq. (32) that $U_n(R, 0, t)$ is a homogeneous polynomial in R and t of degree $2n$ whose coefficients all have the same sign. Near the point $R=1$, $t=0$ the equipotentials $U_n(R, 0, t)$ of the wave field form a system of closed curves encompassing this point. The field $U_n(R, 0, t)$ is not a multipole field. But after one makes the substitution $t=i\tau$ and transforms to imaginary τ in Eq. (32), the field in the coordinates R,τ becomes a circular multipole with a node at $R=1$, $\xi=0$, $\tau=0$.

Thus the spatial–wave structure of the wave field (7) in all space is determined by a time-dependent polynomial whose coefficients $\nu_s(R, \xi)$ are circular multipoles. Initially $(t=0)$ the structure of the wave field $U_n(R, \xi, t)$ will be represented by a non-Laplacian circular multipole $\nu_n(R, \xi)$ $(n \neq 0)$, and in the limit $t \to \infty$ it will be represented by a Laplacian circular multipole $\nu_0(R, \xi)$.

As an example, let us examine the structure of the multipole solution of the wave equation (7) for the case $N=1$. The Laplacian circular multipole is
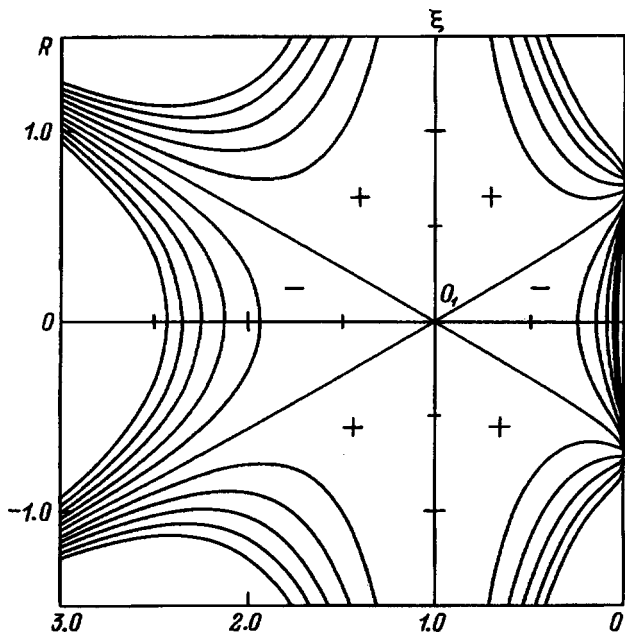
FIG. 1.

FIG. 3.

$$\nu_0(R,\ \xi)=f_1\varphi_0+f_0\varphi_1. \tag{33}$$

According to Eq. (24), the non-Laplacian circular multipoles can be calculated from the formula

$$\nu_s(R,\ \xi)=\frac{(-1)^s}{s!}\sum_{m=0}^{1}\frac{(s+1-m)!}{(1-m)!}\varphi_m(\xi)f_{s+1-m}(R); \tag{34}$$

$$s=1,\quad \nu_1=-[2\varphi_0f_2+\varphi_1f_1],$$

$$s=2,\quad \nu_2=3\varphi_0f_3+\varphi_1f_2,$$
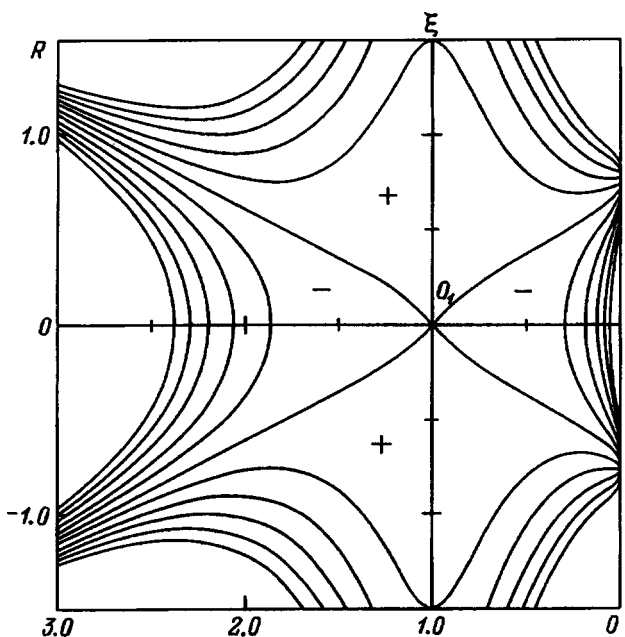
$$s=3,\quad \nu_3=-[4\varphi_0f_4+\varphi_1f_3] \tag{35}$$
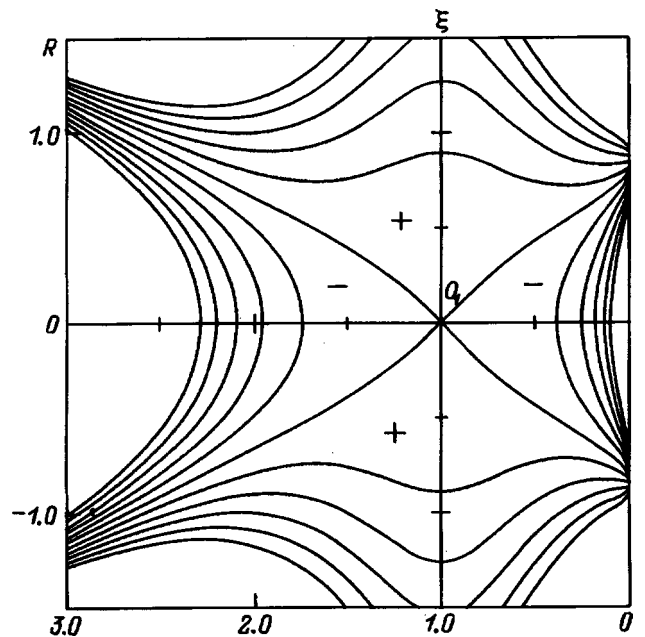
and so on. For even functions $\Phi_n(t)$ we have, according to Eq. (7),

$$n=0,\quad U_0(r,\ \xi,\ t)=\nu_0\Phi_0=f_1\varphi_0+f_0\varphi_1, \tag{36}$$

$$n=1,\quad U_1(R,\ \xi,\ t)=\sum_{s=0}^{1}\nu_s\Phi_{1-s}$$

$$=\frac{1}{2!}(f_1\varphi_0+f_0\varphi_1)t^2-[2\varphi_0f_2+\varphi_1f_1], \tag{37}$$

$$n=2,\quad U_2(R,\ \xi,\ t)=\sum_{s=0}^{2}\nu_s\Phi_{2-s}=\frac{1}{4!}(f_1\varphi_0+f_0\varphi_1)t^4$$

FIG. 2.
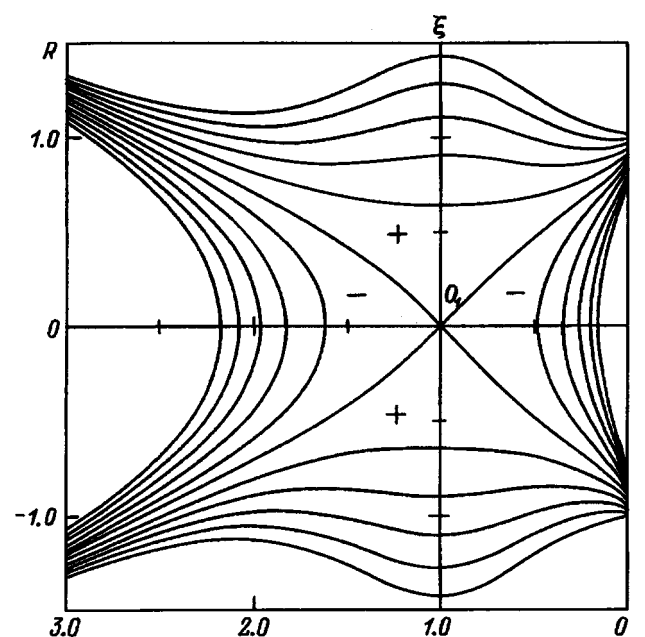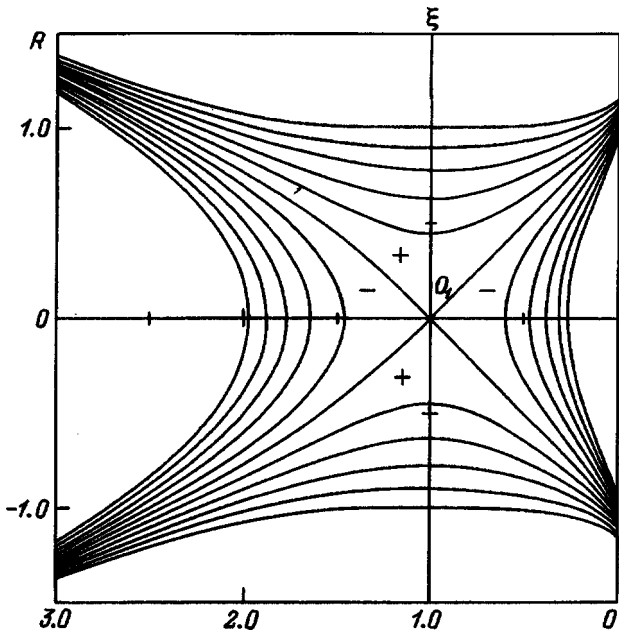

FIG. 4.

FIG. 5.
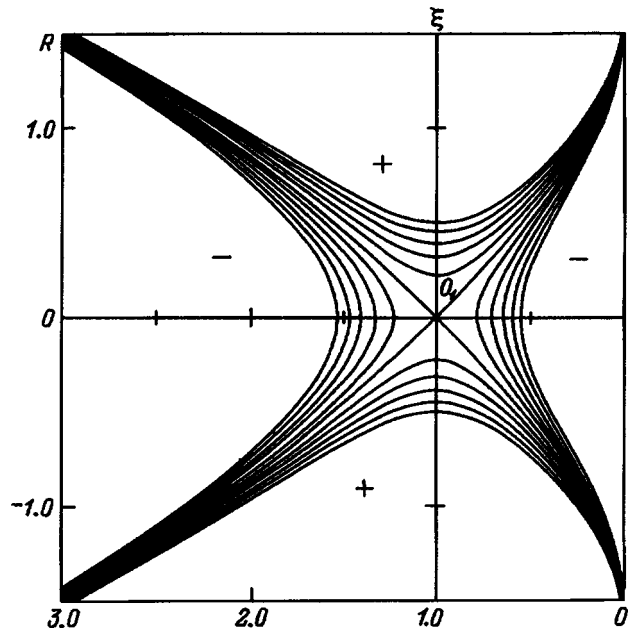

FIG. 6.

$$-\frac{1}{2!}[2\varphi_0 f_2+\varphi_1 f_1]t^2+3\varphi_0 f_3+\varphi_1 f_2 \tag{38}$$

and so on.

To calculate U_1 the radial functions of the second kind[4] were chosen for $f_i(R)$ and the even functions[4] of ξ were chosen for $\varphi_i(\xi)$:

$$f_0(R)=1, \quad f_1(R)=\frac{1}{4}[2\ln R+1-R^2],$$

$$f_2(R)=\frac{1}{64}[-(4+8R^2)\ln R-5+4R^2+R^4]. \tag{39}$$

$$\varphi_0(\xi)=1, \quad \varphi_1(\xi)=\frac{1}{2}\xi^2. \tag{40}$$

Then, according to Eq. (37),

$$U_1(R,\xi,t)=\frac{1}{4}\left[\frac{1}{2}(2\ln R+1-R^2)+\xi^2\right]t^2$$

$$-\frac{1}{32}[-(4+8R^2)\ln 5+4R^2+R^4]$$

$$-\frac{1}{8}\xi^2[2\ln R+1-R^2]. \tag{41}$$

The computational results obtained for $U_1(R,\xi,t)$ using Eq. (41) are presented in Figs. 1–6, which show "snap-shots" of different times in the development of the spatiotemporal process described by the wave multipole $U_1(R,\xi,t)$: $t=0$, 0.3, 0.5, 0.7, and 1.2, respectively. For $t=0$ the wave multipole is an incomplete non-Laplacian circular sextupole (Fig. 1). Even for small t the contribution of the quadrupole $\frac{1}{2}(2\ln R+1-R^2)+\xi^2$ eliminates the zero equipotential $R=1$ and destroys the non-Laplacian sextupole. It is evident from these figures that in time the development of this process leads to the formation of a structure close to a circular Laplacian quadrupole (Figs. 5, 6).

The process unfolds similarly in the other cases which we examined. It can be concluded on this basis that the operator separation method makes it possible to obtain a multipole solution of the wave equation that describes a field evolving from a more complicated to a simpler structure.

[†]Deceased.

[1] V. V. Zashkvara and N. N. Tyndyk, Zh. Tekh. Fiz. **61**(4), 148 (1991) [Sov. Phys. Tech. Phys. **36**, 456 (1991)].

[2] V. V. Zashkvara and N. N. Tyndyk, Nucl. Instrum. Methods A **313**, 315 (1992).

[3] V. V. Zashkvara and N. N. Tyndyk, Nucl. Instrum Methods A **321**, 339 (1992).

[4] V. V. Zashkvara, Nucl. Instrum. Methods **354**, 171 (1995).

[5] V. V. Zashkvara and N. N. Tyndyk, Zh. Tekh. Fiz. **65**(7), 154 (1995) [Tech. Phys. **40**, 717 (1995)].

Translated by M. E. Alferieff

Doppler spectroscopy of a beam in the channel of a H⁻ ion source

V. V. Antsiferov and G. I. Smirnov

Institute of Nuclear Physics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
(Submitted May 21, 1997)
Zh. Tekh. Fiz. **68**, 15–18 (June 1998)

It is shown that high-accuracy contact-free measurements of the divergence and emittance of an accelerated H⁻ ion beam at the exit from the source can in principle be performed by passive Doppler spectroscopy of a beam of excited hydrogen atoms produced by neutralization of the ions with excitation on the residual gas in the source channel. The intensity of the H_α-line radiation detected by the Doppler system is calculated, taking into account the principal processes leading to the excitation and deexcitation of the $3s$, $3p$, and $3d$ levels of the hydrogen atoms in the beam, for residual gas densities of the order of 10^{-4}–10^{-5} Torr in the source channel. The computed H_α-line intensity was confirmed experimentally, making it possible to perform photoelectronic detection of the spectral contour of the line in the current mode rather than the photon-counting mode. © *1998 American Institute of Physics.*
[S1063-7842(98)00406-1]

INTRODUCTION

Monitoring of the parameters of a H⁻ ion beam during continuous operation of the source is an extremely pressing problem. Contact sensors[1-3] do not solve this problem because they introduce perturbations into the beam and increase the beam divergence, and it is virtually impossible to use contact sensors for continuous monitoring of the divergence of an ion beam during the operation of the source. The only method permitting such nonperturbative monitoring is Doppler spectroscopy, which has been used in the past for monitoring the divergence of high-intensity beams of atoms and negative ions of hydrogen in the channel of an accelerator[4,5] and for measuring the temperature of hydrogen atoms in the plasma discharge of surface-plasma sources of H⁻ ions.[6,7]

In the present work we determined the possibilities of performing high-accuracy contact-free monitoring of the divergence and emittance of an accelerated beam of negative hydrogen ions in the source channel by passive Doppler spectroscopy without using a charge-transfer target.

H_α-LINE INTENSITY OF AN ACCELERATED PARTICLE BEAM AT THE SOURCE EXIT

An organically integral component of the ion source of a modern accelerator is the turning magnet (see Fig. 1), which separates particles according to the transverse velocities of the beam extracted from the source and forms an ion beam with a fixed emittance. Transport of the H⁻ ion beam along the channel of the source magnet is accompanied by neutralization of the ions and excitation on the residual gas particles. The number $N(H_\alpha)$ of photons spontaneously emitted in the H_α line by excited atoms in the hydrogen beam during the time t_r that the beam atoms travel over the rectilinear detection section at the exit from the magnet is

$$N(H_\alpha) = N_{3s} + N_{3p} + N_{3d}, \tag{1}$$

where N_m is the number of photons emitted in the transitions $3s \to 2p$, $3p \to 2s$, and $3d \to 2p$, respectively, by hydrogen atoms in the beam and is determined by the number of spontaneous decays occurring over the travel time t_r of excited hydrogen atoms in the detection section which are present in the entire volume of the beam pulse and enter the detection system of the Doppler recording system with angular aperture $\Delta\varphi_A$

$$N_m = \frac{D\Delta\varphi_A V A_m}{4\pi f} \int_0^{t_r} n_m(x)dt. \tag{2}$$

Here D and f are, respectively, the diameter and focal length of the converging lens L_1; A_m are the probabilities of the corresponding spontaneous transitions—$A_{3s} = 0.63 \times 10^7$ s⁻¹, $A_{3p} = 2.2 \times 10^7$ s⁻¹, and $A_{3d} = 6.4 \times 10^7$ s⁻¹; $V = \pi d^2 c\tau\beta/4$ is the volume of a beam pulse of duration τ with d being the beam diameter and $\beta = v/c$, where v is the velocity of the ions in the beam; $n_m(x)$ is the density of excited hydrogen atoms at the detection point in the states $3s$, $3p$, and $3d$

$$n_m(x) = n_m^0(x)\exp(A_m t). \tag{3}$$

The densities $n_m^0(x)$ of the excited hydrogen atoms after the particles have passed through a section of length x in the source can be found from an equation that takes into account the excitation of H⁻ ions and H⁰ atoms in collisions with the residual gas as well as radiative decay of the excited levels and collisional relaxation on the residual gas

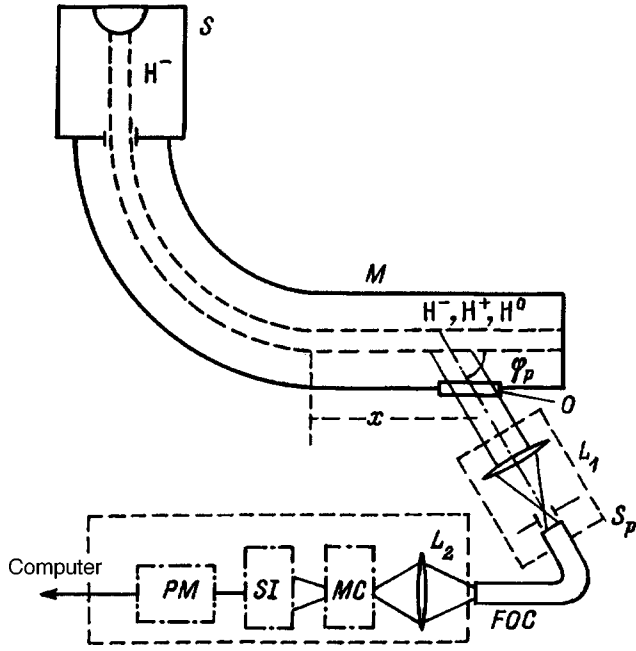$$\frac{dn_m^0(x)}{dx} = n_g[n^0(x)\sigma_{0,m} + n^-\sigma_{-1,m}]$$

FIG. 1. Doppler system for detecting a H⁻ ion beam in the source channel. S—H⁻ ion source, M—turning magnet, O—optical window, L_1—converging lens, S_p—detector slit, FOC—fiber-optic cable, L_2—focusing lens of the spectral unit, MC—monochromator, SI—scanning interferometer, PM—photomultiplier.

$$-\frac{n_m^0(x)A_m}{v} - n_g n_m^0(x)\sigma_{DC}, \qquad (4)$$

where n_g is the density of residual gas particles in the source channel, $n^0(x)$ is the density of hydrogen atoms in the $1s$ ground state in the beam, $n^-(x)$ is the density of H⁻ ions in the beam, $\sigma_{0,m}$ is the cross section for the excitation of hydrogen atoms into the state m in collisions with the residual gas, $\sigma_{-1,m}$ is the cross section for charge transfer from H⁻ ions accompanied by the excitation of hydrogen atoms into the state m in collisions with the residual gas, v is the longitudinal velocity of the accelerated beam of H⁻ ions, $\sigma_{DC} = \sigma_{m,0} + \sigma_{m,+1} + \sigma_{m,-1}$ is the total cross section for the collisional decay of the excited levels m.

Since $\sigma_{-1,1} \ll \sigma_{-1,0}$, the H⁰ atom and H⁻ ion densities in the beam are, from Eqs. (4),

$$n^-(x) = n^-(0)\exp(-x n_g \sigma_{0,1}), \qquad (5)$$

$$n^0(x) = \frac{\sigma_{-1,0} n^-(0)}{\sigma_{-1,0} - \sigma_{0,1}}[\exp(-x n_g \sigma_{0,1}) - \exp(-x n_g \sigma_{-1,0})]. \qquad (6)$$

We introduce the following notation: $\lambda_{DC} = 1/n_g \sigma_{DC}$ is the mean free path of excited hydrogen atoms in the state $m = 3$ due to collisional decay of the excited levels; $\lambda_{DR} = v/A_m$ is the mean free path of excited hydrogen atoms due to the decay of the excited levels as a result of spontaneous emission; $1/\lambda_D = 1/\lambda_{DC} + 1/\lambda_{DR}$; and, $\lambda_1 = 1/n_g \sigma_{-1,0}$, $\lambda_2 = 1/n_g \sigma_{0,1}$, and $B = \sigma_{-1,0}/(\sigma_{-1,0} - \sigma_{0,1})$. Substituting the parameters into Eq. (2), we find

$$\frac{dn_m^0(x)}{dx} + \frac{n^0(x)}{\lambda_D} = n_g[B\sigma_{0,m}n^-(0)[\exp(-x/\lambda_2)$$
$$- \exp(-x/\lambda_1)] + n^-(0)\sigma_{-1,m}$$
$$\times \exp(-x/\lambda_1)]. \qquad (7)$$

Solving this equation with the initial conditions $n_m^0(0) = 0$, we obtain

$$n_m^0(x) = n_g n^-(0)[B\sigma_{0,m}(\lambda_D^{-1} - \lambda_2^{-1})^{-1}[\exp(-x/\lambda_2)$$
$$- \exp(-x/\lambda_D)] + (\sigma_{-1,m} - B\sigma_{0,m})(\lambda_D^{-1} - \lambda_1^{-1})^{-1}$$
$$\times [\exp(-x/\lambda_1) - \exp(-x/\lambda_D)]. \qquad (8)$$

We shall calculate the density $n_m^0(x)$ of excited hydrogen atoms for two values of the residual gas density n_g in the source for a H⁻ ion source with the optimal parameters: ion-beam energy 40 keV, pulse current 0.2 A, pulse duration 0.1 ms, beam diameter 2 cm, beam emittance 10^{-4} rad·cm, and pulse repetition frequency 100 Hz. Then $v = 2.8 \times 10^8$ cm/s, $n^-(0) = 2 \times 10^9$ cm⁻³, and the parameters β are $\beta_\parallel = v/c = 7 \times 10^{-3}$ for the longitudinal component of the beam velocity and $\beta_\perp = v_\perp/c = 1 \times 10^{-4}$ for the transverse component.

1. Let the residual gas pressure in the source be $P = 5 \times 10^{-5}$ Torr. Then $n_g = 1.6 \times 10^{12}$ molecules/cm³. For 40 keV H⁻ ions the charge-transfer cross sections are[8] $\sigma_{-1,0} = 8.7 \times 10^{-16}$ cm², $\sigma_{-1,1} = 4 \times 10^{-17}$ cm², and $\sigma_{0,1} = 1.43 \times 10^{-16}$ cm². Then $\sigma_{DC}^{3s} = \sigma_{DC}^{3p} = \sigma_{DC}^{3d} = 1.6 \times 10^{-15}$ cm². The cross sections for neutralization with excitation are[9] $\sigma_{-1,3s} = 2.1 \times 10^{-18}$ cm² and $\sigma_{-1,3p} + \sigma_{-1,3d} = 1.5 \times 10^{-18}$ cm². For a section of length $x = 20$ cm in the source channel where charge transfer from H⁻ ions occurs the other computational parameters are $\lambda_1 = 0.7 \times 10^3$ cm, $\lambda_2 = 4.4 \times 10^3$ cm, $\lambda_{DC} = 0.4 \times 10^3$ cm, $\lambda_{DR}^{3s} = 44$ cm, $\lambda_{DR}^{3p} = 12$ cm, and $\lambda_{DR}^{3d} = 4.3$ cm. Then Eq. (8) simplifies substantially:

$$n_m^0(x) = n_g n^-(0)[\sigma_{-1,m}\lambda_{DR}^m[1 - \exp(-x/\lambda_{DR}^m)]]. \qquad (9)$$

Substituting expression (9) into Eq. (4), integrating and substituting into Eqs. (2) and (1), we find for the number $N(H_\alpha)$ of photons emitted in the H_α line

$$N(H_\alpha) = \frac{DV\Delta\varphi_A}{4\pi f} \sum_m n_m^0(x)[1 - \exp(-A_m t_r)]. \qquad (10)$$

For a diameter $D = 6$ cm of the lens L_1 in the optical detection system the detection time for the H_α-line radiation is $t_r = D/v = 0.2 \times 10^{-7}$ s. Substituting into Eq. (10) the values of all parameters gives

$$N(H_\alpha) = 3 \times 10^5 \text{ photons/pulse}. \qquad (11)$$

2. For a residual gas pressure $P = 5 \times 10^{-4}$ Torr and a source-channel density $n_g = 1.6 \times 10^{13}$ molecules/cm³, we have $\lambda_1 = 70$ cm, $\lambda_2 = 440$ cm, and $\lambda_{DC} = 40$ cm, the other parameters having similar values. In this case Eq. (8) becomes

$$n_m^0(x) = n_g n^-(0)[\sigma_{0,m} + B\lambda_D^m[1 - \exp(-x/\lambda_D^m)]$$
$$+ (\sigma_{-1,m} - B\sigma_{0,m})(\lambda_D^{-m} + \lambda_1^{-1})^{-1}$$

$$\times[\exp(-x/\lambda_1)-\exp(-x/\lambda_D^m)]]. \tag{12}$$

Here $\sigma_{0,m}$ is the cross section for the excitation of hydrogen atoms into the $3s$, $3p$, and $3d$ states in collisions with residual gas particles in the source: $\sigma_{0,3s}=1\times10^{-17}$ cm^2, $\sigma_{0,3p}=1.5\times10^{-17}$ cm^2, and $\sigma_{0,3d}=7\times10^{-18}$ cm^2 (Ref. 10). The working values of the parameters $n_m^0(x)$ are: $n_{3s}^0(x)=2.8\times10^5$ cm^{-3}, $n_{3p}^0(x)=3.6\times10^5$ cm^{-3}, and $n_{3d}^0(x)=3.4\times10^5$ cm^{-3}. Substituting the values of all parameters into Eq. (10), we have in this case for the number $N(\mathrm{H}_\alpha)$ of photons

$$N(\mathrm{H}_\alpha)=4.2\times10^6 \text{ photons/pulse.} \tag{13}$$

PARAMETERS OF THE DOPPLER DETECTION SYSTEM

The contribution from the spread $\Delta\beta_\parallel$ of the longitudinal particle velocities in the accelerated beam to the Doppler broadening of the H_α line can be neglected for radiation detection angles φ_r close to the ''magic'' angle φ_M, which equals[4]

$$\varphi_M=\arccos\beta_\parallel. \tag{14}$$

Since the average longitudinal velocity of beam particles in the source is $v=2.8\times10^8$ cm/s, the magic angle equals $\varphi_M=89.6°$. When the radiation detection angle equals the magic angle φ_M, two H_α lines will be observed in the Doppler detection system: a line shifted by 0.03 nm, emitted by hydrogen atoms in the beam which are excited in the process of charge transfer from the ions on the residual gas, and an unshifted line, emitted by excited hydrogen atoms formed in the residual gas as a result of the dissociation of molecular hydrogen in collisions with accelerated hydrogen ions and atoms in the beam: $\mathrm{H}^-(\mathrm{H}_0)+\mathrm{H}_2\rightarrow\mathrm{H}^-(\mathrm{H}^0)+\mathrm{H}_0^*(n=3)+\mathrm{H}_0$. The density of the molecular hydrogen in the residual gas in the source channel is quite high ($\sim10^4$ Torr) and the intensity of the unshifted H_α line will be comparable to that of the shifted H_α line, while its spectral width will be

$$\Delta\lambda_{v_T}=2\lambda_0 v_T/c. \tag{15}$$

For the average thermal velocity of the residual gas $v_T=1.38\times10^6[3E(\mathrm{eV})]^{1/2}$, the width of the unshifted H_α line equals 0.1 nm.

The width of the Doppler broadened H_α line emitted by excited hydrogen atoms in the accelerated beam[4]

$$\Delta\lambda_D=2\lambda_0\beta_\perp\cos\theta \tag{16}$$

will equal 0.13 nm in our case. To separate these two H_α lines in the spectrum the detection angle φ_r must be chosen to be somewhat different from the magic angle φ_M. The deviation $\Delta\varphi_r$ of the detection angle from the magic angle must satisfy the condition[4]

$$\pm\Delta\varphi_r\leqslant10^2\frac{\Delta\lambda_D}{\lambda_0\beta_\parallel}\leqslant30°, \tag{17}$$

for which the contribution from the longitudinal velocity spread $\Delta\beta_\parallel$ of the beam to the width of the Doppler contour of the shifted H_α line can be neglected. Therefore the detection angle of the Doppler system can be chosen as $\varphi_r=60°$. In this case the separation of the two H_α lines in the

spectrum will equal 2.3 nm, which makes it possible to separate quite simply the shifted H_α line emitted by an accelerated beam of hydrogen atoms. The contribution to the width $\Delta\lambda_D$ of the Doppler contour of the H_α line with the detection angle different from the magic angle will be of the order of 10^{-3} nm.

The contribution due to the angular aperture $\Delta\varphi_A$ of the receiving unit of the Doppler system to the width of the Doppler contour is eliminated if[4]

$$\Delta\varphi_A\leqslant\frac{0.1\Delta\lambda_D}{\lambda_0\beta_\parallel}\leqslant0.3°. \tag{18}$$

The intensity of the detected H_α line can be increased by using in the detecting unit of the Doppler system (see Fig. 1) a lens L_1 with diameter D and a slit of width

$$\Delta l=f\cdot\Delta\varphi_A \tag{19}$$

at the focal point f.

For a lens with the focal length $f=20$ cm the slit in the detecting unit should be 1 mm wide. For spectral analysis the H_α-line emission from the detecting unit of the Doppler system is extracted with a fiber-optic cable into a spectral unit consisting of a monochromator crossed with a scanning Fabry–Perot interferometer. The monochromator separates the shifted H_α line emitted by the excited hydrogen atoms in the accelerated beam, while the width of the Doppler contour is detected with a scanning interferometer and photomultiplier, the signal from which is fed into a computer.

CONCLUSIONS

The results of the present work show that passive Doppler spectroscopy of a beam of partially excited hydrogen atoms, obtained by neutralization of the ions in a beam and excitation of the hydrogen atoms in the beam on the residual gas in the source channel, can be used for nonperturbative monitoring of the divergence and emittance of a H^- beam extracted from the source. There is no need to introduce a gas target into the source channel, and the Doppler contour of the H_α line can be recorded using photoelectronic detection of the H_α-line contour rather than the photon-counting mode. The method of passive Doppler spectroscopy developed was used to record the divergence of a beam of H^- ions extracted from sources of negative hydrogen ions with planotronic and Penning electrode geometry. Its accuracy was confirmed by comparing with contact slit measurements of the beam divergence. It was shown that for discharge current densities $\leqslant25$ A/cm^2 in the sources the divergence of the H^- ion beam is smallest in the case of a planotronic source of H^- ions (Special Report No. 7504, Sukhumi Physicotechnical Institute, 1989).

[1] G. E. Derevyankin and V. G. Dudnikov, IYaF SO AN SSSR Preprint No. 79-17 [in Russian], Institute of Nuclear Physics, Siberian Branch of the Academy of Sciences of the USSR, Novosibirsk (1979), 15 pp.

[2] G. E. Derevyankin, V. G. Dudnikov, and M. L. Troshkov, IYaF SO AN SSSR Preprint No. 82-110 [in Russian], Institute of Nuclear Physics, Siberian Branch of the Academy of Sciences of the USSR, Novosibirsk (1982), 18 pp.

[3]G. E. Derevyankin and V. G. Dudnikov, AIP Conf. Proc. No. 11, pp. 376–397 (1984).

[4]V. V. Antsiferov, Zh. Tekh. Fiz. **62**(5), 71 (1992) [Tech. Phys. **37**, 525 (1992)].

[5]V. V. Antsiferov, AIP Conf. Proc. No. 87, pp. 616–617 (1993).

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

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

[8]H. Tawara and A. Russec, Rev. Mod. Phys. **45**, 178 (1973).

[9]J. Geddes, J. Hill, and H. B. Gilbody, J. Phys. B **14**, 4837 (1981).

[10]R. H. Hughes, H. M. Petefisch, and H. Kisner, Phys. Rev. **5**, 2103 (1972).

Translated by M. E. Alferieff

Cooperative population dynamics of an ensemble of Λ atoms in a bichromatic field

B. G. Matisov, I. A. Grigorenko, N. Leinfell'ner, I. E. Mazets, and A. Yu. Snegirev

St. Petersburg State Technical University, 195251 St. Petersburg, Russia

Equations are derived which describe the dynamics of three-level atoms with a Λ level scheme, interacting with two coherent resonance fields under conditions such that cooperative relaxation predominates over incoherent spontaneous emission. A numerical calculation of the temporal dynamics of the values of the atomic populations is performed. It is shown that coherent population trapping in the presence of cooperative decay is possible. The quantities characterizing this phenomenon are calculated—the width of the black line and the transition time to coherent trapping in this scheme. © *1998 American Institute of Physics.* [S1063-7842(98)00506-6]

INTRODUCTION

Research on the interaction of coherent electromagnetic radiation with multilevel quantum systems is today a rapidly developing field of nonlinear laser spectroscopy and quantum optics. In multilevel systems the presence of several channels of excitation and induction of coherences between long-lived quantum states by laser fields plays a fundamental role. This leads to the appearance of different quantum interference effects in the internal dynamics of the atoms. The interference of quantum states that arises as a result of coherent excitation and leads in turn to coherent population trapping (CPT) is the basis of many directions of development in modern physics: ultradeep laser cooling of atoms, production of lasers without an inversion, and others (see Refs. 1 and 2 for a more detailed discussion).

It is known[3] that different relaxational processes, for example collisional, strongly influence the evolution of the populations of atomic systems under CPT conditions. However, in previous works concerning CPT in atomic systems, the case $n \cdot \lambda^3 \gtrsim 1$, where n is the atom density and λ is the wavelength of an optical transition, was not studied. Under these conditions cooperative effects have a determining influence on the evolution of the system. As is well known,[4,5] the main such effect is the mutual coordination of the behavior of the atoms. As a result of this, relaxation in such a system is determined not by ordinary spontaneous emission but rather by a coherent process (superradiance). Effects similar to superradiance in a lumped Dicke model[6] also occur in elongated (needle-shaped) samples, where cooperative emission has a narrow directional pattern.

In the present paper we extend the well-known single-mode Bonifacio model[7] to the case of an ensemble of three-level Λ atoms interacting with a coherent bichromatic electromagnetic field in the presence of cooperative relaxation. It is established on the basis of equations derived below and describing the evolution of the system that coherent population trapping is possible and the main parameters characterizing this phenomenon are calculated.

EQUATIONS FOR THE DENSITY MATRIX

Let us list the main assumptions employed in this work. It is assumed that an ensemble of N identical three-level atoms interacts with two resonance electromagnetic waves. Let L be the linear size of the medium. We denote the atomic states by $|0\rangle_l$, $|1\rangle_l$, and $|2\rangle_l$, where $l = 1, \ldots, N$ enumerates the coherent atoms. The corresponding wave function is

$$|i\rangle_l = |i(INT)\rangle \otimes |i(MOT)\rangle_l, \quad i = 0, 1, 2, \tag{1}$$

which consists of two parts describing the internal (*INT*) and translational (*MOT*) degrees of freedom. Let $|0(INT)\rangle$ be the upper excited Λ state and $|1(INT)\rangle$ and $|2(INT)\rangle$ the low-energy states. Absorption of a photon corresponds to a shift of the translational part of the wave function (1) in momentum space by the amount of the photon momentum. For this reason, for further analysis it is convenient to perform the following phase transformation:[4,5]

$$|j(MOT)\rangle_l = \exp(ik_j z_l)|0(MOT)\rangle_l, \quad j = 1,2, \tag{2}$$

where z_l is the coordinate of the lth atom and k_j is the wave number of the jth mode.

This transformation makes it possible to eliminate in the equations below the explicit dependence on the coordinates. We define the collective atomic operator as

$$\hat{R}_{ij} = \sum_{I=1}^{N} |j\rangle_{II}\langle i|. \tag{3}$$

The commutator between the operators $\hat{R}_{ij}$ is

$$[\hat{R}_{\mu\nu}, \hat{R}_{\nu'\mu'}] = \delta_{\nu\nu'}\hat{R}_{\mu\mu'} - \delta_{\mu\mu'}\hat{R}_{\nu'\nu}. \tag{4}$$

The commutation rules follow directly from the definition (3) and the orthonormality of the wave functions

$$_l\langle i|i'\rangle_{l'} = \delta_{ll'}\delta_{ii'}. \tag{5}$$

The electromagnetic field is described by bosonic creation and annihilation operators $\hat{a}_j^+$ and $\hat{a}_j$. The coupling constant characterizing the interaction of the atomic system with the field is defined as

$$g_j = \sqrt{\frac{2\pi\hbar\omega_{0j}}{V}}\, d_{0j}, \quad j = 1, 2, \tag{6}$$

where ω_{0j} is the frequency, d_{0j} is the $|0\rangle - |j\rangle$ transition dipole moment, and V is the quantization volume.

As a simplification, we shall neglect incoherent decay from the upper excited level, which is a consequence of the interaction of the vacuum (zero) modes with the continuum. The reason for this simplification is that the cooperative process is much faster than the incoherent process. Instead of slow atomic relaxation, fast relaxation of the electromagnetic field into a laser-determined stationary state which produces a pure coherent state $|\alpha_1\alpha_2\rangle_f$, i.e.,

$$\hat{a}_j |\alpha_1\alpha_2\rangle_f = \alpha_j |\alpha_1\alpha_2\rangle_f, \tag{7}$$

where α_j are complex numbers, is introduced.

The most important characteristic of the system is the residence time $\tau = L/c$ of a photon in the medium. At times shorter than τ, the electromagnetic field is still in a state that is coherently coupled with the state of the atomic system which has emitted a photon. At times longer than τ the photons leave the region of interaction and coherence between the matter and field states breaks down. The time τ is of the order of 1 ns for $L \approx 30$ cm and can be chosen as the shortest time interval in the system, since it is much shorter than 10^{-7} s—the characteristic spontaneous relaxation time. Thus the field density matrix arrives in its stationary state $|\alpha_1\alpha_2\rangle_{ff}\langle\alpha_1\alpha_2|$ with relaxation time τ.

We shall describe the interaction of two waves with the entire collection of atoms as

$$\hat{H} = -\hbar\Omega_1 \hat{R}_{11} - \hbar\Omega_2 \hat{R}_{22} + g_1 \hat{a}_1 \hat{R}_{01}$$
$$+ g_2^* \hat{a}_2^+ \hat{R}_{20} + g_1^* \hat{a}_1^+ \hat{R}_{10} + g_2 \hat{a}_2 \hat{R}_{02}, \tag{8}$$

where Ω_j is the detuning of the jth field from the frequency of the corresponding atomic transition.

The density matrix $\hat{\sigma}(t)$ describing the atoms–field system satisfies the Liouville–von Neumann equation with the relaxation term

$$\frac{\partial}{\partial t}\hat{\sigma} = -\frac{i}{\hbar}[\hat{H},\hat{\sigma}] - \frac{1}{\tau}(\hat{\sigma} - \hat{\sigma}'), \tag{9}$$

where the equilibrium density matrix $\hat{\sigma}'$ is

$$\hat{\sigma}' = \hat{\rho} \otimes |\alpha_1\alpha_2\rangle_{ff}\langle\alpha_1\alpha_2|, \tag{10}$$

and the reduced density matrix $\hat{\rho}$ describes only the atomic degrees of freedom $\hat{\rho} = \mathrm{Tr}\{\hat{\sigma}\}_f$ (trace extends over the field variables). The equation (9) possesses a formal solution in the form of an infinite series

$$\hat{\sigma}(t) = e^{(-t/\tau)} \sum_{m=0}^{\infty} \frac{t^m}{m!} \hat{M}^m(\hat{\sigma}(0)) + \frac{1}{\tau}\int_0^t e^{-(t-t')/\tau}$$
$$\times \sum_{m=0}^{\infty} \frac{(t-t')^m}{m!} \hat{M}^m(\hat{\sigma}(t'))dt', \tag{11}$$

where

$$\hat{M}(\hat{\sigma}) = -\frac{i}{\hbar}[\hat{H},\hat{\sigma}].$$

Perturbation theory is applicable if

$$\frac{\tau g_j \sqrt{N}}{\hbar} \ll 1, \tag{12}$$

which should hold for both ($j = 1, 2$) modes. Therefore photons escape in a time shorter than the duration of the superradiance (SR) pulse (the duration of the SR pulse is $\tau_s \approx |g_j|^2 / h^2 N\tau$). For this reason, the variation of the field at times $t \gg \tau$ can be excluded adiabatically.

Expanding expression (11) in a series in τ, we obtain up to terms of order τ^2

$$\hat{\sigma}(t) = \hat{\sigma}'(t) - \tau\frac{\partial}{\partial t}\hat{\sigma}'(t) + \tau\hat{M}(\hat{\sigma}'(t)). \tag{13}$$

Substituting expression (13) into Eq. (9) and taking the trace of both parts of Eq. (9) over the field variables, we obtain an equation for the atomic density matrix

$$\frac{\partial}{\partial t}\hat{\rho} - \frac{i\tau}{\hbar}\left[\hat{H}_{at}, \frac{\partial}{\partial t}\hat{\rho}\right] = -\frac{i}{\hbar}[\hat{H}_{at},\hat{\rho}] - \frac{\tau}{\hbar^2}\mathrm{Tr}\{[\hat{H},[\hat{H},\hat{\rho}]]\}_f, \tag{14}$$

where we have introduced the Hamiltonian

$$\hat{H}_{at} = {}_f\langle\alpha_a\alpha_2|\hat{H}|\alpha_1\alpha_2\rangle_f$$

which acts only on the atomic variables. Next, we apply to both parts of Eq. (14) the operator

$$\hat{Q}(\hat{X}) = \hat{X} + \frac{i\tau}{\hbar}[\hat{H}_{at}, \hat{X}], \tag{15}$$

where $\hat{X}$ stands for Eq. (14).

According to the accuracy adopted above, we neglect terms proportional to τ^2

$$\frac{\partial}{\partial\tau}\hat{\rho} = -\frac{i}{\hbar}[\hat{H}_{at}, \hat{\rho}] - \frac{\tau}{\hbar^2}\{(\langle\hat{H}^2\rangle - \hat{H}_{at}^2)\hat{\rho}$$
$$- 2\,\mathrm{Tr}\{\hat{H}(\hat{\rho}\otimes|\alpha_1\alpha_2\rangle_{ff}\langle\alpha_1\alpha_2|)\hat{H}\}\}_f$$
$$- \frac{\tau}{\hbar^2}(\hat{\rho}(\langle\hat{H}^2\rangle - \hat{H}_{at}^2) + 2\hat{H}_{at}\hat{\rho}\hat{H}_{at}), \tag{16}$$

where $\langle\hat{H}^2\rangle = {}_f\langle\alpha_1\alpha_2|\hat{H}^2|\alpha_1\alpha_2\rangle_f$.

In the case of a Λ scheme the equation for the density matrix is

$$\hbar\frac{\partial}{\partial\tau}\hat{\rho} = -i[\hat{H}_{at},\hat{\rho}] + \tau|g_1|^2([\hat{R}_{10},\hat{\rho}\hat{R}_{01}] + [\hat{R}_{10}\hat{\rho},\hat{R}_{01}])$$
$$+ \tau|g_2|^2([\hat{R}_{20},\hat{\rho}\hat{R}_{02}] + [\hat{R}_{20}\hat{\rho},\hat{R}_{02}]). \tag{17}$$

This equation is the extension of the Bonifacio model.[7]

ATOMIC OPERATORS AND THEIR AVERAGE VALUES

Exhaustive information about the dynamics of the atomic system is contained in the average value of the collective atomic operator

$$r_{ij} = \frac{1}{N} \text{Tr}\{\hat{R}_{ij}\hat{\rho}\}, \quad i,j = 0,1,2. \tag{18}$$

Switching to the new variables r_{ij}, we write Eq. (17) as

$$N\frac{\partial}{\partial t}r_{jj'} = -\frac{i}{\hbar}Tr\{[\hat{H}_{at},\hat{\rho}]\hat{R}_{jj'}\} + \frac{\tau}{\hbar^2}$$
$$\times \sum_{j''=1}^{2} |g_{j''}|^2 \text{Tr}\{([\hat{R}_{j''0},\hat{\rho}\hat{R}_{0j''}]$$
$$+ [\hat{R}_{j''0}\hat{\rho},\hat{R}_{0j''}])\hat{R}_{jj'}\}. \tag{19}$$

Permuting cyclically within the trace, we obtain

$$N\frac{\partial}{\partial t}r_{jj'} = -\frac{i}{\hbar}\text{Tr}\{\hat{\rho}[\hat{H}_{at},\hat{R}_{jj'}]\} + \frac{\tau}{\hbar^2}$$
$$\times \sum_{j''=1}^{2} |g_{j''}|^2 \text{Tr}\{\hat{\rho}[\hat{R}_{j'j'},\hat{R}_{0j''}]$$
$$\times \hat{R}_{j''0} + \hat{R}_{0j''}[\hat{R}_{j''0}\hat{\rho},\hat{R}_{0j''}]\}.$$

We now employ the commutation rules (4), which decrease on the right-hand side the order with respect to $\hat{R}_{ij}$. It is easy to see that the commutator with the Hamiltonian is linear

$$[\hat{H}_{at},\hat{R}_{jj'}] = \sum_{m=0}^{2}\sum_{m'=0}^{2} L_{jj'}^{mm'}\hat{R}_{mm'}, \tag{20}$$

where the coefficients $L_{jj'}^{mm'}$ can be calculated easily. The explicit form of $L_{jj'}^{mm'}$ in the Appendix gives

$$\frac{\partial}{\partial t}r_{jj'} = \frac{i}{\hbar}\sum_{m=0}^{2}\sum_{m'=0}^{2} L_{jj'}^{mm'}r_{mm'} - \frac{\tau}{\hbar^2 N}$$
$$\times \left\{ \delta_{j'0}\sum_{j''=1}^{2} |g_{j''}|^2 \text{Tr}\{\hat{\rho}\hat{R}_{jj''}\hat{R}_{j''0}\} + \delta_{j0} \right.$$
$$\times \sum_{j''=1}^{2} |g_{j''}|^2 \text{Tr}\{\hat{\rho}\hat{R}_{0j''}\hat{R}_{j''j'}\} - (|g_j|^2(1-\delta_{j0})$$
$$\left. + |g_{j'}|^2(1-\delta_{j'0}))\text{Tr}\{\hat{\rho}\hat{R}_{0j'}\hat{R}_{j0}\} \right\}. \tag{21}$$

As is customary in the theory of cooperative effects, we expand the correlator using the quasiclassical approximation

$$\frac{1}{N^2}\text{Tr}\{\hat{\rho}\hat{R}_{ij}\hat{R}_{i'j'}\} \approx r_{ij}r_{i'j'}. \tag{22}$$

The approximate nature of the simplification (22) is expressed only in the fact that it is impossible to describe correctly the early stage of the dynamics of the system in the case when there is no external laser field. From Eq. (21) we obtain

$$\frac{\partial}{\partial t}r_{jj'} = \frac{i}{\hbar}\sum_{m=0}^{2}\sum_{m'=0}^{2} L_{jj'}^{mm'}r_{mm'} - \delta_{j0}\sum_{j''=1}^{2} \Gamma_{j''}Nr_{0j''}r_{j''j'}$$

$$-\delta_{j'0}\sum_{j''=1}^{2} \Gamma_{j''}Nr_{jj''}r_{j''0} + [\Gamma_j(1-\delta_{j0})$$
$$+ \Gamma_{j'}(1-\delta_{j'0})]Nr_{0j'}r_{j0}, \tag{23}$$

where we have introduced the notation

$$\Gamma_\mu = \frac{\tau}{\hbar^2}|g_\mu|^2.$$

It is important to note that

$$r_{11} + r_{22} + r_{33} = 1 \tag{24}$$

and the approximation (22) has no effect on this condition. After calculating the coefficients $L_{jj'}^{mm'}$, which are either the detunings Ω_j or the atom–field coupling constants $V_j = g_j\alpha_j/\hbar$, we can write Eq. (23) as

$$\dot{r}_{11} = iV_1r_{01} - iV_1^*r_{10} + 2\Gamma_1 N|r_{01}|^2,$$
$$\dot{r}_{22} = iV_2r_{02} - iV_2^*r_{20} + 2\Gamma_2 N|r_{02}|^2,$$
$$\dot{r}_{01} = i\Omega_1 r_{01} + iV_1^*(r_{11}-r_{00}) + iV_2^*r_{21}$$
$$- \Gamma_1 Nr_{01}(r_{11}-r_{00}) - \Gamma_2 Nr_{02}r_{21},$$
$$\dot{r}_{02} = i\Omega_2 r_{02} + iV_1^*r_{12} - iV_2^*(r_{00}-r_{22})$$
$$- \Gamma_1 Nr_{01}r_{12} - \Gamma_2 Nr_{02}(r_{22}-r_{00}),$$
$$\dot{r}_{21} = i(\Omega_1-\Omega_2)r_{21} + iV_2r_{01} - iV_1^*r_{20}$$
$$+ (\Gamma_1+\Gamma_2)Nr_{20}r_{01}. \tag{25}$$

The Hermiticity condition $r_{j'j} = r_{jj'}^*$, then holds. The system of equations (25) for the average values of the collective atomic operators is similar to the system of equations for the single-particle density matrix in the case of incoherent decay. Indeed, the diagonal elements r_{ii} give us the populations of the corresponding states (per atom). Nonetheless, there is an important distinction: In the case of the cooperative effect the relaxation terms are nonlinear and proportional to the total number N of atoms. As the initial condition in the system (25), we assume that all atoms initially occupy the lower energy level $|1(INT)\rangle$, i.e.,
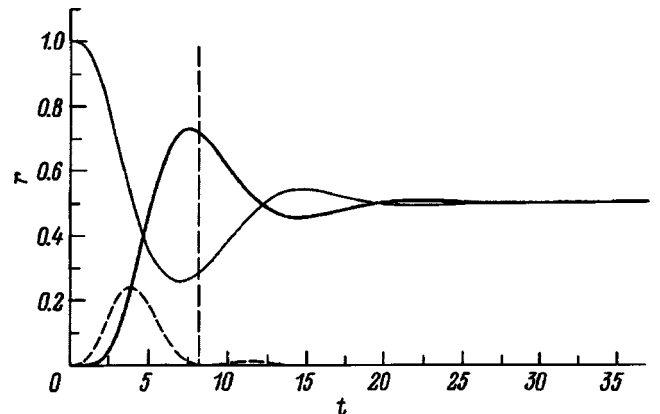

FIG. 1. Temporal evolution of populations in the Λ system for $\Omega = 0$, $V = 0.3\Gamma N$. The dimensionless time $t/\Gamma N$ is plotted along the abscissa. Thin curve—level 1, thick curve—level 2, dashed curve—level zero.
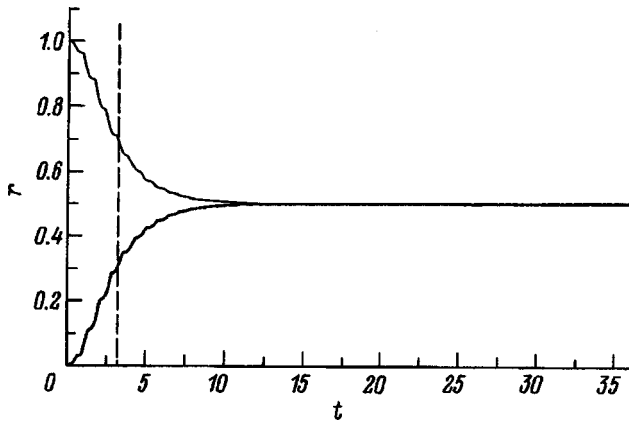
FIG. 2. Same as Fig. 1 but for $\Omega = 0$, $V = 3\Gamma N$.

FIG. 3. Same as Fig. 1 but for $\Omega = 2\Gamma N$, $V = 3\Gamma N$.

$$r_{jj'}|_{t=0} = \delta_{j1}\delta_{j'1}. \tag{26}$$

The off-diagonal elements of the matrix r_{ij} equal zero.

COMPUTATIONAL RESULTS

We shall present the results obtained by integrating numerically the system of equations (25) with the initial conditions (26). As a simplification we set $\Gamma_1 = \Gamma_2 \equiv \Gamma$, $V_1 = V_2 \equiv V$, and $\Omega_1 = \Omega_2 \equiv \Omega$.

The temporal evolution of the atomic populations r_{ii} is displayed in Figs. 1–4. In the case $\Omega = 0$ the behavior of the system corresponds to a transition to CPT,[8] where the states $|1\rangle$ and $|2\rangle$ form a noninteracting superposition, while the population of the upper level $|0\rangle$ becomes zero (Figs. 1 and 2). There is an important distinction from the ordinary transition to CPT: In the presence of cooperative relaxation in the system the population of the upper level reaches a steady
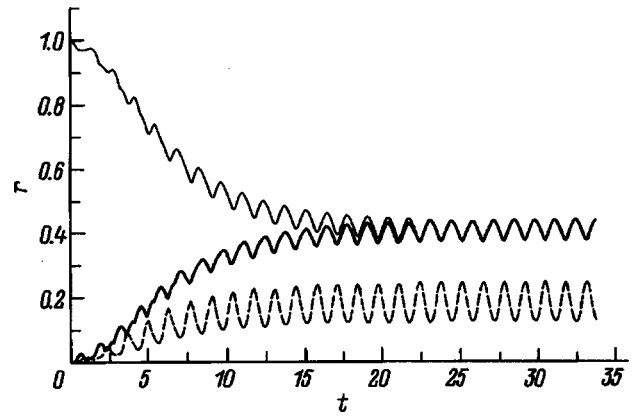
regime much earlier than the population of the lower levels (Figs. 1 and 2). In the case of incoherent relaxation, however, the system of equations describing the population dynamics is linear,[8] so that the evolution of all populations is determined by the roots of the same characteristic equation. In the nonlinear case of cooperative dynamics the oscillations of the populations predominate at the initial stage of evolution for $V < \Gamma N$ (Fig. 1). In the case $V > \Gamma N$ oscillations are suppressed (Fig. 2). Conversely, in the case when incoherent relaxation predominates the oscillations are suppressed at Rabi frequencies less than the relaxation rate from the upper level and are developed in the opposite case.[8] We note also that the stationary state $r_{11} = 0.5$, $r_{22} = 0.5$, $r_{12} = -0.5$, and all other elements $r_{ij} = 0$ is an exact solution of the system of equation (25).

In the case $\Omega \neq 0$ the population dynamics is more complicated. The steady state of the system is an oscillatory re-
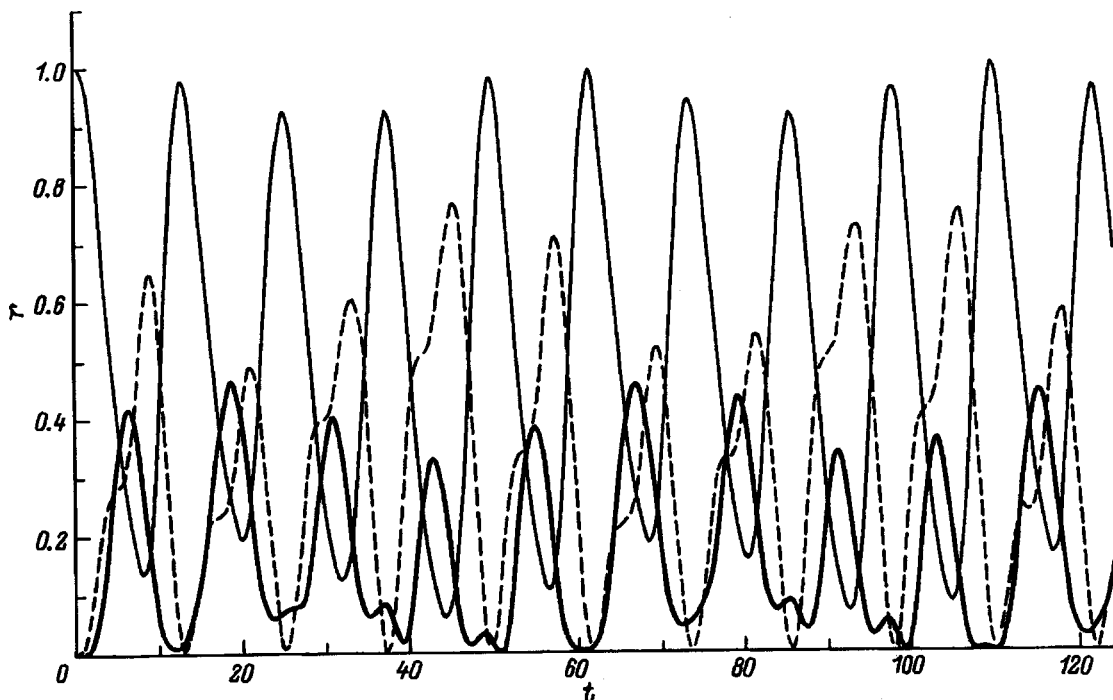

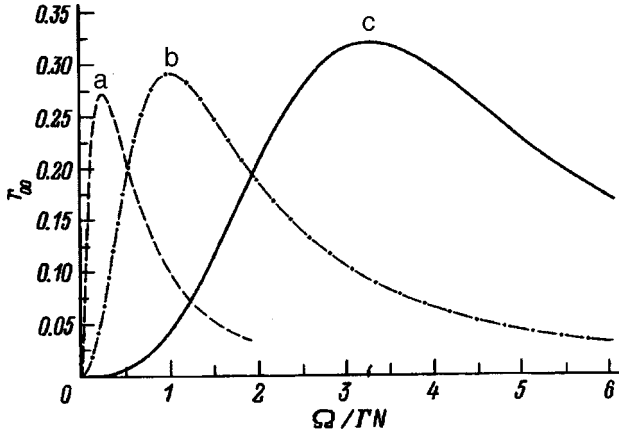FIG. 4. Same as Fig. 1 but for $\Omega = 1.5\Gamma N$, $V = 1.5\Gamma N$.

FIG. 5. Time-averaged population r_{00} of the upper level $|0\rangle$ versus the dimensionless detuning $\Omega/\Gamma N$. $V/\Gamma N$: a—0.5, b—1.0, c—2.0.

FIG. 6. τ_{rel}^{-1}, $\delta\Omega_{BL}$, and F (in units of ΓN) versus the dimensionless coupling parameter k. Computed values: $\bigcirc$—τ_{rel}^{-1}, $\triangle$—$\delta\Omega_{BL}$, $\blacksquare$—F.

gime (Figs. 3 and 4), in contrast to CPT in the ordinary, noncooperative case, where the steady state corresponds to constant populations.[1,8] These oscillations in the case of cooperative dynamics are periodic (Fig. 3). When Ω and V become equal to one another, the period increases—nonlinearity appears in the behavior of the quantum system (Fig. 4).

Figure 5 shows the time-averaged value of the population of the level $|0\rangle$ versus the dimensionless detuning $\Omega/\Gamma N$ for different values of the atom-field coupling constant. One can see that the width of the black line increases with the coupling constant V.

We now introduce the following three characteristics: the characteristic time τ_{rel} during which coherent population trapping is established in the system, the width $\delta\Omega_{BL}$ of the black line, and the width F of the overall contour. Analysis of the dependences of the introduced quantities on the value of the parameter $k = V/\Gamma N$ gives simple approximation formulas for all three quantities characterizing the quantum system:

$$\frac{\tau_{\mathrm{rel}}^{-1}}{\Gamma N} = \frac{0.03k^2}{0.17 + k^2}, \qquad \frac{\delta\Omega_{BL}}{\Gamma N} = 0.005 + \frac{0.70k^2}{0.90 + k},$$

$$\frac{F}{\Gamma N} = 2.3k^{0.92}. \tag{27}$$

One can see that the transition rate τ_{rel}^{-1} to CPT under two-photon resonance conditions (i.e., $\Omega_1 = \Omega_2$) behaves just as in the ordinary single-atom case: It increases quadratically for $k \ll 1$ and is constant in the opposite case. Further, the dependence of the width of the black line is similar to the ordinary, noncooperative case.[1] The width of the overall contour satisfying a power law with an exponent close to 1 in a wide range of values of the parameter k differs from the noncooperative dynamics, where it is constant for $k \ll 1$. The computed points for the three characteristics τ_{rel}^{-1}, $\delta\Omega_{BL}$, and F and the curves corresponding to the approximation formulas (27) are presented in Fig. 6.
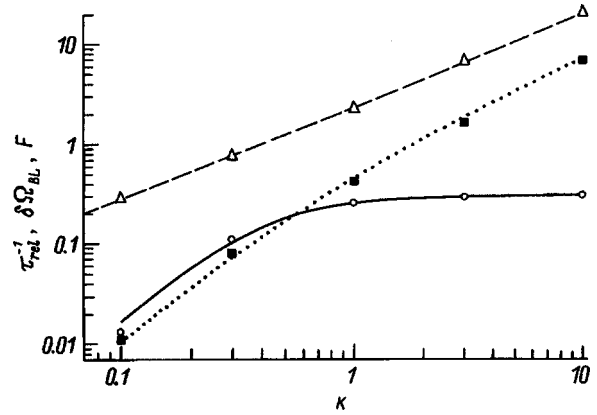
CONCLUSIONS

The transition to CPT in an atomic ensemble interacting with two coherent resonance electromagnetic fields under cooperative relaxation conditions was investigated theoretically. The results obtained make it possible to apply the phenomenon of coherent population trapping, for example, for investigating electromagnetically induced transparency for purposes of coherent bleaching[9,10] in dense optical media with average interatomic distance of the order of the wavelength of the atomic transition $n\lambda^3 \sim 1$. It was established that cooperative relaxation introduces substantial changes in the dynamics of the ensemble of atoms. The features that distinguish the cooperative case from the case where incoherent spontaneous relaxation predominates were clarified and parameters such as the width of the black line, the contour width, and the transition rate to CPT in the system studied above were calculated.

APPENDIX

We present here the explicit form of the nonzero coefficients $L_{jj'}^{mm'}$ appearing in expression (20): $L_{11}^{01} = iV_1$, $L_{11}^{10} = -iV_1^*$, $L_{22}^{02} = iV_2$, $L_{22}^{20} = iV_2^*$, $L_{01}^{01} = i\Omega_1$, $L_{01}^{11} = iV_1^*$, $L_{01}^{00} = -iV_1^*$, $L_{02}^{02} = i\Omega_2$, $L_{02}^{12} = iV_1^*$, $L_{02}^{00} = -iV_2^*$, $L_{02}^{22} = iV_2^*$, $L_{21}^{21} = i(\Omega_1 - \Omega_2)$, $L_{21}^{01} = iV_2$, $L_{21}^{20} = iV_1$. The coefficients $L_{jj'}^{mm'}$ satisfy the relation $L_{jj'}^{mm'} = (L_{j'j}^{m'm})^*$.

[1] B. G. Agap'ev, M. B. Gornyĭ, B. G. Matisov *et al.*, Usp. Fiz. Nauk **163**, 1 (1993).

[2] E. Arimondo, in *Progress in Optics*, edited by E. Wolf, North-Holland, Amsterdam, 1995, No. 35, pp. 257–269.

[3] E. Arimondo, Phys. Rev. A **4**, 2216 (1996).

[4] A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Effects in Optics* [in Russian], Nauka, Moscow, 1988, 288 pp.

[5] M. M. Al'perin, Ya. D. Klubis, and A. I. Khizhnyak, *Introduction to the Physics of Two-Level Systems* [in Russian], Naukova Dumka, Kiev, 1987, 256 pp.

[6] R. H. Dicke, Phys. Rev. **93**, 99 (1954).

[7] R. Bonifacio, P. Schwendimann, and F. Haake, Phys. Rev. A **93**, 302 (1971).

[8] E. A. Korsunskiĭ, B. G. Matisov, and Yu. V. Rozhdestvenskiĭ, Zh. Éksp. Teor. Fiz. **102**, 1096 (1992) [Sov. Phys. JETP **75**, 595 (1992)].

[9] M. B. Gornyĭ, B. G. Matisov, and Yu. V. Rozhdestvenskiĭ, Zh. Éksp. Teor. Fiz. **95**, 1263 (1989) [Sov. Phys. JETP **68**, 728 (1989)].

[10] K. J. Boller, A. Imamoglu, and S. E. Harris, Phys. Rev. Lett. **66**, 2593 (1991).

Thermophoresis of touching solid spheres in the direction along the line joining their centers

S. N. D'yakonov and Yu. I. Yalamov

Moscow Pedagogical University, 107005 Moscow, Russia

A theory of the steady motion of an aggregate of two touching solid, nonvolatile, low-thermal-conductivity, spherical particles in the direction along the line joining their centers in a nonuniformly heated viscous gas is constructed in a hydrodynamic regime with slipping at low Reynolds and Peclet numbers. The thermophoretic transport velocity of an aggregate is determined in an approximation linear in the small parameters. The small parameters are the relative deviations of the thermal conductivity of the constituent particles of an aggregate from the thermal conductivity of the external medium. © *1998 American Institute of Physics.* [S1063-7842(98)00606-0]

INTRODUCTION

The study of particle dynamics in a viscous medium with nonuniform temperature is of interest in connection with investigations of thermophoretic motion in the physics of aerodispersion systems for analysis of the interaction of hot particles, in mechanics and rheology of suspensions, and in a number of other problems. Thus far the characteristics of thermophoretic motion in viscous media of single solid and liquid aerosol particles have been studied in greatest detail. A general bibliography concerning these questions is presented in Ref. 1.

It is more important to study the motion of a particle ensemble, since particles which are sufficiently close to one another strongly influence the relative particle motions. In aerosol systems, in practice, pairs of particles are most likely to approach one another. It is thus of interest to study the dynamics of such pairs.

The thermophoresis of two spherical aerosol particles in the direction along the line joining their centers was investigated in Refs. 2–5. Exact analytical solutions for particles located at an arbitrary distance but quite far from one another were obtained in a bispherical coordinate system on the basis of linearized stationary equations of hydrodynamics and heat transfer. A numerical comparison of these solutions with approximate solutions obtained by the method of reflections shows that for close-lying particles there is a degradation of the convergence of the approximate solutions. In the limiting case when the particles touch the exact analytical solutions are suitable only for estimating the instantaneous velocities of the steady motion of the particles. These estimates are strongly limited by the conditions of applicability of the linear stationary equations of slow flow.

In the present paper we construct on the basis of a hydrodynamic analysis a theory of the motion of two solid nonvolatile contiguous spheres in the direction along the line joining their centers in a nonuniformly heated viscous gas. This work is necessitated by the fact that the limiting problem of two touching spheres cannot be solved on the basis of theories employing a bispherical coordinate system and the Stimson–Jeffery approach to the hydrodynamic problem.[6] The endeavor to obtain an analytical solution of this problem for the linearized stationary equations of hydrodynamics and heat transfer makes it possible to obtain more reliable results for larger temperature differentials than those obtained earlier in the limiting case of touching particles.

FORMULATION OF THE PROBLEM

We shall study the slow motion of an aggregate of two touching, solid, nonvolatile, low-thermal-conductivity spherical particles in the direction along the line joining their centers in a temperature-nonuniform viscous gaseous medium.

The problem of determining the thermophoretic velocity $\mathbf{U}_T$ of an aggregate can be solved in tangential spherical coordinates (ζ, η, φ) related to the circular cylindrical coordinates (γ, z, φ) as

$$\gamma = \frac{2\zeta}{\zeta^2 + \eta^2}, \quad z = \frac{2\eta}{\zeta^2 + \eta^2}, \quad \varphi = \varphi. \tag{1}$$

The origin of the cylindrical coordinate system is rigidly fixed at the contact point of the particles. In this coordinate system the center of gravity of the exterior medium moves with the velocity $\mathbf{U} = -\mathbf{U}_T$, which is sought, relative to the stationary aggregate.

A constant temperature gradient $\mathbf{A}_T = (\nabla T^{(e)})_\infty$ is maintained in the gas infinitely far away from the particle aggregate. Here and below the superscripts e and i denote physical quantities in a region outside and inside an aggregate, respectively, while a subscript α ($\alpha = 1, 2$) refers to a definite particle.

Let the axis $z = (r \cos \theta)$ pass through the centers of the touching spheres and be directed parallel to the vector $\mathbf{A}_T$. A nonuniform temperature distribution near the particles results in the appearance of a directed (thermophoretic) motion of the aggregate (on account of the thermal slipping of the gas along the surfaces of the solid spheres).

The external medium is assumed to be single-component, isotropic, incompressible, and continuous—the Knudsen number $Kn = \lambda/R_0 \ll 1$ (where λ, $R_0 = R_1 + R_2$, $R_1 = \eta_1^{-1}$, and $R_2 = \eta_2^{-1}$ are, respectively, the average mean free path length of the gas molecules, the unit of length, the radii of curvature of the surfaces $\eta = \eta_1 > 0$ and $\eta = -\eta_2 < 0$ of the particles in an aggregate).

We assume that each particle consists of a material which is uniform and has isotropic properties.

Thermophoresis of an aggregate occurs at low Reynolds and Peclet numbers $Re^{(e)} = UR_0/\nu^{(e)} \ll 1$ and $Re_T^{(e)} = UR_0/\chi^{(e)} \ll 1$. This makes it possible to drop the nonlinear (inertial and convection) terms in the equations of hydrodynamics and heat transfer. The external mass forces are neglected. There are no heat sources inside or outside the particles.

The relative temperature differentials under the conditions of the problem are small and the temperature variation of the coefficients of molecular transport can be neglected. The density, kinematic viscosity, and thermal conductivity are assumed to be constants ($\rho_0^{(e,i)}, \nu_0^{(e)}, \varkappa_0^{(e,i)}$) at the unperturbed temperature $T_0^{(e)}$ (the temperature of the external medium at the location of the point of contact of the particles in an aggregate in the absence of the aggregate). However, the existing temperature differentials are large enough so that in comparison the temperature variations due to heating as a result of energy dissipation by internal friction can be neglected in the heat-transfer equation.

Since the thermal and hydrodynamic relaxation times of the system are short, the motions of an aggregate can be described in a quasistationary approximation (a slow axisymmetric motion of the gas medium and a steady temperature distribution inside and outside the particles).

The following boundary conditions hold at infinity and on the surface of the particles:

$$r \to \infty: \quad \mathbf{v}^{(e)} = U\mathbf{i}_z, \quad T^{(e)} = T_0^{(e)} + A_z z, \tag{2}$$

$$S_\alpha(\alpha = 1,2): \quad (\mathbf{i}_\eta \cdot \mathbf{v}^{(e)}) = 0 \tag{3}$$

$$(\mathbf{i}_\zeta \cdot \mathbf{v}^{(e)}) = K_{TSL}^{(e)} \frac{\nu_0^{(e)}}{T_o^{(e)}} (\mathbf{i}_\zeta \cdot \nabla T^{(e)}), \tag{4}$$

$$T^{(e)} = T_\alpha^{(i)}, \tag{5}$$

$$\varkappa_0^{(e)}(\mathbf{i}_\eta \cdot \nabla T^{(e)}) = \varkappa_{0\alpha}^{(i)}(\mathbf{i}_\eta \cdot \nabla T_\alpha^{(i)}), \tag{6}$$

$$F_z(\text{aggregate}) = 0. \tag{7}$$

The conditions (2)–(7) physically signify the following.

At infinity the axisymmetric gas flow is uniform in space and its velocity $\mathbf{U}$ is in the positive direction along the z axis, while the external temperature field is unperturbed.

On the gas-impermeable surface S_α of the solid nonvolatile particles the normal velocity component $v_\eta^{(e)}$ of the external medium vanishes, while the tangential component $v_\zeta^{(e)}$ equals the velocity of thermal slipping (it is characterized by the coefficient $K_{TSL}^{(e)}$, determined by methods of the kinetic theory of gases); the normal heat flux and temperature are continuous.

The resultant force $\mathbf{F}$ exerted on an aggregate by the incident flow of the external medium equals zero.

It is convenient to write the equations of hydrodynamics and heat transfer and the boundary conditions in a reduced form. The physical quantities entering in the equations are made dimensionless as follows:

$$\tilde{\gamma} = \frac{\gamma}{R_0}, \quad \tilde{z} = \frac{r}{R_0}, \quad \tilde{r} = \frac{r}{R_0}, \quad \tilde{\zeta} = \zeta R_0, \quad \tilde{\eta} = \eta R_0,$$

$$\tilde{v}_\zeta^{(e)} = \frac{v_\zeta^{(e)}}{U_{[0]}}, \quad \tilde{v}_\eta^{(e)} = \frac{v_\eta^{(e)}}{U_{[0]}}, \quad \tilde{U} = \frac{U}{U_{[0]}}, \quad \tilde{\Psi}^{(e)} = \frac{\Psi^{(e)}}{R_0^2 U_{[0]}},$$

$$\tilde{F}_z = \frac{F_z}{6\pi \eta_0^{(e)} R_0 U_{[0]}}, \quad \tilde{T}^{(e)} = \frac{T^{(e)} - T_0^{(e)}}{A_T R_0},$$

$$\tilde{T}_\alpha^{(i)} = \frac{T_\alpha^{(i)} - T_0^{(e)}}{A_T R_0},$$

$\Psi^{(e)}$ is the stream function and $U_{[0]}$ is the velocity of the gas flow at infinity in the zeroth approximation in the small parameters (it is determined in the course of the solution).

In what follows the tilde is dropped and the initial equations and boundary conditions are written in the reduced form as follows:

$$E^4\Psi^{(e)} = 0, \tag{8}$$

$$\Delta T^{(e)} = \Delta T_\alpha^{(i)} = 0, \tag{9}$$

$$r \to \infty: \quad \Psi^{(e)} = -\frac{1}{2}\gamma^2 U, \tag{10}$$

$$T^{(e)} = Z, \tag{11}$$

$$S_\alpha \ (\alpha = 1,2): \quad \Psi^{(e)} = 0, \tag{12}$$

$$U_{[0]} \frac{\partial \Psi^{(e)}}{\partial \eta} = -K_{TSL}^{(e)} \frac{\nu_0^{(e)}}{T_0^{(e)}} A_T \gamma \frac{\partial T^{(e)}}{\partial \zeta}, \tag{13}$$

$$T^{(e)} = T_\alpha^{(i)}, \tag{14}$$

$$\frac{\varkappa_0^{(e)}}{\varkappa_{0\alpha}^{(i)}} \nabla_\eta T^{(e)} = \nabla_\eta T_\alpha^{(i)}, \tag{15}$$

$$F_z(\text{aggregate}) = 0,$$

$$E^2 = \frac{\partial^2}{\partial\gamma^2} - \frac{1}{\gamma}\frac{\partial}{\partial\gamma} + \frac{\partial^2}{\partial z^2}, \quad \Delta = \frac{\partial^2}{\partial\gamma^2} + \frac{1}{\gamma}\frac{\partial}{\partial\gamma} + \frac{\partial^2}{\partial z^2}. \tag{16}$$

The reduced velocity U of the gas flow at infinity is sought in the form of a power series in the small parameters ε_1 and ε_2

$$0 \le \varepsilon_1 = (\varkappa_{01}^{(i)} - \varkappa_0^{(e)})/\varkappa_{01}^{(i)} \ll 1,$$

$$0 \le \varepsilon_2 = (\varkappa_{02}^{(i)} - \varkappa_0^{(e)})/\varkappa_{02}^{(i)} \ll 1,$$

$$U(\varepsilon_1, \varepsilon_2) = 1 + \varepsilon_1 U_{[1]}^{(1)} + \varepsilon_2 U_{[1]}^{(2)} + \varepsilon_1^2 U_{[2]}^{(1)}$$

$$+ \varepsilon_1\varepsilon_2 U_{[2]}^{(2)} + \varepsilon_2^2 U_{[3]}^{(2)} + \dots \,. \tag{17}$$

We shall confine ourselves below to determining the quantities $U_{[0]}$, $U_{[1]}^{(1)}$, and $U_{[2]}^{(2)}$, which characterize the thermophoretic velocity of an aggregate in the zeroth and first approximations (the bracketed subscripts).

THERMAL PROBLEM

An exact analytical solution of the thermal problem (9), (11), (14), and (15) cannot be obtained for arbitrary values of the reduced thermal conductivities $\varkappa_1 = \varkappa_0^{(e)}/\varkappa_{01}^{(i)}$ and $\varkappa_2 = \varkappa_0^{(e)}/\varkappa_{02}^{(i)}$. An approximate solution of the thermal (hydrodynamic) problem can be constructed by the method of successive approximations in the small parameters ε_1 and ε_2 (the case of an aggregate of touching low-thermal-conductivity particles $\varkappa_1 \sim \varkappa_2 \sim 1$). The reduced axisymmetric temperature inside and outside the particles is sought in the form

$$T^{(e)}(\zeta,\eta) = T_{[0]}^{(e)}(\zeta,\eta) + T_{[1]}^{(e)}(\zeta,\eta) + T_{[2]}^{(e)}(\zeta,\eta) + \ldots,$$

$$T_\alpha^{(i)}(\zeta,\eta) = T_{\alpha,[0]}^{(i)}(\zeta,\eta) + T_{\alpha,[1]}^{(i)}(\zeta,\eta) + T_{\alpha,[2]}^{(i)}(\zeta,\eta) + \ldots .$$

Here the temperature perturbations $T_{[1]}^{(e)}(\zeta,\eta)$ and $T_{\alpha,[1]}^{(i)} \times (\zeta,\eta)$ in an approximation linear in the small parameters can be written as

$$T_{[1]}^{(e)}(\zeta,\eta) = \varepsilon_1 t_1^{(e)}(\zeta,\eta) + \varepsilon_2 t_2^{(e)}(\zeta,\eta),$$

$$T_{\alpha,[1]}^{(i)}(\zeta,\eta) = \varepsilon_1 t_{\alpha,1}^{(i)}(\zeta,\eta) + \varepsilon_2 t_{\alpha,2}^{(i)}(\zeta,\eta).$$

Since the Laplacian is a linear operator, the functions

$$T_{[0]}^{(e)}(\zeta,\eta), \quad T_{\alpha,[0]}^{(i)}(\zeta,\eta),$$

$$t_1^{(e)}(\zeta,\eta), \quad t_2^{(e)}(\zeta,\eta), \quad t_{\alpha,1}^{(i)}(\zeta,\eta), \quad t_{\alpha,2}^{(i)}(\zeta,\eta)$$

are solutions of Eqs. (9) in a tangential spherical system (ζ,η,φ) of coordinates of revolution[7]

$$T_{[0]}^{(e)}(\zeta,\eta) = z + (\zeta^2 + \eta^2)^{1/2} \int_0^\infty (A_{[0]}(\lambda)\cosh(\lambda\eta) + B_{[0]}$$
$$\times (\lambda)\sinh(\lambda\eta)) J_0(\lambda\zeta) d\lambda, \tag{18}$$

$$T_{1,[0]}^{(i)}(\zeta,\eta) = (\zeta^2 + \eta^2)^{1/2} \int_0^\infty C_{[0]}(\lambda) e^{-\lambda\eta} J_0(\lambda\zeta) d\lambda, \tag{19}$$

$$T_{2,[0]}^{(i)}(\zeta,\eta) = (\zeta^2 + \eta^2)^{1/2} \int_0^\infty D_{[0]}(\lambda) e^{+\lambda\eta} J_0(\lambda\zeta) d\lambda, \tag{20}$$

$$t_j^{(e)}(\zeta,\eta) = (\zeta^2 + \eta^2)^{1/2} \int_0^\infty (A_{[1]}^{(j)}(\lambda)\cosh(\lambda\eta) + B_{[1]}^{(j)}$$
$$\times (\lambda)\sinh(\lambda\eta)) J_0(\lambda\zeta) d\lambda, \tag{21}$$

$$t_{1,j}^{(i)}(\zeta,\eta) = (\zeta^2 + \eta^2)^{1/2} \int_0^\infty C_{[1]}^{(j)}(\lambda) e^{-\lambda\eta} J_0(\lambda\zeta) d\lambda, \tag{22}$$

$$t_{2,j}^{(i)}(\zeta,\eta) = (\zeta^2 + \eta^2)^{1/2} \int_0^\infty D_{[1]}^{(j)}(\lambda) e^{+\lambda\eta} J_0(\lambda\zeta) d\lambda$$

$$(j = 1; 2). \tag{23}$$

The functions (18)–(20) satisfy the conditions

$$T_{[0]}^{(e)}(\zeta,\eta)|_{S_\alpha} = T_{\alpha,[0]}^{(i)}(\zeta,\eta)|_{S_\alpha}, \tag{24}$$

$$(\nabla_\eta T_{[0]}^{(e)}(\zeta,\eta))|_{S_\alpha} = (\nabla_\eta T_{\alpha,[0]}^{(i)}(\zeta,\eta))|_{S_\alpha}, \tag{25}$$

$$T_{[0]}^{(e)}(\zeta,\eta)|_{r\to\infty} = z, \quad T_{\alpha,[0]}^{(i)}(\zeta,\eta)|_{\substack{\zeta\to\infty \\ \eta\to\infty}} < \infty. \tag{26}$$

The following algebraic equations follow from the boundary conditions (24) and (25) with allowance for the integral transformations (A2) and (A3) in the Appendix:

$$2\lambda \exp(-\lambda\eta_1) + A_{[0]}(\lambda)\cosh(\lambda\eta_1) + B_{[0]}$$
$$\times (\lambda)\sinh(\lambda\eta_1) = C_{[0]}(\lambda)\exp(-\lambda\eta_1),$$

$$-2\lambda \exp(-\lambda\eta_1) + A_{[0]}(\lambda)\sinh(\lambda\eta_1) + B_{[0]}$$
$$\times (\lambda)\cosh(\lambda\eta_1) = -C_{[0]}(\lambda)\exp(-\lambda\eta_1),$$

$$-2\lambda \exp(-\lambda\eta_2) + A_{[0]}(\lambda)\cosh(\lambda\eta_2) - B_{[0]}$$
$$\times (\lambda)\sinh(\lambda\eta_2) = D_{[0]}(\lambda)\exp(-\lambda\eta_2),$$

$$2\lambda \exp(-\lambda\eta_2) + A_{[0]}(\lambda)\sinh(\lambda\eta_2) - B_{[0]}$$
$$\times (\lambda)\cosh(\lambda\eta_2) = -D_{[0]}(\lambda)\exp(-\lambda\eta_2),$$

whence we have

$$A_{[0]}(\lambda) = B_{[0]}(\lambda) = 0, \quad C_{[0]}(\lambda) = -D_{[0]}(\lambda) = 2\lambda, \tag{27}$$

$$T_{[0]}^{(e)}(\zeta,\eta) = T_{\alpha,[0]}^{(i)}(\zeta,\eta) = z. \tag{28}$$

Thus in the zeroth approximation $\varepsilon_1 = \varepsilon_2 = 0$ the temperature gradient is constant everywhere in space.

The functions (21)–(23) satisfy the conditions

$$t_1^{(e)}(\zeta,\eta)|_{S_1} = t_{1,1}^{(i)}(\zeta,\eta)|_{S_1}, \tag{29}$$

$$(\nabla_\eta t_1^{(e)}(\zeta,\eta) - \nabla_\eta T_{[0]}^{(e)}(\zeta,\eta))|_{S_1} = (\nabla_\eta t_{1,1}^{(i)}(\zeta,\eta))|_{S_1}, \tag{30}$$

$$t_2^{(e)}(\zeta,\eta)|_{S_1} = t_{1,2}^{(i)}(\zeta,\eta)|_{S_1}, \tag{31}$$

$$(\nabla_\eta t_2^{(e)}(\zeta,\eta))|_{S_1} = (\nabla_\eta t_{1,2}^{(i)}(\zeta,\eta))|_{S_1}, \tag{32}$$

$$t_1^{(e)}(\zeta,\eta)|_{S_2} = t_{2,1}^{(i)}(\zeta,\eta)|_{S_2}, \tag{33}$$

$$(\nabla_\eta t_1^{(e)}(\zeta,\eta))|_{S_2} = (\nabla_\eta t_{2,1}^{(i)}(\zeta,\eta))|_{S_2}, \tag{34}$$

$$t_2^{(e)}(\zeta,\eta)|_{S_2} = t_{2,2}^{(i)}(\zeta,\eta)|_{S_2}, \tag{35}$$

$$(\nabla_\eta t_2^{(e)}(\zeta,\eta) - \nabla_\eta T_{[0]}^{(e)}(\zeta,\eta))|_{S_2} = (\nabla_\eta t_{2,2}^{(i)}(\zeta,\eta))|_{S_2},$$

$$t_j^{(e)}(\zeta,\eta)|_{r\to\infty} = 0, \quad t_{1,j}^{(i)}(\zeta,\eta)|_{\substack{\zeta\to\infty \\ \eta\to\infty}} < \infty, \tag{36}$$

$$t_{2,j}^{(i)}(\zeta,\eta)|_{\substack{\zeta\to\infty \\ \eta\to\infty}} < \infty. \tag{37}$$

From the boundary conditions (29)–(36) with relations (A2) and (A3) taken into account it is easy to obtain the algebraic equations

$$A_{[1]}^{(1)}(\lambda)\cosh(\lambda\,\eta_1)+B_{[1]}^{(1)}(\lambda)\sinh(\lambda\,\eta_1)$$
$$-C_{[1]}^{(1)}(\lambda)\exp(-\lambda\,\eta_1)=0,$$

$$A_{[1]}^{(1)}(\lambda)\sinh(\lambda\,\eta_1)+B_{[1]}^{(1)}(\lambda)\cosh(\lambda\,\eta_1)+C_{[1]}^{(1)}(\lambda)$$
$$\times\exp(-\lambda\,\eta_1)=-\frac{2}{3}\left(2\lambda-\frac{1}{\eta_1}\right)\exp(-\lambda\,\eta_1),$$

$$A_{[1]}^{(2)}(\lambda)\cosh(\lambda\,\eta_1)+B_{[1]}^{(2)}(\lambda)\sinh(\lambda\,\eta_1)$$
$$-C_{[1]}^{(2)}(\lambda)\exp(-\lambda\,\eta_1)=0,$$

$$A_{[1]}^{(2)}(\lambda)\sinh(\lambda\,\eta_1)+B_{[1]}^{(2)}(\lambda)\cosh(\lambda\,\eta_1)$$
$$+C_{[1]}^{(2)}(\lambda)\exp(-\lambda\,\eta_1)=0,$$

$$A_{[1]}^{(1)}(\lambda)\cosh(\lambda\,\eta_2)-B_{[1]}^{(1)}(\lambda)\sinh(\lambda\,\eta_2)$$
$$-D_{[1]}^{(1)}(\lambda)\exp(-\lambda\,\eta_2)=0,$$

$$A_{[1]}^{(1)}(\lambda)\sinh(\lambda\,\eta_2)-B_{[1]}^{(1)}(\lambda)\cosh(\lambda\,\eta_2)$$
$$+D_{[1]}^{(1)}(\lambda)\exp(-\lambda\,\eta_2)=0,$$

$$A_{[1]}^{(2)}(\lambda)\cosh(\lambda\,\eta_2)-B_{[1]}^{(2)}(\lambda)\sinh(\lambda\,\eta_2)$$
$$-D_{[1]}^{(2)}(\lambda)\exp(-\lambda\,\eta_2)=0,$$

$$A_{[1]}^{(2)}(\lambda)\sinh(\lambda\,\eta_2)-B_{[1]}^{(2)}(\lambda)\cosh(\lambda\,\eta_2)+D_{[1]}^{(2)}(\lambda)$$
$$\times\exp(-\lambda\,\eta_2)=\frac{2}{3}\left(2\lambda-\frac{1}{\eta_2}\right)\exp(-\lambda\,\eta_2),$$

Hence we have

$$A_{[1]}^{(1)}(\lambda)=B_{[1]}^{(1)}(\lambda)=D_{[1]}^{(1)}(\lambda)$$
$$=-\frac{1}{3}\left(2\lambda-\frac{1}{\eta_1}\right)\exp(-2\lambda\,\eta_1),\tag{38}$$

$$C_{[1]}^{(1)}(\lambda)=-\frac{1}{3}\left(2\lambda-\frac{1}{\eta_1}\right),\tag{39}$$

$$A_{[1]}^{(2)}(\lambda)=-B_{[1]}^{(2)}(\lambda)=C_{[1]}^{(2)}(\lambda)$$
$$=\frac{1}{3}\left(2\lambda-\frac{1}{\eta_2}\right)\exp(-2\lambda\,\eta_2),\tag{40}$$

$$D_{[1]}^{(2)}(\lambda)=\frac{1}{3}\left(2\lambda-\frac{1}{\eta_2}\right).\tag{41}$$

Using the results (28), (38), and (40), we have

$$T^{(e)}(\zeta,\eta)=z-\frac{1}{3}\varepsilon_1(\zeta^2+\eta^2)^{1/2}\int_0^{\infty}\left(2\lambda-\frac{1}{\eta_1}\right)$$
$$\times\exp(-2\lambda\,\eta_1+\lambda\,\eta)J_0(\lambda\zeta)d\lambda$$
$$+\frac{1}{3}\varepsilon_2(\zeta^2+\eta^2)^{1/2}\int_0^{\infty}\left(2\lambda-\frac{1}{\eta_2}\right)$$
$$\times\exp(-2\lambda\,\eta_2-\lambda\,\eta)J_0(\lambda\zeta)d\lambda.\tag{42}$$

HYDRODYNAMIC PROBLEM

We seek the solution of the Stokes equation (8) in the form

$$\Psi^{(e)}(\zeta,\eta)=-\frac{1}{2}\gamma^2 U+\bar{\Psi}^{(e)}(\zeta,\eta).$$

An expression for the distortion $\bar{\Psi}^{(e)}(\zeta,\eta)$ of the stationary velocity field near an aggregate, bounded on the symmetry axis ($\zeta=0$) of the flow and vanishing at infinity ($\zeta=\eta=0$), was first obtained in Ref. 8

$$\bar{\Psi}^{(e)}(\zeta,\eta)=\frac{\zeta}{(\zeta^2+\eta^2)^{3/2}}\int_0^{\infty}W(\lambda,\eta)J_1(\lambda\zeta)d\lambda,$$

$$W(\lambda,\eta)=[a(\lambda)+\eta c(\lambda)]\sinh(\lambda\,\eta)$$
$$+[b(\lambda)+\eta d(\lambda)]\cosh(\lambda\,\eta).$$

We seek the functions $X(\lambda)=a(\lambda)$, $b(\lambda)$, $c(\lambda)$, and $d(\lambda)$ of the parameter λ in the form of a power series in ε_1 and ε_2

$$X(\lambda,\varepsilon_1,\varepsilon_2)=X_{[0]}(\lambda)+\varepsilon_1 X_{[1]}^{(1)}(\lambda)+\varepsilon_2 X_{[1]}^{(2)}(\lambda)+\dots\,.$$

Each term in these expansions is determined from the boundary conditions on the surface of an aggregate of touching particles. Evidently, we can write

$$W(\lambda,\eta,\varepsilon_1,\varepsilon_2)=W_{[0]}(\lambda,\eta)+\varepsilon_1 W_{[1]}^{(1)}(\lambda,\eta)$$
$$+\varepsilon_2 W_{[1]}^{(2)}(\lambda,\eta)+\dots,$$

$$W_{[0]}(\lambda,\eta)=(a_{[0]}(\lambda)+\eta c_{[0]}(\lambda))\sinh(\lambda\,\eta)$$
$$+(b_{[0]}(\lambda)+\eta d_{[0]}(\lambda))\cosh(\lambda\,\eta),$$

$$W_{[1]}^{(j)}(\lambda,\eta)=(a_{[1]}^{(j)}(\lambda)+\eta c_{[1]}^{(j)}(\lambda))\sinh(\lambda\,\eta)$$
$$+(b_{[1]}^{(j)}(\lambda)+\eta d_{[1]}^{(j)}(\lambda))\cosh(\lambda\,\eta).$$

An expression for the resultant force exerted by the external medium on an aggregate of touching spheres moving in the direction along the line joining their centers is presented in Ref. 8:

$$F_z(\text{aggregate})=\int_0^{\infty}\lambda b(\lambda)d\lambda.$$

Then, since the small parameters $\varepsilon_1\ll 1$ and $\varepsilon_2\ll 1$ are arbitrary, we have from the condition (16) of uniform motion of an aggregate

$$\int_0^{\infty}\lambda b_{[0]}(\lambda)d\lambda=0,\tag{43}$$

$$\int_0^{\infty}\lambda b_{[1]}^{(1)}(\lambda)d\lambda=\int_0^{\infty}\lambda b_{[1]}^{(2)}(\lambda)d\lambda=0.\tag{44}$$

In what follows we employ the temperature distribution (42) near an aggregate of touching solid spheres. Using Eqs. (A4) and (A5), we find from the boundary conditions (12) and (13)

$$W_{[0]}(\lambda,\eta)\big|_{\eta=\eta_1}=2(\eta_1+\lambda^{-1})\exp(-\lambda\,\eta_1),\tag{45}$$

$$\frac{\partial W_{[0]}(\lambda, \eta)}{\partial \eta}\Bigg|_{\eta=\eta_1} = 2\delta\eta_1\lambda \, \exp(-\lambda\eta_1), \tag{46}$$

$$W_{[0]}(\lambda, \eta)\big|_{\eta=-\eta_2} = 2(\eta_2+\lambda^{-1})\exp(-\lambda\eta_2), \tag{47}$$

$$\frac{\partial W_{[0]}(\lambda, \eta)}{\partial \eta}\Bigg|_{\eta=-\eta_2} = -2\delta\eta_2\lambda \, \exp(-\lambda\eta_2), \tag{48}$$

$$\delta=4\frac{K_{TSL}^{(e)}\dfrac{\nu_0^{(e)}}{T_0^{(e)}}A_T}{U_{[0]}}-1. \tag{49}$$

The dimensional thermophoretic velocity $\mathbf{U}_{[0]}^T = -\mathbf{U}_{[0]}$ is determined in the zeroth approximation by solving the system (45)–(48) of linear inhomogeneous algebraic equations, taking Eq. (43) into account,

$$\mathbf{U}_{[0]}^T = -\frac{4}{1+\delta}K_{TSL}^{(e)}\frac{\nu_0^{(e)}}{T_0^{(e)}}\mathbf{A}_T, \quad \delta=\frac{\displaystyle\int_0^\infty\frac{\lambda}{D}\Phi_1(\lambda)d\lambda}{\displaystyle\int_0^\infty\frac{\lambda^2}{D}\Phi(\lambda)d\lambda}, \tag{50}$$

$$\Phi_1(\lambda)=(\eta_1+\lambda^{-1})\exp(-\lambda\eta_1)\tilde{D}_1$$
$$+(\eta_2+\lambda^{-1})\exp(-\lambda\eta_2)\tilde{D}_3,$$

$$\Phi(\lambda)=\eta_1\exp(-\lambda\eta_1)\tilde{D}_2-\eta_2\exp(-\lambda\eta_2)\tilde{D}_4,$$

$$D=-\sinh^2(\lambda(\eta_1+\eta_2))+(\eta_1+\eta_2)^2\lambda^2,$$

$$\tilde{D}=-2(\eta_1+\lambda^{-1})\exp(-\lambda\eta_1)\tilde{D}_1+2\delta\eta_1\lambda$$
$$\times\exp(-\lambda\eta_1)\tilde{D}_2-2(\eta_2+\lambda^{-1})$$
$$\times\exp(-\lambda\eta_2)\tilde{D}_3-2\delta\eta_2\lambda \, \exp(-\lambda\eta_2)\tilde{D}_4,$$

$$\tilde{D}_1=-\eta_2\lambda\{(\eta_1+\eta_2)\lambda \, \cosh(\lambda\eta_1)+\sinh(\lambda\eta_1)\}$$
$$+\{\eta_1\lambda \, \cosh(\lambda(\eta_1+\eta_2))$$
$$+\sinh(\lambda(\eta_1+\eta_2))\}\sinh(\lambda\eta_2),$$

$$\tilde{D}_2=-(\eta_1+\eta_2)\eta_2\lambda \, \sinh(\lambda\eta_1)$$
$$+\eta_1\sinh(\lambda(\eta_1+\eta_2)\sinh(\lambda\eta_2),$$

$$\tilde{D}_3=-\eta_1\lambda\{(\eta_1+\eta_2)\lambda \, \cosh(\lambda\eta_2)+\sinh(\lambda\eta_2)\}$$
$$+\{\eta_2\lambda \, \cosh(\lambda(\eta_1+\eta_2))$$
$$+\sinh(\lambda(\eta_1+\eta_2))\}\sinh(\lambda\eta_1),$$

$$\tilde{D}_4=(\eta_1+\eta_2)\eta_1\lambda \, \sinh(\lambda\eta_2)$$
$$-\eta_2\sinh(\lambda(n_1+\eta_2))\sinh(\lambda\eta_1).$$

The corrections $\varepsilon_1 U_{[1]}^{(1)}$ and $\varepsilon_2 U_{[1]}^{(2)}$ (due to the difference of the thermal conductivities of the external medium and the particles in the aggregate) are finally found with the aid of the transformations (A1), (A2), (A4)–(A6) from the system of algebraic equations (51)–(58):

$$W_{[1]}^{(1)}(\lambda, \eta)\big|_{\eta=\eta_1} = 2U_{[1]}^{(1)}(\eta_1+\lambda^{-1})\exp(-\eta_1\lambda), \tag{51}$$

$$\frac{\partial W_{[1]}^{(1)}(\lambda, \eta)}{\partial \eta}\Bigg|_{\eta=\eta_1} = -2\eta_1\left\{\frac{1}{3}(1+\delta)+U_{[1]}^{(1)}\right\}\lambda \, \exp(-\eta_1\lambda), \tag{52}$$

$$W_{[1]}^{(1)}(\lambda, \eta)\big|_{\eta=-\eta_2} = 2U_{[1]}^{(1)}(\eta_2+\lambda^{-1})\exp(-\eta_2\lambda), \tag{53}$$

$$\frac{\partial W_{[1]}^{(1)}(\lambda, \eta)}{\partial \eta}\Bigg|_{\eta=-\eta_2} = 2U_{[1]}^{(1)}\eta_2\lambda \, \exp(-\eta_2\lambda)-\frac{2}{3}$$
$$\times(1+\delta)(\eta_1+\eta_2)\left\{2+\frac{\eta_1}{\eta_1+\eta_2}\right.$$
$$\left.-2\eta_1\lambda\right\}\lambda \, \exp(-\alpha_1\lambda), \tag{54}$$

$$W_{[1]}^{(2)}(\lambda, \eta)\big|_{\eta=\eta_1} = 2U_{[1]}^{(2)}(\eta_1+\lambda^{-1})\exp(-\eta_1\lambda), \tag{55}$$

$$\frac{\partial W_{[1]}^{(2)}(\lambda, \eta)}{\partial \eta}\Bigg|_{\eta=\eta_1}$$
$$= -2U_{[1]}^{(2)}\eta_1\lambda \, \exp(-\eta_1\lambda)+\frac{2}{3}$$
$$\times(1+\delta)(\eta_1+\eta_2)\left\{2+\frac{\eta_2}{\eta_1+\eta_2}-2\eta_2\lambda\right\}\lambda$$
$$\times\exp(-\alpha_2\lambda), \tag{56}$$

$$W_{[1]}^{(2)}(\lambda, \eta)\big|_{\eta=-\eta_2} = 2U_{[1]}^{(2)}(\eta_2+\lambda^{-1})\exp(-\eta_2\lambda), \tag{57}$$

$$\frac{\partial W_{[1]}^{(2)}(\lambda, \eta)}{\partial \eta}\Bigg|_{\eta=-\eta_2} = -2\eta_2\left\{\frac{1}{3}(1+\delta)+U_{[1]}^{(2)}\right\}\lambda$$
$$\times\exp(-\eta_2\lambda),$$

$$\alpha_1=2\eta_1+\eta_2, \quad \alpha_2=\eta_1+2\eta_2. \tag{58}$$

The additional conditions (44) make it possible to write

$$U_{[1]}^{(j)}=(-1)^j\frac{1+\delta}{3}\beta^j, \quad \beta^j=\frac{\displaystyle\int_0^\infty\frac{\lambda^2}{D}\Omega_j(\lambda)d\lambda}{\displaystyle\int_0^\infty\frac{\lambda}{D}\Omega(\lambda)d\lambda}, \tag{59}$$

$$\Omega(\lambda)=(\eta_1+\lambda^{-1})\exp(-\lambda\eta_1)\tilde{D}_1+\eta_1\lambda$$
$$\times\exp(-\lambda\eta_1)\tilde{D}_2+(\eta_2+\lambda^{-1})$$
$$\times\exp(-\lambda\eta_2)\tilde{D}_3-\eta_2\lambda \, \exp(-\lambda\eta_2)\tilde{D}_4,$$

$$\Omega_1(\lambda)=\eta_1\exp(-\lambda\eta_1)\tilde{D}_2+(\eta_1+\eta_2)$$
$$\times\left(2+\frac{\eta_1}{\eta_1+\eta_2}-2\eta_1\lambda\right)\exp(-\alpha_1\lambda)\tilde{D}_4,$$

$$\Omega_2(\lambda)=(\eta_1+\eta_2)\left(2+\frac{\eta_2}{\eta_1+\eta_2}-2\eta_2\lambda\right)$$
$$\times\exp(-\alpha_2\lambda)\tilde{D}_2+\eta_2\exp(-\lambda\eta_2)\tilde{D}_4.$$

TABLE I.

R_1/R_2	δ	$\beta^{(1)}$	$\beta^{(2)}$
1	4.69309	7.61846×10^{-2}	-7.61846×10^{-2}
2	4.77724	1.31308×10^{-1}	-2.54592×10^{-2}
3	4.85906	1.49923×10^{-1}	-1.08006×10^{-2}
4	4.90815	1.57435×10^{-1}	-5.50003×10^{-3}
5	4.93750	1.61028×10^{-1}	-3.16567×10^{-3}
6	4.95576	1.62964×10^{-1}	-1.98498×10^{-3}
7	4.96762	1.64103×10^{-1}	-1.32536×10^{-3}
8	4.97563	1.64816×10^{-1}	-9.28434×10^{-4}
9	4.98121	1.65287×10^{-1}	-6.75435×10^{-4}
10	4.98522	1.65610×10^{-1}	-5.06618×10^{-4}
20	4.99735	1.66499×10^{-1}	-7.22619×10^{-5}
30	4.99911	1.66613×10^{-1}	-2.24074×10^{-5}
40	4.99960	1.66643×10^{-1}	-9.67350×10^{-6}

TABLE II.

R_2/R_1	δ	$\beta^{(1)}$	$\beta^{(2)}$
1	4.69309	7.61846×10^{-2}	-7.61846×10^{-2}
2	4.77724	2.54592×10^{-2}	-1.31308×10^{-1}
3	4.85906	1.08006×10^{-2}	-1.49923×10^{-1}
4	4.90815	5.50003×10^{-3}	-1.57435×10^{-1}
5	4.93750	3.16567×10^{-3}	-1.61028×10^{-1}
6	4.95576	1.98498×10^{-3}	-1.62964×10^{-1}
7	4.96762	1.32536×10^{-3}	-1.64103×10^{-1}
8	4.97563	9.28434×10^{-4}	-1.64816×10^{-1}
9	4.98121	6.75435×10^{-4}	-1.65287×10^{-1}
10	4.98522	5.06618×10^{-4}	-1.65610×10^{-1}
20	4.99735	7.22619×10^{-5}	-1.66499×10^{-1}
30	4.99911	2.24074×10^{-5}	-1.66613×10^{-1}
40	4.99960	9.67350×10^{-6}	-1.66643×10^{-1}

ANALYSIS OF RESULTS

The dimensional thermophoretic velocity of an aggregate in an approximation linear in the small parameters can be written as follows taking expressions (50) and (59) into account:

$$\mathbf{U}^T = -4K_{TSL}^{(e)} \frac{\nu_0^{(e)}}{T_0^{(e)}} \mathbf{A}_T \left\{ \frac{1}{1+\delta} - \frac{1}{3}\varepsilon_1\beta^{(1)} + \frac{1}{3}\varepsilon_2\beta^{(2)} \right\}. \tag{60}$$

It is of interest to check in the course of the numerical analysis the agreement between Eq. (60) in the limiting cases of thermophoresis of an aggregate of touching solid nonvolatile spheres ($R_1 \gg R_2$ or $R_2 \gg R_1$) with the result obtained earlier for a single particle[1]

$$\mathbf{U}^T = -\frac{2\varkappa_\alpha}{1+2\varkappa_\alpha} K_{TSL}^{(e)} \frac{\nu_0^{(e)}}{T_0^{(e)}} \mathbf{A}_T. \tag{61}$$

Let us expand the expression $2\varkappa_\alpha/(1+2\varkappa_\alpha)$ in a power series in ε_α

$$\frac{2\varkappa_\alpha}{1+2\varkappa_\alpha} = \frac{2}{3}\left(1 - \frac{1}{3}\varepsilon_\alpha - \frac{2}{9}\varepsilon_\alpha^2 - \dots \right).$$

Then, comparing the right-hand sides of Eqs. (60) and (61) for the limiting cases $R_1 \gg R_2$ or $R_1 \ll R_2$ shows that the following equalities are satisfied:

$$\frac{4}{1+\delta} = \frac{2}{3} \text{ (whence } \delta = 5) \tag{62}$$

in the zeroth approximation and

$$\varepsilon_1\beta^{(1)} - \varepsilon_2\beta^{(2)} = \frac{1}{6}\begin{cases} \varepsilon_1 & \text{for } R_1 \gg R_2, \\ \varepsilon_2 & \text{for } R_2 \gg R_1 \end{cases} \tag{63}$$

in the first approximation.

The limiting cases (62) and (63) are confirmed by the results of a numerical analysis (presented in Tables I and II).

It should be noted that in the zeroth approximation in the small parameters (corresponding to the thermal conductivity of the external medium being the same as that of the touching nonvolatile solid spheres) the thermophoretic velocity of an aggregate remains unchanged when the particles are made to change places ($R_1 \to R_2$, $R_2 \to R_1$). In the case $\varepsilon_1 = \varepsilon_2$

$= 0$ the velocity of an aggregate is higher than that of any single particle and the effect of the form of the aggregation is greatest when equal spheres touch (it equals at least 5%).

Evidently, the theory of thermophoresis constructed on the basis of a hydrodynamic analysis is also valid for an aggregate of touching solid hydrosol particles and high-viscosity pure drops. This case is most important for practical applications.

APPENDIX

$$\int_0^\infty \exp(-\lambda\eta)J_0(\lambda\zeta)d\lambda = (\zeta^2 + \eta^2)^{-1/2}, \tag{A1}$$

$$\int_0^\infty \exp(-\lambda\eta)J_0(\lambda\zeta)\lambda d\lambda = \eta(\zeta^2 + \eta^2)^{-3/2}, \tag{A2}$$

$$\int_0^\infty \exp(-\lambda\eta)J_0(\lambda\zeta)\lambda^2 d\lambda = 3\eta^2(\zeta^2 + \eta^2)^{-5/2}$$
$$- (\zeta^2 + \eta^2)^{-3/2}, \tag{A3}$$

$$\int_0^\infty \exp(-\lambda\eta)(\eta + \lambda^{-1})J_1(\lambda\zeta)d\lambda = \frac{\zeta}{(\zeta^2 + \eta^2)^{1/2}}, \tag{A4}$$

$$\int_0^\infty \exp(-\lambda\eta)J_1(\lambda\zeta)\lambda d\lambda = \frac{\zeta}{(\zeta^2 + \eta^2)^{3/2}}, \tag{A5}$$

$$\int_0^\infty \exp(-\lambda\eta)J_1(\lambda\zeta)\lambda^2 d\lambda = 3\frac{\zeta\eta}{(\zeta^2 + \eta^2)^{5/2}}. \tag{A6}$$

[1] Yu. I. Yalamov and V. S. Galoyan, *Dynamics of Droplets in Inhomogeneous Viscous Media* [in Russian], Luĭs, Erevan, 1985, 208 pp.

[2] Yu. I. Yalamov, M. N. Gaĭdukov, and A. P. Melekhov, Dokl. Akad. Nauk SSSR **287**, 337 (1986) [Sov. Phys. Dokl. **31**, 251 (1986)].

[3] Yu. I. Yalamov, E. R. Shchukin, and S. I. Grashchenkov, *Thermophoresis and Diffusiophoresis of Two Aerosol Particles with Allowance for Internal Heat Sources* [in Russian], MOPI, Moscow, 1989, 118 pp.; Deposited in VINITI, No. 7212-V89.

[4] Yu. I. Yalamov, M. N. Gaĭdukov, and V. V. Levin, *Thermophoresis of Two Particles at Low Reynolds Numbers* [in Russian], MPU, Moscow, 1993, 9 pp.; Deposited in VINITI, No. 1029-V93.

[5] L. D. Reed and F. A. Morrison, J. Aerosol Sci. **6**, 349 (1975).

[6] M. Stimson and G. B. Jeffrey, Proc. R. Soc. London, Ser. A **111**, 110 (1926).

[7] O. O. Olaru, Vestn. Mosk. Univ. Ser. 1, Matem. Mekh., No. 3, 73 (1992).

[8] M. D. A. Cooley and M. E. O'Neill, Proc. Cambridge Philos. Soc. **66**, 407 (1969).

Determination of nonlinear aerodynamic characteristics from trajectory data of an object: modification of the method for complicated cases

A. B. Podlaskin

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted March 26, 1997)
Zh. Tekh. Fiz. **68**, 32–36 (June 1998)

The method of calculating the nonlinear aerodynamic characteristics of objects from trajectory data, based on the differential correction principle, is well known [N. P. Mende, FTI Preprint No. 1326, A. F. Ioffe Physicotechnical Institute, Leningrad (1989); G. T. Chapman and D. B. Kirk, AIAA J. **8**, 753 (1970)]. A modification of this method is proposed here, in which the solution is to be obtained in the form of a spline. This new approach, which has been tested on model problems, can provide a more reliable guarantee of adequacy of the solutions and an improvement in accuracy in cases where the functional relations sought have a complicated form. © *1998 American Institute of Physics.* [S1063-7842(98)00706-5]

INTRODUCTION

One problem of experiments on ballistic apparatus is to determine the aerodynamic characteristics, i.e., the forces and torques acting on a body in free flight. In this context, researchers have at their disposal trajectory data—an ordered set of readings of the coordinates of the object at fixed instants of time. These coordinates and the desired functional dependences of the aerodynamic forces and torques are interrelated by a system of nonlinear differential equations,[1] which are not analytically integrable. Thus the problem takes the general form

$$Az(x) = u(x), \tag{1}$$

where $u(x) = x$ is the vector of measured coordinates, z is the vector of aerodynamic coefficients (thus $z(x)$ are the desired functional relations), and A is the integral operator of the problem. This ill-posed problem is prescribed by an operator equation of the first kind.[2]

Only the inverse operator A^{-1} can be written down analytically. To solve this problem, the trial-and-error method has traditionally been used, i.e., for a series of approximations of the unknown functions the direct problem is solved and the deviation of the calculated trajectory from the experimental trajectory is calculated. The first approximation is chosen on the basis of an *a priori* expert assessment. As the algorithm for constructing the minimizing sequence the differential correction method has proved to be successful[3] (it is also known as the Gauss–Newton method). The class of functions among which the stable solution is to be sought (the correctness class of the problem) can be restricted in different ways. The only condition is that the functional subspace under consideration be compact.[2] Convergence of the method in the class of power-law polynomials containing a finite number of terms of even or odd powers was confirmed in Ref. 1, where examples are given of successful determination of aerodynamic coefficients on ballistic experiments with sharp cones. The practical importance of a statistical assessment of the results was also demonstrated[1] on the basis of the variance matrix of the regression coefficients of the unknown functions. The complexity of the question of adequacy of the solution found by the trial-and-error method is well known, since the efficiency of the method in a given case is determined by the errors of the experimental data.

STATEMENT OF PROBLEM

Among the objects of ballistic studies, a few are encountered that possess a quite complicated form of dependence $z(x)$ of the aerodynamic coefficients on the generalized coordinate. Although the aerodynamic coefficients, as a rule, are described by smooth functions, they can, for example, be substantially nonmonotonic. Complexities of this kind usually show up when the range of variation of the independent variable is increased. In such cases a model description of the motion of the object using power series segments as the regression form can be assumed to be inefficient. The more complicated the form of the unknown function, the further out must its approximating series be carried. As the number of terms of the series increases, the conditionality of the least-squares matrices deteriorates. This is a well-known fact in regression analysis.[4] Computational algorithms exist (e.g., Ref. 5) which allow one to solve the corresponding underdetermined systems of equations. However, other shortcomings of higher-degree polynomials do not allow one to go this route in the case of a complicated form of $z(x)$. Experience shows that the confidence interval for the sum of the series segment increases severalfold at the end of the approximation interval due to errors in the coefficients of the higher powers of the independent variable. In addition, since a power-law basis is not orthogonal, introducing higher powers in the approximation of the unknown function on large intervals alters the values of the regression coefficients of the lower powers. The presence of a large number of terms can lead to the appearance of nonphysical oscillations. All this motivates us to look for other forms of representation of the unknown functions to use in the trial-and-error search for the solution of the ill-posed problem under consideration here.
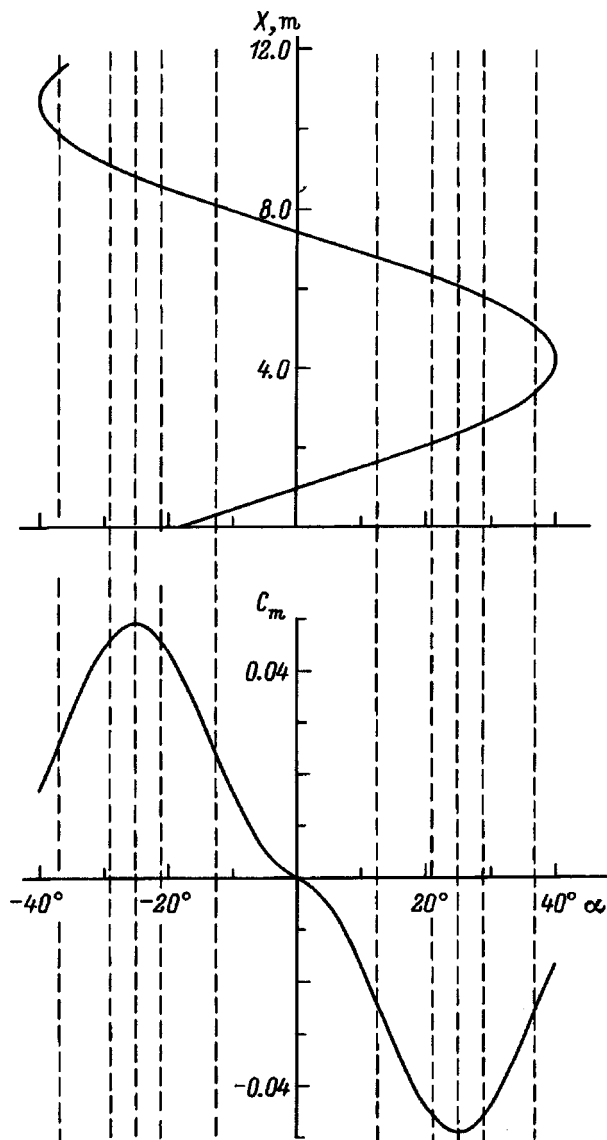
FIG. 1. Model functional dependence of the static aerodynamic moment coefficient (lower graph) and example of the corresponding oscillations of the object as a function of the attack angle (upper graph). Dashed lines—partition of the $\{\alpha\}$ axis into intervals.

One variant that has been proposed is to partition the range of variation of the independent variable x and construct a regression of power series segments "matched" at the boundaries of the intervals in respect to the value of the function and the first derivative. In this case, within the limits of comparatively large intervals in x the unknown function will be described by substantially shorter polynomials.

CONSTRUCTION OF THE METHOD

In light of the above considerations regarding the difficulty of elucidating the adequacy of the desired model and the effect of hard-to-control experimental errors at the processing stage of the algorithm, model problems were examined, i.e., trajectory data calculated on the basis of prescribed aerodynamic characteristics of an "object." Consider Fig. 1.

As a particular case of $z(x)$, the lower graph plots the static aerodynamic moment coefficient as a function of the angle of attack, $C_m(\alpha)$. This function

$$C_m(\alpha) = \text{sign}(\alpha)[C_{\max} \sin^3(|\alpha| + \alpha_0) + C_0],$$

which I used as my model example, cannot in general be described by a power series with a finite number of terms. In regard to the problem under consideration this means that the Chapman–Kirk algorithm[3] does not allow us to seek out a solution that would adequately describe the given moment function. Increasing the degree of the polynomial to 13 does not achieve the desired model adequacy even though the elements of the matrix of normal equations at this point have already reached the limit of machine accuracy. If we consider segments of its domain of definition, e.g., as shown by the dashed lines, then it is obvious that on these intervals the function can be well fitted by lower-degree polynomials. The upper curve is the "trajectory" of the oscillations of a body in flight, i.e., the dependence of the attack angle on the longitudinal coordinate along the trajectory. From the matching condition for the segments of the unknown function using a quadratic polynomial as the regression function to fit the functional dependence $C_m(\alpha)$ on the domain $|\alpha| \in [\alpha_i, \alpha_{i+1}]$ we have

$$C_m(\alpha) = \text{sign}(\alpha) C_m(\alpha_i) + [dC_m(\alpha_i)/d\alpha]$$
$$\times (\alpha - \alpha_i) \text{sign}(\alpha) k_{i+1}(\alpha - \alpha_i)^2 \quad (2)$$

(the sign function was introduced in accordance with the physical requirement that the static aerodynamic moment be odd).

Thus, the adopted regression form, possessing a piecewise-continuous second derivative, corresponds to a first-order spline.

Many of the advantages of the previously developed technique for finding the aerodynamic characteristics determined the factors of continuity. Only, the use of the sign function sign (α) corresponds to the use of polynomials of even or odd degree in the Chapman–Kirk algorithm, which is yet another example in the trial-and-error method of the necessity of invoking *a priori* information about the nature of the unknown function $z(x)$. As in the case of an all-at-once polynomial approximation of the aerodynamic characteristics, the unknown number of degrees of freedom can be achieved by coprocessing the data of several experiments with one model. In this case the number of new nodes of the approximation increases more rapidly than the number of unknowns (the regression coefficients remain general, only the initial conditions of the launch, which are treated as unknowns, are added on). Moreover, the coprocessing of the data of several launches of one object with different amplitudes of oscillations in the attack angle allows one to make use of the nonisochronicity of these oscillations, which gives more information about the nonlinear characteristics of the moment than can be had from an analysis of the form of the oscillations.

As before, the main logical thrust of the algorithm consists in constructing a sequence of approximations of the desired solution which minimize the deviation of the calcu-

lated trajectory from the experimental trajectory in a quadratic metric. To linearize the target function, i.e., the residual sum of squares of differences of the attack angle α, we expand it in a multidimensional Taylor series in the desired parameters in the vicinity of the solution, where this series is truncated at the level of the linear terms. In such a representation the deviation is written in terms of the derivatives of α with respect to the regression coefficients of the unknown function and the initial conditions. The values of the derivatives are found from the equations of sensitivity, which are obtained by differentiating the equation of motion with respect to all the unknowns and are integrated together with the equation of motion. This approach allows one to compose the system of normal equations of the method of least squares—linear algebraic equations based on the equations of motion and sensitivity for all points of the scheme. However, the solution of such a system of equations, which only approximate the given nonlinear problem, gives only a set of corrections to the coefficients. Therefore, the search for a solution reaches its goal after several iterations. Calculation of the variance matrix[6] allows one in the course of the differential correction to estimate the error of the found approximations of the coefficients and assess its significance.

The new form of representation of the unknown function necessitated certain modifications in the algorithm. In the trial-and-error search for a solution of the operator equation (1), when we are solving the direct problem, i.e., integrating the equations of motion, at each step the current value of α is monitored with the aim of determining to which of the designated intervals it belongs. Depending on the interval to which the argument belongs, the values of $C_m(\alpha)$ are calculated according to the formulas of the respective interval in the procedures for calculating these values. Since on different intervals of $\{\alpha\}$ the regression coefficients affect the target function in a different way: some of them directly, some via their effect on the preceding segments [via formula (2)], the calculation of the coefficients of the equations of sensitivity also depends on which interval α belongs to. However, from the derivatives calculated in this way a single matrix of normal equations of the method of least squares is constructed, which of course corresponds to a single equation prescribing the motion of the object, and all coefficients used to describe the unknown function turn out to be correlated.

ANALYSIS OF THE SOLUTION

By varying the number and positions of the interval boundaries α_i, it is possible to achieve acceptable results in determining $C_m(\alpha)$. Acceptability is determined, on the one hand, by the errors of the found regression coefficients with allowance for their correlation (when the relative error of the coefficient of the square of the independent variable exceeds 100% for a prescribed confidence level, rejection of such an insignificant coefficient is indicated and the moment on this interval is taken to be linear). On the other hand, the adequacy of the found solution for the model problem is easily monitored since the exact form of the desired function is known. Experience shows that the problem does not have a unique solution in the considered class of functions and for
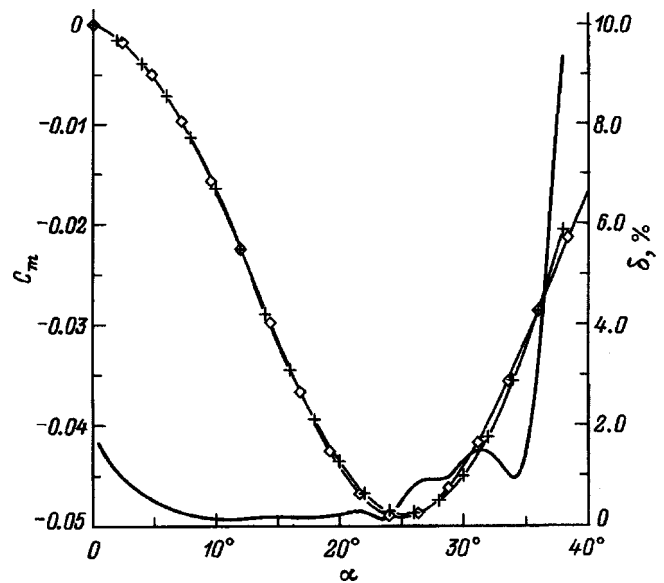

FIG. 2. Spline-reconstructed dependence (crosses) of the static aerodynamic moment coefficient from Fig. 1. Original function—squares. Solid line—estimate of the residual error of the δ approximation.

an unpropitious choice of the partition of the domain $\{\alpha\}$ it is possible to find a spurious solution. In practice it has been found that good results in terms of efficiency of the algorithm are had by assigning the boundaries at the extrema and inflection points of the unknown function. For the indicated example (in Fig. 1 the moment is prescribed as the cube of the sine of the attack angle), which has two inflection points on the positive side and on the negative side in the range $|\alpha| \leqslant 40°$ ($\alpha \cong \pm 13.25°$ and $\alpha \cong \pm 36.75°$) and one extremum ($\alpha = \pm 25°$), acceptable results are achieved by starting not with four intervals in the absolute value of α, but with six: it was necessary to divide up the regions of large curvature symmetrically near an extremum (increasing the number of intervals above the minimum necessary serves no purpose). Trajectory data of four ''launches'' were used, with amplitudes of oscillation in the attack angle equal to 7, 14, 27, and 38°.

The plan of the ''experiment'' called for 34 points. Approximation with quadratic polynomials on six intervals had 14 degrees of freedom. The accumulation of error of the found nonlinear regression coefficients k_i went from 1 ($|\alpha| \leqslant \alpha_1$) to 65% ($|\alpha| \geqslant \alpha_5$). Here the achieved accuracy of determination of the static moment was within 9% (Fig. 2). The error was calculated from the variance matrix of the regression coefficients with their covariances taken into account.

Finding the solution of an ill-posed problem by the trial-and-error method[2] is equivalent to minimizing the distance ρ (in the quadratic metric) from the intermediate approximate solution z_δ to the exact solution z_T. The result of this process, as a rule, looks like this: the local deviations have opposite signs at the boundaries of the approximation region and in its middle while its integral $\rho(z_\delta, z_T)$ over the entire region tends to zero. In the given case on each interval $\alpha_i < |\alpha| \leqslant \alpha_{i+1}$ the higher (one) coefficient of the approximating parabolic segment is sought independently. This finds
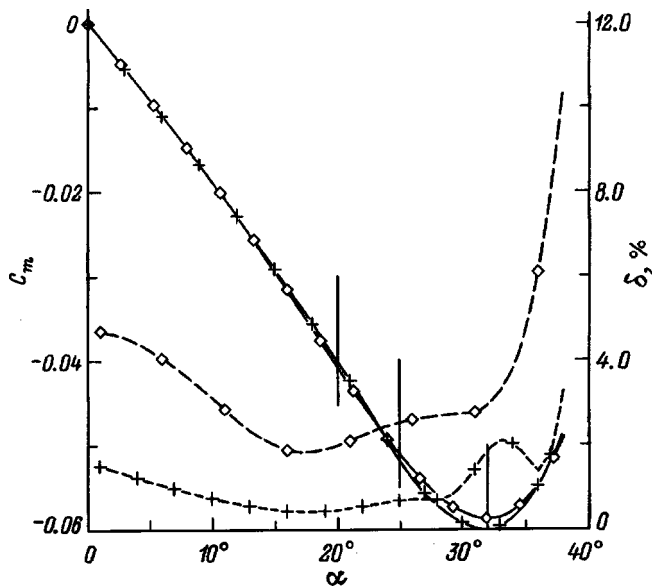
FIG. 3. Comparison of the results of all-at-once (squares) and piecewise (crosses) approximation in the reconstruction of the nonlinear static aerodynamic moment of a sharp cone. Solid lines—values of the moment coefficient, dashed lines—estimate of the relative error δ of the corresponding approximations.

expression in the complex character of the dependence of the error function of the found moment coefficient on the attack angle: the error reflects the character of the local deviation of z_δ from z_T. Consideration of a model problem (where the exact solution is known) makes it possible visually to trace out the variation of ρ from the relative location of the curves of the prescribed and found functions $C_m(\alpha)$.

The case of four launches of a sharp cone with a total vertex angle of 30°, considered in Ref. 1, was also analyzed. In this case it is possible to compare the results for an all-at-once approximation of the desired moment coefficient using the Chapman–Kirk algorithm (odd fifth-degree polynomial) and for a piecewise-polynomial approximation. However, use of the available experimental data had to be rejected by virtue of the sensitivity of the method to the systematic and anomalous errors contained in them for a relatively small sample space and it was necessary to resort to modeled input data. Model trajectories, calculated from the characteristics taken from Ref. 1, was "noised up" by adding normally distributed random error with a standard deviation of 0.5°, which corresponds to the error of the actual measurements. For this version of the function $C_m(\alpha)$ it turned out to be sufficient to use four intervals (indicated by vertical line segments in Fig. 3). Comparison of the obtained curves reveals good agreement (the same solution is found). Estimates of the residual mean-square deviations of the attack angle α of the trajectories calculated by the two different methods ($\sigma_\alpha \cong 0.51°$ for the all-at-once approximation and $\sigma_\alpha \cong 0.13°$ for the new method) favor the piecewise polynomial approximation. Construction of the variance matrix of regression coefficients in both cases allows us to estimate the total error of the moment within the given range $\{\alpha\}$. Values of the confidence interval of the static moment for a confidence level of 95% are shown in Fig. 3. It can be seen that the piecewise

description of the moment function allows it to be found with less error. However, finding it is fraught with greater difficulties than finding the all-at-once polynomial approximation. There are additional problems associated with choice of the boundaries of the segments into which to divide the range of variation of $\{\alpha\}$. Overcoming these difficulties will require time, but probably with accumulation of experience it will be possible to shorten the amount of time needed.

CONCLUSION

Analyzing the difficulties arising during processing of the algorithm for determining the aerodynamic characteristics from trajectory measurements using splines to describe the unknown functions, we can form the following picture regarding application of the method to actual experimental data. It is unrealistic to choose the initial approximation in the form proposed using the given approach if all that is available are raw data. Therefore it is advisable to begin processing the experimental results with the help of the Chapman–Kirk algorithm, for example. Having obtained a representation of the unknown functions, it is possible to shift over to the new algorithm and carry out a comparative analysis of the solutions found by the two approaches as to their adequacy (to the extent that it is possible to judge, generally) and their errors. Features of the unknown functions revealed in the course of fitting using the Chapman–Kirk algorithm can help construct initial spline approximations. Here an additional optimization factor arises— deciding on the number of intervals in the independent variable and the positions of their boundaries. As has already been noted, for the observations just dealt with the most successful locations, in the sense of convergence, for the boundaries are at the extrema and inflection points of the unknown functions. When working with experimental data, the search for such points must be carried out one after the other. A future check of the fruitfulness of the proposed approach in this case will be possible by obtaining qualitative experimental data.

[1] N. P. Mende, FTI Preprint No. 1326 [in Russian], A. F. Ioffe Physico-technical Institute, Leningrad (1989), 44 pp.

[2] A. N. Tikhonov and V. Ya. Arsenin, *Methods of Solving Ill-Posed Problems* [in Russian] (Nauka, Moscow, 1979), 228 pp.

[3] G. T. Chapman and D. B. Kirk, AIAA Pap. **8**, 753 (1970).

[4] G. A. Seber, *Linear Regression Analysis* [Wiley, New York, 1977; Mir, Moscow, 1980, 456 pp.].

[5] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in Pascal: The Art of Scientific Programming* (Cambridge University Press, New York, 1989), 759 pp.

[6] S. M. Ermakov and A. A. Zhiglyavskiǐ, *Mathematical Theory of the Optimal Experiment: A Textbook* [in Russian] (Nauka, Moscow, 1987), 320 pp.

Translated by Paul F. Schippnick

Steady-state leader breakdown. Nitrogen atmosphere

A. V. Ivanovskiĭ

Physical arguments about the possible mechanism of formation of a leader channel are presented. A mathematical model describing steady-state leader breakdown is constructed. An algorithm for determining the propagation velocity, dimensions, and electric field in the streamer zone is developed. A numerical simulation of the channel formation stage in the plasma of a streamer zone a nitrogen atmosphere is performed. The dependence of the leader velocity on the potential is obtained. The tentative model proposed here can be used to describe the leader breakdown of a long gap at high positive potentials. © *1998 American Institute of Physics.* [S1063-7842(98)00806-X]

INTRODUCTION

The problem of steady-state breakdown of an extended gas-filled gap with a prescribed potential U_0 on one of the electrodes is considered. It is well known (see, e.g., Ref. 1) that such a discharge has a complicated structure. The potential U_0 is transmitted along a highly conductive leader channel at a temperature high enough for thermal ionization to take place. The channel is formed in the poorly conducting streamer zone, whose dimensions are quite large $L \geqslant 1$ m. There are a number of conceptions about the mechanism of formation of the leader channel. According to one of them, it is formed as a result of heating of a solitary streamer.[2] Hypotheses have been developed about the formation of the channel against the background of a huge number of continuously forming streamers.[3] We will proceed from the first.

If we neglect the potential drop in a highly conductive channel, then for gaps with large dimensions of the streamer zone it is natural to consider the steady-state breakdown regime. In this case the problem of the formation of a leader channel in a streamer discharge plasma is one-dimensional—all quantities depend only on the transverse coordinate r. The z dependence is determined parametrically in terms of the retarded time $\tau = t - z/v_x$. For a given medium the described breakdown regime is completely determined by one parameter—the potential U_0, on which depend the propagation velocity v_z and the parameters of the streamer zone and channel. Determination of the dependence on U_0 of the velocity v_z and of the dimensions L and electric field E_0 in the streamer zone is the aim of the present work. Here it is assumed that v_z is much greater than the electron drift velocity, i.e., v_z is determined by the charge transfer rate along the channel. An important aspect of these studies is the model of the streamer zone of the leader.

MODEL OF THE STREAMER ZONE

The situation here is unclear.[1] Reference 4 calls attention to the fact that broadening of the current-carrying region of the cylindrical ionization wave plays a substantial role in the propagation of a streamer. In Ref. 5 a regime of streamer propagation is constructed in which the conductance per unit length, $G = 2\pi \int_0^\infty \sigma r \, dr$, is formed by a cylindrical ionization wave with nearly constant propagation velocity v_r. The conductivity of the streamer (i.e., the electron concentration) depends on the coordinate r as $\sigma \approx v_z^2 / v_r r$ $(G \approx v_z^2 \tau)$. It has been shown that such a regime can support the propagation of streamers many meters in length. Here the electric field varies weakly along the length of the streamer and is equal to ~ 10 kV/m. This picture permits an explanation of the experimentally observed[3] linear dependence of the streamer velocity on the rate of growth of the potential for $dU/dt > 400$ kV/μs ($v_z > 4 \times 10^7$ cm/s). We will proceed from the concept of a streamer given in Ref. 5.

It is assumed that the leader channel is formed as a result of the development of an ionization–overheating instability[1] near the axis, where the ionization is maximum. For moderate rates of current growth it is possible to restrict the discussion to processes at atmospheric pressure (without shock waves). The description of channel formation is complicated in that the dimensions of the region in which the instability develops are uncertain. The scale of this region can depend on the initial conditions, e.g., the size of the streamer tip. We will neglect this scale $\leqslant 10^{-3}$ cm (Ref. 6), assuming that the channel parameters are determined by transport processes in the streamer plasma.

In a quasineutral streamer plasma diffusion is ambipolar. However, heating of the electron gas is governed by the electron diffusion coefficient. On this basis, we neglect diffusion of charged particles in comparison with the electronic thermal conductivity. Note that by virtue of the low electron concentration at the stage of channel formation the electronic thermal conductivity does not contribute significantly to the overall balance of energy. Its influence is manifested in an averaged way, through the removal of energetic electrons capable of ionizing the gas, into the colder regions.

Below, on the basis of the concepts expounded above we construct a calculational model of channel formation in a streamer discharge plasma. Here the main mechanism of cold gas entrainment into the channel is the electronic thermal conductivity.

BASIC EQUATIONS

We assume that the plasma consists of molecular nitrogen, the corresponding ions and electrons with concentrations N, n_+, n_e. This assumption is justified up to temperatures of the medium $T \leqslant 4000$ K, where it is possible to neglect dissociation.[1] The plasma is electrically neutral $n_e \cong n_+$ and weakly ionized $10^{-3} \times N \geqslant n_e$, i.e., it is possible to neglect Coulomb collisions.[1] Neglecting diffusive drift of the charged particles, the equation of continuity for the electron component takes the form

$$\frac{\partial n_e}{\partial \tau} + \frac{1}{r}\frac{\partial}{\partial r}(rv_0 n_e) = k_{\mathrm{ion}} N n_e - k_r n_e^2 + S_{\mathrm{str}}, \tag{1}$$

where v_0 is the transverse hydrodynamic velocity.

The source term $S_{\mathrm{str}}(v_z, v_r, E_0)$ describes the creation of electron–ion pairs at the front of the cylindrical ionization wave forming the streamer discharge. To get the streamer discharge to propagate, it is necessary to set[5]

$$S_{\mathrm{str}} = \frac{C_0}{2\pi} \frac{1 v_z v_z}{e k_e(E_0) v_r (r + r_0)} \delta(\tau - r/v_r), \tag{2}$$

where $k_e(E_0) = W_e/E_0$ is the electron mobility, $W_e(E_0)$ is the drift velocity, the capacitance of the discharge per unit length is $C_0 = 1/2 \mathrm{Ln}(2v_z/v_r)$, and the quantity r_0 has been introduced to eliminate the singularity in the limit $r \to 0$.

For a given electron energy distribution function $\varepsilon - f(\varepsilon)$ the ionization k_{ion} and recombination k_r rate constants are given by

$$k_{\mathrm{ion}} = \sqrt{\frac{2}{m_e}} \int_I^\infty \varepsilon q_{\mathrm{ion}}(\varepsilon) f(\varepsilon) d\varepsilon,$$

$$k_r = \sqrt{\frac{2}{m_e}} \int_0^\infty \varepsilon q_r(\varepsilon) f(\varepsilon) d\varepsilon, \quad \int_0^\infty \sqrt{\varepsilon} f(\varepsilon) d\varepsilon, \tag{3}$$

where q_{ion} and q_r the ionization and recombination cross sections (I is the ionization potential); it is assumed that ionization occurs from the ground state.

To determine N and v_0 it is necessary to solve the hydrodynamic equations which at constant atmospheric pressure P_0 take the form[7]

$$\frac{\partial N}{\partial \tau} + \frac{1}{r}\frac{\partial}{\partial r}(rNv_0) = 0, \quad P_0 = NkT,$$

$$\frac{\partial}{\partial \tau}(N\varepsilon) + \frac{1}{r}\frac{\partial}{\partial r}[rv_0(N\varepsilon + P_0)] - \frac{1}{r}\frac{\partial}{\partial r}(rJ) = n_e S_e, \tag{4}$$

where we have neglected the kinetic energy in comparison with the internal energy ε, J is the energy flux, S_e is the energy transferred by an electron to the medium per second in elastic and inelastic collisions.

By virtue of their small mobility the contribution of the ions to heating of the medium can be neglected. The internal energy ε is equal to the sum of the kinetic energy $\varepsilon_k = (3/2)kT$, the energy of rotational degrees of freedom, which under conditions of thermodynamic equilibrium is equal to $\varepsilon_R = kT$, and the vibrational energy ε_V

$$\varepsilon = \frac{5}{2}kT + \varepsilon_V. \tag{5}$$

In writing Eq. (5) we have neglected the energy of the electron-excited particles, assuming that it relaxes quite rapidly into thermal energy.

By virtue of retarded deactivation of the vibrational degrees of freedom, to determine ε_V it is necessary to solve the equation[7]

$$\frac{\partial(N\varepsilon_V)}{\partial \tau} + \frac{1}{r}\frac{\partial}{\partial r}(rv_0 N\varepsilon_V) - \frac{1}{r}\frac{\partial}{\partial r}\left(r\chi N \frac{\partial \varepsilon_V}{\partial r}\right)$$

$$= n_e N S_V + N \frac{\varepsilon_V^p - \varepsilon_V}{\tau_V}, \tag{6}$$

where S_V is the energy lost by an electron to excitation of vibrations of a molecule, τ_V is the vibrational relaxation time, $\varepsilon_V^p = h\nu/[\exp(h\nu/kT) - 1]$ is the vibrational energy under conditions of thermodynamic equilibrium ($h\nu = 0.291$ eV is the energy of a vibrational quantum), $\chi = \lambda/(Nc_p)$ is the thermal diffusivity [$c_p = (7/2)k$ is the specific heat at constant pressure], and λ is the thermal conductivity.

Transforming Eqs. (4)–(6), we obtain a system of hydrodynamic equations in the temperature T, concentration N, and mass velocity v_0

$$Nc_p\left(\frac{\partial T}{\partial \tau} + v_0 \frac{\partial T}{\partial r}\right) = \frac{1}{r}\frac{\partial}{\partial r}\left(r\lambda \frac{\partial T}{\partial r}\right) + n_e(S_e - NS_V)$$

$$- N\frac{\varepsilon_V^p - \varepsilon_V}{\tau_V},$$

$$\frac{c_p P_0}{k}\frac{1}{r}\frac{\partial}{\partial r}(rv_0) = \frac{1}{r}\frac{\partial}{\partial r}\left(r\lambda \frac{\partial T}{\partial r}\right) + n_e(S_e - NS_V)$$

$$- N\frac{\varepsilon_V^p - \varepsilon_V}{\tau_V},$$

$$\frac{\partial \varepsilon_V}{\partial \tau} + v_0 \frac{\partial \varepsilon_V}{\partial r} = \frac{1}{Nr}\frac{\partial}{\partial r}\left(r\chi N \frac{\partial \varepsilon_V}{\partial r}\right) + n_e S_V + \frac{\varepsilon_V^p - \varepsilon_V}{\tau_V},$$

$$N = \frac{P_0}{kT}. \tag{7}$$

To determine the energy scattered by the electrons, it is necessary to solve the equation of energy balance, which can be written in the form[7]

$$\frac{\partial}{\partial \tau}\left(\frac{3}{2}n_e kT_e\right) + \frac{1}{r}\frac{\partial}{\partial r}\left[r\left(\frac{5}{2}n_e kT_e v_0 - \lambda_e \frac{\partial T_e}{\partial r}\right)\right]$$

$$= n_e \frac{e^2 E_0^2}{m_e v_m^e} - n_e N(S_{\mathrm{el}} - S_{\mathrm{in}} - S_i) - n_e^2 S_r + S_T, \tag{8}$$

where the electron temperature is understood as

$$\frac{3}{2}kT_e = \int_0^\infty \varepsilon^{3/2} f(\varepsilon) d\varepsilon;$$

the kinetic energy $m_e v_0^2/2$ has been neglected in comparison with kT_e; λ_e is the electronic thermal conductivity;

$e^2 E_0^2 / m_e \nu_m^e$ is the energy transferred to the electrons by the field; S_{el}, S_{in}, and S_i are the rates of energy loss per molecule by an electron in elastic and inelastic collisions to ionization of the gas; S_r is the rate of energy loss per ion by the electrons during recombination; ν_m^e is the elastic scattering frequency of the electrons; the source term $S_T = (3/2) k T_e^0 S_{str}$ describes the energy acquired by the electrons at the front of the cylindrical ionization wave; and T_e^0 is the equilibrium electron temperature in nitrogen at normal density in the field E_0.

The frequency ν_m^e and the electron energy loss rate are given by

$$\nu_m^e = \sqrt{\frac{2}{m_e}} N \int_0^\infty \varepsilon q_m(\varepsilon) f(\varepsilon) d\varepsilon,$$

$$S_{el} = \sqrt{\frac{2}{m_e}} \frac{2m_e}{M} \int_0^\infty \varepsilon^2 q_m(\varepsilon) \left[f(\varepsilon) + kT \frac{df}{d\varepsilon} \right] d\varepsilon,$$

$$S_{in} = \sqrt{\frac{2}{m_e}} \sum_i \left[\varepsilon_i \int_{\varepsilon_i}^\infty \varepsilon Q_i(\varepsilon) f(\varepsilon) d\varepsilon \right.$$
$$\left. - \varepsilon_i^- \int_0^\infty \varepsilon Q_i^-(\varepsilon) f(\varepsilon) d\varepsilon \right],$$

$$S_i = \sqrt{\frac{2}{m_e}} I \int_I^\infty \varepsilon q_{ion}(\varepsilon) f(\varepsilon) d\varepsilon,$$

$$S_r = \sqrt{\frac{2}{m_e}} \int_0^\infty \varepsilon^2 q_r(\varepsilon) f(\varepsilon) d\varepsilon, \tag{9}$$

where q_m is the transport cross section of electron scattering by a nitrogen molecule; $Q_i = P_i q_i(\varepsilon)$ is the excitation cross section of the ith level with energy ε_i; P_i is the population of the ith level; $Q_i^- = P_i q_i^-(\varepsilon)$ is the deactivation cross section of the ith level with energy ε_i^-; M and m_e are the mass of a nitrogen molecule and an electron, respectively.

To determine the rates of ionization, recombination, and electron energy loss and the collision frequency, it is necessary to know the electron energy distribution function. It is usually assumed to be Maxwellian

$$f(\varepsilon) = f_M(\varepsilon) = \frac{2}{\sqrt{\pi}} \frac{1}{(kT_e)^{3/2}} \exp\left(-\frac{\varepsilon}{kT_e} \right). \tag{10}$$

Here the expression for the electronic thermal conductivity takes the form

$$\lambda_e = \frac{5}{2} k n_e D_e = \frac{5}{2} n_e k T_e \frac{k}{m_e \nu_m^e}, \tag{11}$$

where D_e is the electron diffusion coefficient.

The contribution of recombination to the energy transferred to the medium per unit time S_e can be calculated from the formula

$$S_e^r = k_r n_e \left(\frac{3}{2} k T_e + I_N - I_{N_2^+}^d \right), \tag{12}$$

where $I_N = 14.5$ eV is the ionization potential of atomic nitrogen and $I_{N_2^+}^d = 8.7$ eV is the binding energy of the N_2^+ molecule.

INTERACTION CROSS SECTIONS OF THE ELECTRON WITH THE MEDIUM

A detailed analysis of the interaction cross sections of the electrons with molecular nitrogen was carried out in Ref. 8. Let us describe briefly the set of constants which we will use here. To describe the elastic collisions, we employ the dependence of the transport cross section q_m on the electron energy given in Ref. 9. We describe rotational excitation in line with Refs. 10 and 11. Invoking the smallness of the rotational constant $B_0 = 2.5 \times 10^{-4}$ eV $\ll kT$, kT_e, we can take the distribution over levels to be quasicontinuous. This allows us to replace the sum over levels j by an integral over the corresponding continuum dj from $j = 0$ to $j = \infty$. In particular, for the braking cross section we obtain

$$S_{in}^R = 4 B_0 \sigma_0 \left(1 - \frac{kT}{\varepsilon} \right), \tag{13}$$

where $\sigma_0 = 8\pi/15 (a_0 Q)^2$, a_0 is the Bohr radius, and $Q = 1.05$ is the electric quadrupole moment.

In the description of excitation of vibrations, we take account of only collisions of the first kind, i.e., we assume that the molecules are found in the ground state. The excitation cross sections q_i^V of the first eight levels are given in Ref. 12.

The electron excitation cross sections were taken from the following sources: $A^3\Sigma_u^+$ and $a^1\Pi_g$ from Ref. 13; $B^3\Pi_g$ and $C^3\Pi_u$ from Ref. 14; $b^1\Pi_i$, $b^1\Sigma_u^+$, and the sums of cross sections of the higher-lying states from Ref. 15.

In describing electron impact ionization we restrict ourselves to the process originating in the ground state. The corresponding cross sections were taken from Ref. 16.

In line with Ref. 17, we assign the dissociative recombination rate constants as

$$k_r = 3 \times 10^{-7} \left(\frac{300}{T_e} \right)^{1/2}, \quad \frac{cm^3}{s} ([T_e] = K). \tag{14}$$

For the vibrational relaxation time we use the expression[18]

$$\tau_V = 6.5 \times 10^{-9} \exp\left(\frac{137}{T^{1/3}} \right), \quad s([T] = K). \tag{15}$$

The dependence of the thermal conductivity of molecular nitrogen on the temperature of the medium is taken from Ref. 19.

JUSTIFICATION OF THE HYDRODYNAMIC APPROACH

At high electron concentrations scattering takes place in Coulomb collisions and use of the Maxwell distribution function for the electrons is justified by the intense energy exchange in electron–electron collisions. At low electron concentrations the distribution function is governed by the interaction of the electrons with the neutral gas and use of the Maxwell energy distribution requires justification.
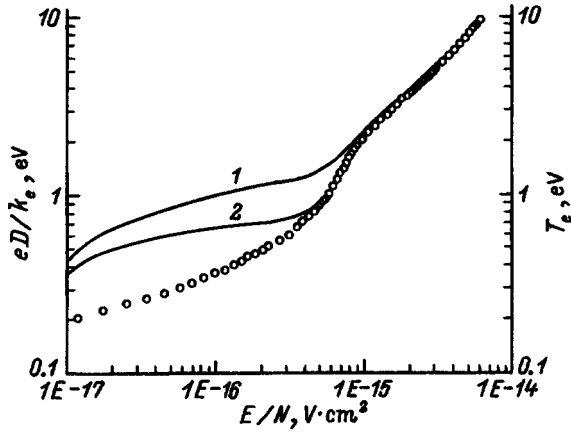
FIG. 1. Dependence on the parameter E/N of the ratio eD/k_e (*1*) and the electron temperature T_e (*2*). $\bigcirc$—the hydrodynamic approximation ($T_e = eD/k_e$).
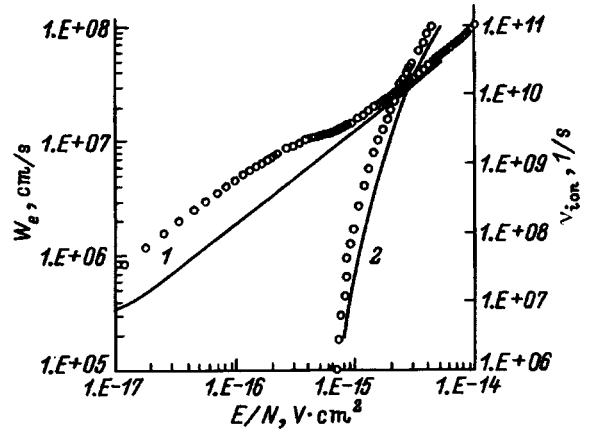

FIG. 2. Dependence on the parameter E/N of the electron drift velocity (*1*) and the ionization rate (*2*). $N = 2.5 \times 10^{19}$ cm^3. Curves—the kinetic approximation, $\bigcirc$—the hydrodynamic approximation.

To describe the electron distribution function f under conditions of interaction with the neutral medium we use the following approach, which provides a good description of the experimental data.[20] Expanding f in a series of Legendre polynomials and limiting it to the first two terms

$$f = f_0(\varepsilon) + \cos \Theta f_1(\varepsilon), \tag{16}$$

where Θ is the angle between the field vector **E** and the electron velocity, we obtain the following equation for the steady-state value of f_0:

$$\frac{\partial}{\partial \varepsilon} \left\{ \left[\frac{e^2 E^2}{N^2} \frac{\varepsilon}{3 q_m(\varepsilon)} + kT\varepsilon^2 \frac{m_e}{M} q_m(\varepsilon) \right] \right.$$
$$\times \frac{\partial}{\partial \varepsilon} f_0(\varepsilon) + \varepsilon^2 \frac{2m_e}{M} q_m(\varepsilon) f_0(\varepsilon) \Big\}$$
$$= \sum_i \left[\varepsilon Q_i(\varepsilon) f_0(\varepsilon) - (\varepsilon + \varepsilon_i) Q_i(\varepsilon + \varepsilon_i) f_0 \right.$$
$$\times (\varepsilon + \varepsilon_i) + \varepsilon Q_i^-(\varepsilon) f_0(\varepsilon) - (\varepsilon - \varepsilon_i^-) Q_i^-(\varepsilon - \varepsilon_i^-)$$
$$\times f_0(\varepsilon - \varepsilon_i^-) \right] + \varepsilon q_{\text{ion}}(\varepsilon) f_0(\varepsilon) - 2 \int_{\varepsilon + I}^{\infty} \varepsilon' q_{\text{ion}}(\varepsilon')$$
$$\times f_0(\varepsilon') \psi(\varepsilon', \varepsilon) d\varepsilon'. \tag{17}$$

The quantity $\psi(\varepsilon', \varepsilon) d\varepsilon$ is the probability that one of the two electrons coming out of the collision will have energy in the interval from ε to $\varepsilon + d\varepsilon$. It is normalized by the condition $\int_0^{\varepsilon' - I_j} \psi(\varepsilon', \varepsilon) d\varepsilon = 1$. For definiteness, we assume the energies of the electrons to be identical $\psi(\varepsilon', \varepsilon) = \delta(\varepsilon - (\varepsilon' - I_j)/2)$. For purposes of comparison of the two approaches to description of the electron dynamics such an assumption is admissible.

Using the distribution function f_0 the ionization and recombination rate constants are found from formulas (3), and the energy loss rate—from formula (9). The drift velocity, electron diffusion coefficient, and electron temperature can be calculated from the formulas

$$\mathbf{W}_e = -\frac{e\mathbf{E}}{3} \sqrt{\frac{2}{m_e}} \int_0^{\infty} \frac{\varepsilon}{N q_m(\varepsilon)} \frac{df_0}{d\varepsilon} d\varepsilon,$$

$$D_e = \frac{1}{3} \sqrt{\frac{2}{m_e}} \int_0^{\infty} \frac{\varepsilon}{N q_m(\varepsilon)} f_0(\varepsilon) d\varepsilon,$$

$$T_e = \frac{2}{3k} \int_0^{\infty} \varepsilon^{3/2} f_0(\varepsilon) d\varepsilon. \tag{18}$$

Under similar conditions, in the hydrodynamic approximation the electron temperature [T_e^0, see Eq. (8)] can be found from the equation of energy balance

$$\frac{e^2 E^2}{m_e \nu_e} = N(S_{\text{el}} + S_{\text{in}} + S_i), \tag{19}$$

where the dependence of the collision frequency and the energy loss rates on the temperature are found from formulas (9) for $f = f_M(T_e, \varepsilon)$ (we have neglected the energy losses to dissociation, assuming the electron concentration to be small).

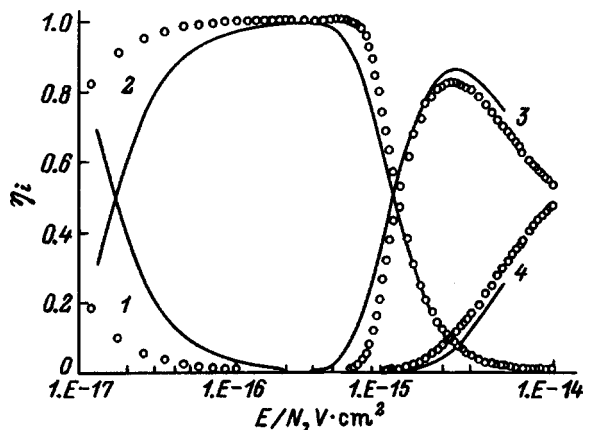Figures 1, 2, and 3 compare the dependence on E/N of

FIG. 3. Dependence of the relative energy losses η_i on E/N. *1*—excitation of rotations together with elastic collisions, *2*—excitation of vibrations, *3*—excitation of electron levels, *4*—ionization. Curves—the kinetic approximation, $\bigcirc$—the hydrodynamic approximation.
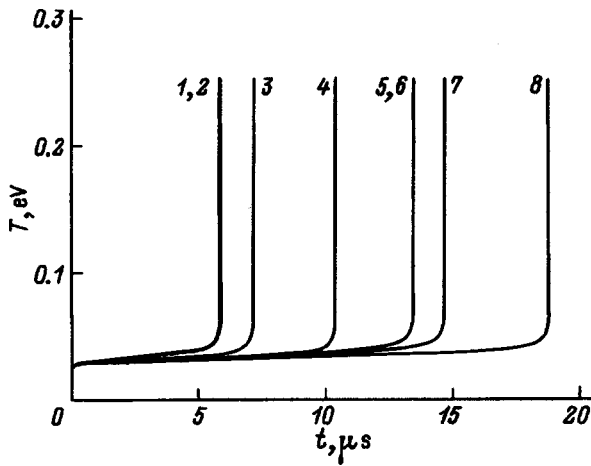
FIG. 4. Time dependence of the temperature of the medium along the discharge axis. r_0, cm: $1—10^{-5}$, $2—10^{-4}$, $3—10^{-3}$, $4—2.5\times10^{-3}$ (for $E_0 = 12.6$ kV/cm and $\nu_z = 2\times10^7$ cm/s), $5—10^{-5}$, $6—10^{-4}$, $7—10^{-3}$, $8—2.5\times10^{-3}$ (for $E_0 = 10.8$ kV/cm and $\nu = 2\times10^8$ cm/s).
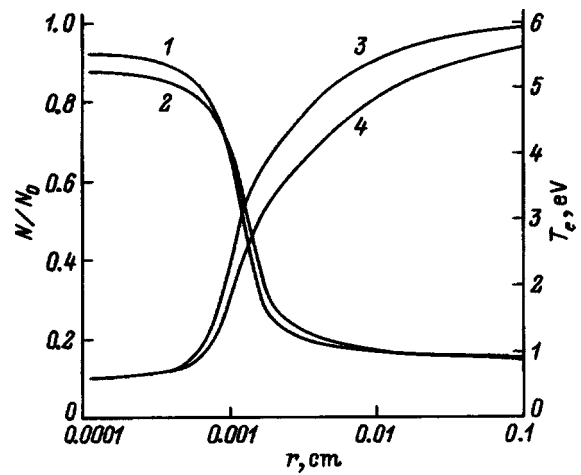

FIG. 5. Spatial distributions of the electron temperature and relative density of the medium. $1,2—T_e$; $3,4—N/N_0$; E_0, kV/cm ν, cm/s: $1,3—12.6$ and 2×10^7; $2,4—10.8$ and 2×10^8.

the electron temperature, the drift velocities, the diffusion coefficients, the ionization frequencies, and electron energy loss rates for molecular nitrogen, obtained by numerical solution of the kinetic equation (17), (18) and in the hydrodynamic approximation (9), (11), (19). Equation (17) was solved by the method of Ref. 21. For $E/N \leqslant 4 \times 10^{-16}$ V·cm² the difference in the quantities is large. In the region $E/N \geqslant 4\times10^{-16}$ V·cm² the hydrodynamic approach gives reasonable agreement with the results obtained from the kinetic treatment.

ANALYSIS OF THE CALCULATED RESULTS

The solution of the problem is defined by Eqs. (1)–(3) and (7)–(12). At the initial time instant the nitrogen is found under normal conditions: $N|_{\tau=0} = N_0 = 2.5\times10^{19}$ cm⁻³, $T|_{\tau=0} = 293$ K, $\varepsilon_V|_{\tau=0} = 0$, the electron concentration $n_e|_{\tau=0} = 0$, and the electron temperature $T_e|_{\tau=0} = 293$ K. The boundary conditions follow from the requirement of axial symmetry: at $r = 0$ the particle and energy fluxes are equal to zero. The ionization and recombination rate constants, energy loss rates, collision frequency, and electronic thermal conductivity follow from formulas (3), (9), (11), and (12) assuming a Maxwellian electron energy distribution (10).

The ionization source S_{str} (2) and the field in the streamer zone are determined by the streamer parameters E_0, v_z/v_r, and v_z, which are unknown. We take the value $v_z/v_r = 10$ (Ref. 5). To find E_0 and v_z, we proceed from the fact that for a given potential U_0, of all possible values of E_0 that one will be realized for which v_z is maximum.

Figure 4 plots the time dependence of the medium temperature along the discharge axis for different values of r_0 (2) in two series of calculations with $E_0 = 12.6$ kV/cm, $v_z = 2\times10^7$ cm/s, and $E_0 = 10.8$ kV/cm, $v_z = 2\times10^8$ cm/s. As r_0 is decreased, the dimensions of the streamer discharge zone decrease and for $r_0 \leqslant 10^{-4}$ cm the solutions are indistinguishable, i.e., for $r_0 \leqslant 10^{-4}$ cm the solution is observed

to converge to the true solution obtained in the limit $r_0 \to 0$. This testifies to the correctness of the mathematical statement of the problem.

Figure 5 plots the electron temperature distribution and medium density distribution across the channel at the instant of time when the value $N/N_0 = 0.1$ is reached. It can be seen that at the earlier stage the scale of leader channel formation is $\sim 10^{-3}$ cm and depends weakly on E_0 and v_z. To describe the further evolution of the channel requires that we introduce Coulomb collisions limiting growth of the electron temperature, take into account nitrogen dissociation limiting growth of the temperature of the medium, describe radiative transfer entraining new masses of gas into the channel due to heating of the cold periphery, etc. In addition, the growing conductivity of and consequently current through the channel affect the magnitude of the electric field (feedback effect). This requires a self-consistent approach to the description of the dynamics of channel formation and the magnitude of the electric field. However, basing ourselves on a string of assumptions it is possible to obtain the dependence of the propagation velocity of the leader and the dimensions of the streamer zone on the potential without complicating the problem by bringing in a description of the enumerated effects.

The nature of the solution—an abrupt growth of the temperature of the medium (decrease in the density) at the stage of channel formation—makes it possible to introduce the concept of a length of the streamer zone L and a potential of leader channel formation $V = E_0 L$. For definiteness we take $L = v_z \cdot t_{0.1}$, where $t_{0.1}$ is the time it takes the channel density to reach the value $N/N_0 = 0.1$. Proceeding from the fact that the leader channel possesses a high temperature (high conductivity), we estimate the value of the leader potential $U_0 \approx V$, i.e., we neglect the potential drop in the channel.

Figure 6 plots the dependence of the potential V on the propagation velocity v_z, calculated for $E_0 = 11.4, 11.64, 12$ kV/cm. The curves have a bell shape: growth of the potential at small velocities gives way to a fall at large velocities. Qualitatively, this may be understood
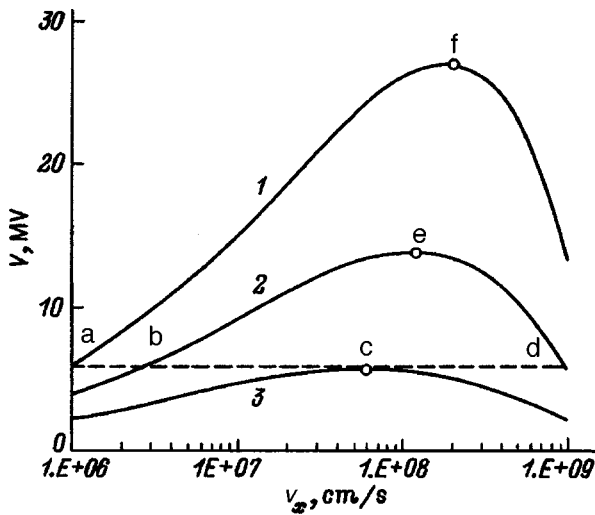
FIG. 6. Dependence of the potential V on the propagation velocity v_z. E_0, kV/cm: *1*—11.4, *2*—11.64, *3*—12.

FIG. 7. Diagram of the process: *1*—leader channel, *2*—streamer zone.

from the following considerations. On the one hand, growth of the velocity is accompanied by an increase of $L = v_z \cdot t_{0.1}$ and consequently also of $V = E_0 L$. On the other, in the adopted streamer zone model an increase in v_z leads to growth of S_{str} (2), i.e., n_e in the streamer zone, and consequently to increased energy release and a decrease in $t_{0.1}$ and L, V. At small velocities the first of these prevails while at large velocities, the second, which leads to the appearance of a maximum in the dependence of V on v_z.

Thus, for a given potential V there is a family of curves $v_z(E_0)$ assigning the propagation velocity of the leader as a function of the electric field in the streamer zone. Thus, for $V = V^*$ the field $E_0 = 11.4$ kV/cm corresponds to the velocity at the point a, i.e., v_a, and at the corresponding point for large v_z; and the field $E_0 = 11.64$ kV/cm corresponds to the velocities v_b and v_d; the field $E_0 = 12$ kV/cm corresponds to the velocity v_c, etc. (Fig. 6), i.e., beginning its motion with velocity v_a, the leader can accelerate to the velocities v_b, v_c, v_d, and so on.

Generally speaking, the propagation velocity and consequently the field in the streamer zone are determined by the propagation velocity of the channel. In turn, the channel parameters and consequently the channel propagation velocity are determined by processes in the streamer zone, i.e., by the field E_0 and the velocity v_z. Let us qualitatively consider the process of acceleration of a leader due to an increase in the channel propagation velocity (Fig. 7). Let at the initial time the channel velocity v_z^c and streamer zone velocity v_z^{str} be identical and equal to the leader propagation velocity v_z. Increasing the field E_0 will lead to an increase in the rate of energy release, i.e., to a growth of v_z^c, and, conversely, a growth of v_z^c leads to a shrinkage of the dimensions of the streamer zone L and consequently to a growth of E_0 ($E_0 = V^*/L$). In other words, growth of the leader velocity is possible with simultaneous growth of E_0 and vice versa. Having started its motion with velocity v_a, a streamer can accelerate to v_c since this acceleration is accompanied by a growth of E_0. Further growth of the velocity, e.g., to v_d, is
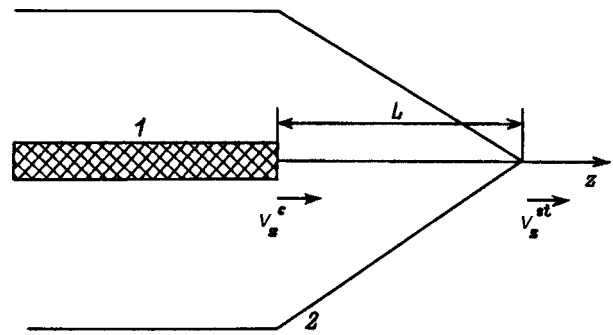
impossible since it is associated with a decrease in the field strength in the streamer zone E_0.

Thus, if we assume that for a given potential the leader propagates with the maximum possible velocity, then its velocity is determined by the maxima in the $V(v_z)$ curves (the points c, e, f in Fig. 8). The dependence $v_z(U_0)$, where U_0 is the potential determined by the condition $\partial V / \partial v_z |_{E_0} = 0$ (E_0 is a parameter here), is plotted in Fig. 8. With growth of U_0 the leader velocity grows roughly as $v_z \sim U_0^{0.75}$. This is clear from comparison with the interpolating v_z dependence also shown there:

$$v_z = 1.47 \times 10^7 U_0^{0.75} \ \mathrm{cm/s}([U_0] = \mathrm{MV}). \tag{20}$$

Thus, within the framework of the ideas expounded above it is possible to construct the steady-state regime of leader propagation, describe the initial stage of channel formation, and obtain the dependence on the potential of the propagation velocity, the field, and the dimensions of the streamer zone. These ideas depend in an essential way on the parameters of the plasma in the streamer zone. These parameters were chosen on the basis of a model of streamer breakdown.[5] Presumably, in the case of a positive streamer this model works at propagation velocities $v_z \geqslant 3 - 4 \times 10^7$ cm/s. At least it makes it possible to explain the linear dependence of the streamer velocity on the rate of growth of
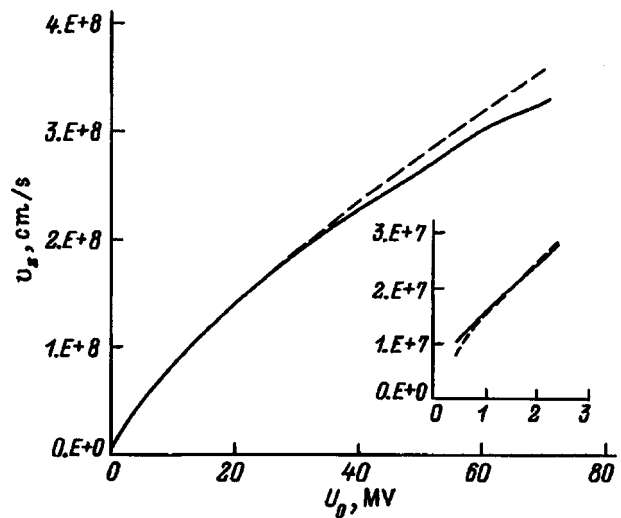

FIG. 8. Dependence of the velocity v_z on the potential U_0. Solid curve—calculated; dashed curve—interpolation based on formula (20).

the potential observed in this region.[3,6] Therefore we can assume that what has been said above is valid for positive leaders for $U_0 \geqslant 3$ MV (Fig. 8), i.e., at high potentials. It cannot be ruled out that at small potentials formation of a leader channel takes place via another mechanism, e.g., against the background of a huge number of continuously forming streamers.[3] The experimental data known to us do not allow us to conjecture about the applicability of the model to negative leaders.

It is interesting to note that the streamer zone fields E_0 depend weakly on the potential U_0 (on the velocity v_z). As U_0 varies from 6 to 26 MV, the field E_0 varies within the range $11.4-12$ kV/cm. This has to do with a peculiarity of energy dissipation in the given range of field strengths (Fig. 3), specifically with a shift of the main channel of energy scattering by the electrons from excitation of molecular vibrations to excitation of electron levels. By reason of retarded deactivation of the vibrational levels of the molecules heating of the medium and consequently channel formation are hindered in the first case. In this sense one may speak of a threshold field for leader breakdown in nitrogen (air) equal to $\sim 11-12$ kV/cm.

CONCLUSION

1. A physical picture of the formation of a leader channel has been presented. A leader channel is formed in the plasma of a streamer zone as a consequence of an ionization–heating instability. The streamer zone consists of an isolated streamer formed by a cylindrical ionization wave.[5] The main mechanism determining the channel parameters at the early stage is the electronic thermal conductivity.

2. A mathematical model has been constructed, describing the leader channel formation stage in a streamer zone plasma. By comparing with the solution of the kinetic equations in the spatially homogeneous case it has been shown that use of the hydrodynamic approximation is justified for $E/N \geqslant 4 \times 10^{-16}$ V·cm^2.

3. Assuming smallness of the potential drop in a highly conductive channel, an algorithm has been constructed for determining the leader propagation velocity v_z and the dimensions L and electric field E_0 in the streamer zone. The algorithm is based on the assumption that for fixed U_0 the value of E_0 is determined that provides the maximum possible propagation velocity.

4. Steady-state leader breakdown in a nitrogen atmosphere has been modeled by a numerical simulation. It was found that the leader propagation velocity is proportional to $U_0^{3/4}$.

5. It has been shown that the field in the streamer zone depends weakly on the potential U_0 and is determined by the value of E/N corresponding to the shift of the electron energy dissipation channel from excitation of vibrations to excitation of electron levels. Under normal conditions this corresponds to the value $E_0 \cong 12$ kV/cm, i.e., in some sense it is possible to speak of the existence of a threshold field for leader breakdown.

6. Presumably in the case of positive leaders the model works at potentials $U_0 \geqslant 3$ MV. It is possible that at low potentials channel formation takes place against a background of a huge number of continuously forming streamers.[3] The experimental data known to us does not allow us to conjecture about the applicability of the model to negative leaders.

7. The model can be used to describe a leader breakdown of a long air gap. When additional processes are added to the model, it will be possible to obtain leader channel parameters matched to the potential U_0.

[1] Yu. P. Raĭzer, *Physics of Gas Discharge* [in Russian] (Nauka, Moscow, 1987), 591 pp.

[2] M. V. Kostenko (Ed.), *High-Voltage Technique* [in Russian] (Vysshaya Shkola, Moscow, 1973), 351 pp.

[3] É. M. Bazelyan and A. Yu. Goryunov, Elektrichestvo, No. 11, pp. 27–33 (1986).

[4] A. É. Bazelyan and É. M. Bazelyan, Teplofiz. Vys. Temp. **32**, 354 (1994).

[5] A. V. Ivanovskiĭ, Zh. Tekh. Fiz. **65**(12), 48 (1995) [Tech. Phys. **40**, 1230 (1995)].

[6] É. M. Bazelyan and I. M. Razhanskiĭ, *Spark Discharge in Air* [in Russian] (Nauka, Novosibirsk, 1988), 164 pp.

[7] Ya. B. Zel'dovich and Yu. P. Raĭzer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, 2nd ed. [in Russian] (Fizmatgiz, Moscow, 1966), 686 pp.

[8] N. L. Aleksandrov, A. M. Konchakov, and É. E. Son, Fiz. Plazmy **4**, 169 (1978) [Sov. J. Plasma Phys. **4**, 98 (1978)].

[9] I. Shimamura, Sci. Pap. Inst. Phys. Chem. Res. (Jpn.) Vol. **82**, pp. 1–51 (1989).

[10] E. Gerjuoy and S. Stein, Phys. Rev. **97**, 1671 (1955).

[11] E. Gerjuoy and S. Stein, Phys. Rev. **98**, 1848 (1955).

[12] G. J. Schulz, Phys. Rev. **135**, A988 (1964).

[13] W. L. Borst, Phys. Rev. A **5**, 648 (1972).

[14] P. N. Stanton and R. M. St. John, J. Opt. Soc. Am. **59**, 252 (1969).

[15] A. E. S. Green and R. S. Stolarski, J. Atmos. Terr. Phys. **34**, 1703 (1972).

[16] D. Rapp and P. Englander-Golden, J. Chem. Phys. **42**, 4081 (1965).

[17] A. V. Eletskiĭ and B. M. Smirnov, Usp. Fiz. Nauk **136**, 254 (1982) [Sov. Phys. Usp. **25**, 13 (1982)].

[18] J. D. Lambert, *Vibrational and Rotational Relaxation in Gases* (Oxford, Clarendon Press, 1977).

[19] V. S. Éngel'sht and B. A. Uryukov (Eds.), *Low-Temperature Plasma. Theory of the Column of an Electric Arc* [in Russian] (Nauka, Novosibirsk, 1990), 215 pp.

[20] L. G. H. Huxley and R. W. Crompton, *The Diffusion and Drift of Electrons in Gases* (Wiley, New York, 1974); Mir, Moscow, 1977, 672 pp.

[21] B. Sherman, J. Math. Anal. Appl. **1**, 342 (1960).

Translated by Paul F. Schippnick

Lowering of ionization potentials in a nonideal plasma

S. I. Anisimov and Yu. V. Petrov

L. D. Landau Institute of Theoretical Physics, Russian Academy of Sciences, 142432 Chernogolovka, Moscow District, Russia
(Submitted February 10, 1997)
Zh. Tekh. Fiz. **68**, 45–50 (June 1998)

Analytical expressions are obtained for the ionization potentials of neutral atoms and ions in the screened Coulomb potential of a nonideal plasma. Among all the chemical elements considered, cesium exhibits the greatest relative lowering of the ionization potentials in comparison with the case of an unscreened interaction. © *1998 American Institute of Physics.*
[S1063-7842(98)00906-4]

INTRODUCTION

Many characteristics of plasmas, in particular the equation of state and the kinetic coefficients, depend substantially on the electron density.[1-11] In a not-too-dense plasma ionization is caused by thermal excitation of electrons to states of the continuum, and at high enough temperatures there exist in the plasma ions in different states of ionization. Such multistep ionization in an ideal plasma is described the the Saha equations,[12] according to which the main parameters determining ionization equilibrium at given plasma temperature and density are the ionization potentials of the atoms and ions. In an ideal plasma these are the ionization potentials of the isolated neutral atoms and of ions in the various charge states. With growth of the density of the plasma and its deviation from ideality a renormalization of the ionization potentials takes place due to effects of interactions between the particles of the plasma. For a weakly nonideal plasma effects of the Coulomb interaction can be taken into account within the framework of the Debye approximation. The correction to the free energy per unit volume due to interaction in the continuum can be expressed as[12] (here and below we use atomic units)

$$\Delta f = -\frac{2}{3}\sqrt{\pi/T}\left(\sum_i n_i Z_i^2\right)^{3/2} = -\frac{1}{3\lambda}\sum_i n_i Z_i^2. \qquad (1)$$

Here n_i is the density of particles of the plasma with charge Z_i, T is the temperature, λ is the Debye screening length leading to an effective interaction of particles with charges Z_1 and Z_2 at the distance r

$$U_D(r) = \frac{Z_1 Z_2}{r} e^{-r/\lambda}. \qquad (2)$$

Screening of the Coulomb interaction for bound electron states in atoms and ions leads to a lowering of their ionization potentials.

LOWERING OF THE IONIZATION POTENTIALS AS A CONSEQUENCE OF SCREENING

The Schrödinger equation for the valence electron with potential (2) does not have an analytical solution in terms of known functions (approximate techniques of solution by the

variational method and by perturbation theory can be found in Ref. 4). Analytical formulas convenient for practical calculations can be obtained by using, instead of Eq. (2), the similar Hulthén potential[13,14]

$$U_H(r) = -\frac{Z^*/\lambda}{e^{r/\lambda} - 1}. \qquad (3)$$

Like the Debye potential, potential (3) for $r \ll \lambda$ goes over to the Coulomb potential $U_c(r) = -Z^*/r$ and falls off exponentially at large r. Z^* is the effective charge of the atomic radical for the valence electron with principal and orbital quantum numbers n and l. We will determine Z^* from the experimentally measured ionization potential $I_{n,l}^0$ of this valence electron in an isolated atom or ion

$$I_{n,l}^0 = \frac{Z^{*2}}{2n^2}. \qquad (4)$$

This gives the n- and l-dependent effective charge

$$Z^* = n\sqrt{2I_{n,l}^0}. \qquad (5)$$

Let us determine the energy levels of the bound states of the electron in a centrally symmetric potential (3). Representing the electron wave function in the usual way in the form of a product of a radial function $R(r)$ and angular functions $Y_{lm}(\theta,\varphi)$ and then transforming to the variable $x = r/\lambda$, we obtain the radial Schrödinger equation for the function $\chi(x) = R(x)/x$

$$-\frac{1}{2\lambda^2}\frac{d^2\chi}{dx^2} + \left(-\frac{Z^*}{\lambda}\frac{1}{e^x - 1} + \frac{l(l+1)}{2\lambda^2 x^2}\right)\chi = E\chi. \qquad (6)$$

In this equation the states with $l = 0$ are a special case. For them the centrifugal term $U_l(x) = l(l+1)/(2\lambda^2 x^2)$ is generally absent, and Eq. (6) can be solved by transforming to the new argument $u = e^{-x}$ (Ref. 13). The energy levels of the discrete spectrum for $l = 0$ are determined by the principal quantum number n and can be written as

$$E_n = -\frac{Z^{*2}}{2n^2}\left(1 - \frac{n^2}{2Z^*\lambda}\right)^2. \qquad (7)$$

Replacing Z^* by the expression for it given in formula (5), we obtain

$$E_n = -I_{n,0}^0 \left(1 - \frac{n}{2\lambda \sqrt{2I_{n,0}^0}} \right)^2. \tag{8}$$

To obtain an analytical expression for the energy levels of the bound states with $l \neq 0$, we replace the centrifugal potential $U_l(x)$ by

$$\tilde{U}_l(x) = \frac{l(l+1)}{2\lambda^2} \frac{1}{(e^x - 1)^2}. \tag{9}$$

The function $\tilde{U}_l(x)$ approximates the function $U_l(x)$ very well. Indeed, for $x \ll 1$ ($r \ll \lambda$) the two functions coincide, but in the localization region of the bound electron $r \sim 1$ ($x \sim 1/\lambda$). Therefore, in a plasma with large λ $\tilde{U}_l(x)$ is a good approximation for $U_l(x)$. With the true centrifugal term $U_l(x)$ replaced by the approximate centrifugal term (9) the radial equation for $\chi(x)$ takes the following form:

$$-\frac{1}{2\lambda^2} \frac{d^2\chi}{dx^2} + \left(-\frac{Z^*}{\lambda(e^x - 1)} + \frac{l(l+1)}{2\lambda^2(e^x - 1)^2} \right)\chi = \tilde{E}\chi \tag{10}$$

and can be solved exactly for any l. We make the transformation of variable $u = e^{-x}$ in Eq. (10) and introduce the notation

$$\alpha = \lambda\sqrt{-2\tilde{E}}, \quad \beta^2 = 2Z^*\lambda.$$

Taking the asymptotic behavior of $\chi(x)$

$$\chi(x) \sim x^{l+1} \sim (1-u)^{l+1} \quad \text{for} \quad x \to 0,$$

$$\chi(x) \sim e^{-\alpha x} \sim u^\alpha \quad \text{for} \quad x \to \infty,$$

into account, we seek the solution of Eq. (10) in the form

$$\chi(u) = u^\alpha (1-u)^{l+1} w(u). \tag{13}$$

Here for the function $w(u)$ we obtain the hypergeometric equation

$$u(1-u)w'' + (2\alpha + 1 - (2\alpha + 2l + 3)u)w' - ((2\alpha + 1)$$
$$\times (l+1) - \beta^2)w = 0. \tag{14}$$

Its finite solution in the limit $u \to 1$ has the form

$$w(u) = F(\alpha + 1 + l + \gamma, -n_r, 2l + 1, u). \tag{15}$$

Here $\gamma = (\alpha^2 + \beta^2 + l(l+1))^{1/2}$, $n_r = -(\alpha + 1 + l + \gamma)$ is the radial quantum number taking nonnegative integer values, and $F(\xi, \eta, \zeta, u)$ is the hypergeometric function

$$F(\xi, \eta, \zeta, u) = \sum_{m=0}^{\infty} \frac{(\xi)_m (\eta)_m}{(\zeta)_m} \frac{u^m}{m!}. \tag{16}$$

The symbol $(\xi)_m$ is defined as

$$(\xi)_m = \xi(\xi+1)\dots(\xi+m-1) \quad \text{for} \quad m > 0,$$

$$(\xi)_m = 1 \quad \text{for} \quad m = 0. \tag{17}$$

For $n_r = 0, 1, 2, \dots$ the hypergeometric series (15) degenerates into a polynomial with degree n_r. Introducing as in the case of the Coulomb potential the principal quantum number $n = n_r + l + 1$, which takes only positive integer values, we obtain the discrete spectrum for any orbital quantum number l

$$\tilde{E}_{n,l} = -\frac{Z^{*2}}{2n^2} \left(1 - \frac{n^2 - l(l+1)}{2Z^*\lambda} \right)^2 \tag{18}$$

$(n = 1, 2, 3, \dots, \ 0 \leq l \leq n-1)$.

Expressing the effective charge Z^* in terms of the ionization potential $I_{n,l}^0$ of an isolated atom or ion, we obtain the ionization potential for an electron with quantum numbers n and l with screening taken into account

$$\tilde{I}_{n,l} = -\tilde{E}_{n,l} = I_{n,l}^0 \left(1 - \frac{n^2 - l(l+1)}{2n\lambda} \sqrt{\frac{1}{2I_{n,l}^0}} \right)^2. \tag{19}$$

As the particles of the plasma we may consider negative ions of the chemical elements. Loss of electrons by such ions can also be examined within the framework of the model expounded above. In this case the ionization energy $I_{n,l}^0$ is the electron affinity ϵ^0 of isolated atoms of these chemical elements. As a rule, $\epsilon^0 \ll 1$ (e.g., for hydrogen $\epsilon^0 = 0.028$); therefore, as follows from Eq. (19), the relative lowering of the electron affinity as a result of screening turns out to be still more significant than the lowering of the ionization potentials. Similarly, the critical value of the screening length, λ_c, at which the renormalized electron affinity vanishes,

$$\lambda_c = \frac{n^2 - l(l+1)}{2n\sqrt{2\epsilon^0}} \tag{20}$$

turns out to be substantially greater than the critical screening length for vanishing of the renormalized first ionization potential (the Mott screening length). This leads to the result that with growth of the temperature the density of negative ions in the plasma falls significantly faster in comparison with the situation where screening of the Coulomb interaction is not taken into account.

The above expressions for the ionization potentials and the electron affinity were obtained by using the approximate centrifugal potential (9). The difference between $U_l(x)$ and $\tilde{U}_l(x)$ can be taken into account in perturbation theory. In the first-order theory

$$\delta U_l(x) = U_l(x) - \tilde{U}_l(x) \tag{21}$$

and the correction to the energy of an electron with quantum numbers n and l is equal to

$$\delta E_{n,l} = \frac{l(l+1)}{2\lambda^2} \frac{\int_0^\infty \chi_{n,l}^2 \left(\frac{1}{x^2} - \frac{1}{(e^x - 1)^2} \right) dx}{\int_0^\infty \chi_{n,l}^2 dx}. \tag{22}$$

Substituting the radial function $\chi_{n,l}$ (13) into expression (22), we write $w^2(u)$ as

$$w^2(u) = \sum_{j=0}^{2n_r} b_j u^j. \tag{23}$$

Here

$$b_j = \begin{cases} \displaystyle\sum_{m=0}^{j} a_m a_{j-m}, & \text{if} \quad j \leq n_r, \\[2ex] \displaystyle\sum_{m=j-n_r}^{n_r} a_m a_{j-m}, & \text{if} \quad j > n_r. \end{cases} \tag{24}$$

The coefficients a_m here are equal to

$$a_m = \frac{(2\gamma - n_r)_m (-n_r)_m}{(2\alpha + 1)_m m!}. \tag{25}$$

Calculating the integrals arising in expression (22), we obtain

$$\delta E_{n,l} = \frac{l(l+1)}{2\lambda^2}$$

$$\times \frac{\Sigma_{m=0}^{2n} c_m (m+2\alpha) \ln(m+2\alpha) - \Sigma_{m=0}^{2(n-1)} \frac{d_m}{m+2\alpha+2}}{\Sigma_{m=0}^{2n} \frac{c_m}{m+2\alpha}}. \tag{26}$$

The coefficients c_m can be expressed in terms of the coefficients b_m by way of the following formulas:
$l \leqslant n_r - 1$

$$c_m = \begin{cases} \sum_{i=o}^{m} (-1)^i \binom{2l+2}{i} b_{m-i}, & \text{if } m \leqslant 2n_r, \\ \sum_{i=m-2n_r}^{2l+2} (-1)^i \binom{2l+2}{i} b_{m-i}, & \text{if } m > 2n_r, \end{cases} \tag{27}$$

for $l > n_r - 1$

$$c_m = \begin{cases} \sum_{i=o}^{m} (-1)^{m-i} \binom{2l+2}{m-i} b_i, & \text{if } m \leqslant 2l+2, \\ \sum_{i=m-2l-2}^{2n_r} (-1)^{m-i} \binom{2l+2}{m-i} b_i, & \text{if } m > 2l+2. \end{cases} \tag{28}$$

In an analogous way the coefficients d_m can be expressed in terms of the coefficients b_m: for $l \leqslant n_r$

$$d_m = \begin{cases} \sum_{i=0}^{m} (-1)^i \binom{2l}{i} b_{m-i}, & \text{if } m \leqslant 2n_r, \\ \sum_{i=m-n_r}^{2l} (-1)^i \binom{2l}{i} b_{m-i}, & \text{if } m > 2n_r, \end{cases} \tag{29}$$

for $l > n_r$

$$d_m = \begin{cases} \sum_{i=0}^{m} (-1)^{m-i} \binom{2l}{m-i} b_i, & \text{if } m \leqslant 2l, \\ \sum_{i=m-2l}^{2n_r} (-1)^{m-i} \binom{2l}{m-i} b_i, & \text{if } m > 2l. \end{cases} \tag{30}$$

In expressions (27)–(30) we have used the standard binomial coefficient notation

$$\binom{i}{j} = \frac{i!}{j!(i-j)!}. \tag{31}$$

The expression for $\delta E_{n,l}$ looks especially simple in the case $l = n - 1$ ($n_r = 0$) corresponding to circular Coulomb orbitals. For electrons with such quantum numbers

$$\delta E_{n,l} = \delta E_{n,n-1} = \frac{l(l+1)}{2\lambda^2} \frac{\Sigma_{m=0}^{2n}(-1)^m \binom{2n}{m}(2\alpha+m)\ln(2\alpha+m) - \Sigma_{m=0}^{2n-2} \frac{(-1)^m}{2\alpha+2+m}\binom{2n-2}{m}}{\Sigma_{m=0}^{2n} \frac{(-1)^m}{2\alpha+m}\binom{2n}{m}}. \tag{32}$$

For large λ the expression for $\delta E_{n,l}$ for all l is simply

$$\delta E_{n,l} = \frac{\alpha l}{2\lambda^2} \tag{33}$$

and the ionization potential with screening taken into account for large λ can be written in the form

$$I_{n,l} = I_{n,l}^0 \left(1 - \frac{n^2 - l(l-1)}{2n\lambda} \sqrt{\frac{1}{2I_{n,l}^0}}\right)^2 \left(1 - \frac{l}{\lambda} \sqrt{\frac{1}{2I_{n,l}^0}}\right). \tag{34}$$

RESULTS

The analytical expressions obtained above for the ionization potentials were used to calculate the first three ionization potentials of the alkali-metal atoms (Li, K, Na, Rb, and Cs)

and inert gases (He, Ne, Ar, Kr, and Xe). The relative lowering of the first ionization potential of the alkali metals in comparison with the case of an unscreened interaction in the isolated atoms is plotted in Fig. 1 as a function of screening length, and of the second and third potentials, in Figs. 2 and 3, respectively. For the first ionization potential the relative lowering turns out to be especially significant for cesium. This is due to the fact that the principal quantum number of its valence electron ($n = 6$) is the largest of all the above-listed alkali metals for the orbital quantum number $l = 0$ (which is the case for all of them) and the first ionization potential of an isolated atom is the lowest ($I_{6.0}^0 = 3.893$ V). The relative lowering of the ionization potential of cesium turns out to be the largest for all of the chemical elements considered.
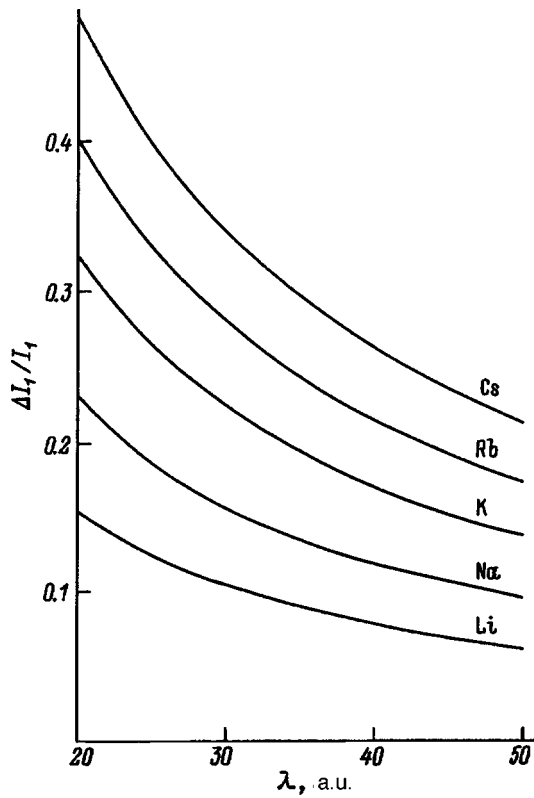
S. I. Anisimov and Yu. V. Petrov

FIG. 1. Relative lowering of the first ionization potential of lithium, sodium, potassium, rubidium, and cesium as a function of the interaction screening length.

FIG. 3. The same as in Fig. 1 for the third ionization potential.

The relative lowering of the two ionization potentials of helium and the first three ionization potentials of the atoms of the other inert gases (Ne, Ar, Kr, and Xe) are plotted in Figs. 4, 5, and 6 as functions of the screening length. Among the elements of this group the greatest relative lowering of

the ionization potentials obtains for Xe. For all the atoms considered this lowering decreases as the charge state of the ion increases: the lowering of the first ionization potential is the most significant. The ionization potentials of isolated atoms of all the elements are taken from Ref. 15.

Figure 7 plots the relative lowering of the electron affinity due to screening of the Coulomb interaction for the series of elements (H, He, Cs). The relative lowering of the electron affinity turns out to be substantially greater than the lowering of the first ionization potential of these elements.

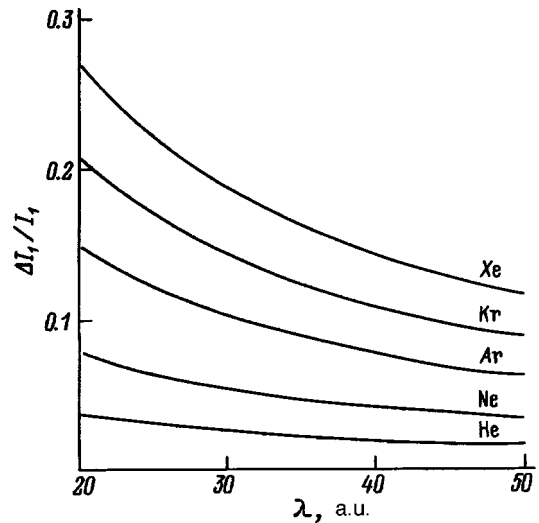FIG. 2. The same as in Fig. 1 for the second ionization potential.

FIG. 4. Relative lowering of the first ionization potential of helium, neon, argon, krypton, and xenon as a function of the screening length in the plasma.
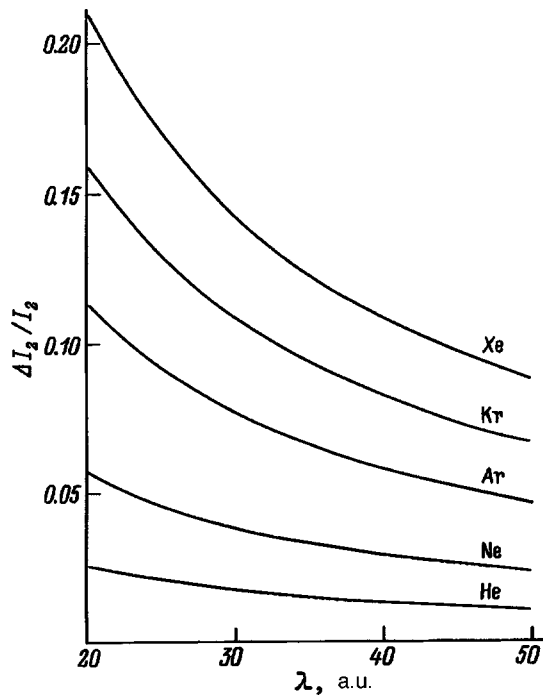
FIG. 5. The same as in Fig. 4 for the second ionization potential.
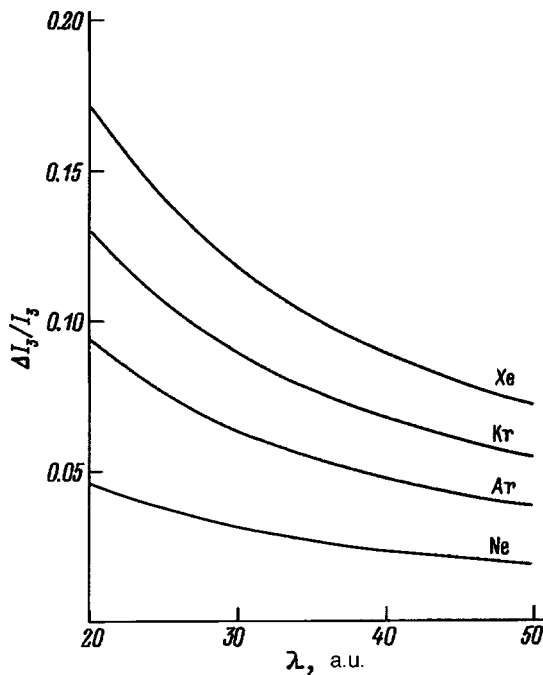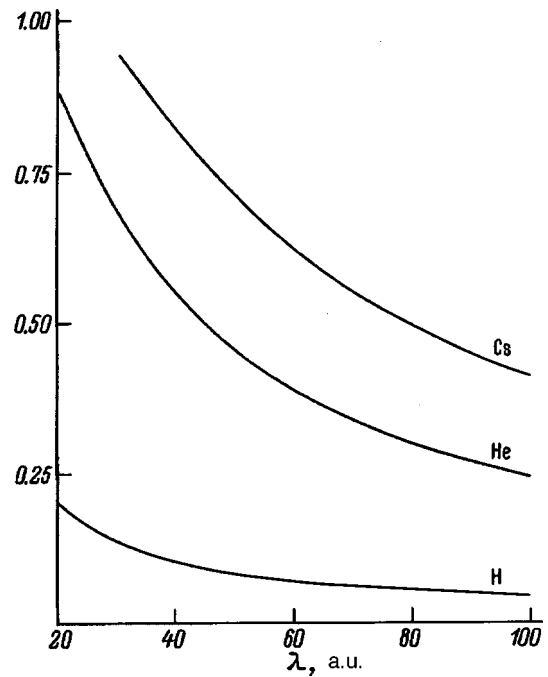

FIG. 7. Relative lowering of the electron affinity of hydrogen, helium, and cesium atoms as a function of the interaction screening length in the plasma.

[1] H. Renkert, F. Hensel, and E. U. Frank, Phys. Lett. A **30**, 494 (1969).
[2] W. Hefner and F. Hensel, Phys. Rev. Lett. **48**, 1026 (1982).
[3] A. Förster, T. Kahlbaum, and W. Ebeling, High Press. Res. **7**, 375 (1991).
[4] W. Ebeling, A. Förster, V. Fortov *et al.*, *Thermophysical Properties of Hot Dense Plasma*, Teubner-Texte zur Physik, Vol. 25 (Teubner, Stuttgart–Leipzig, 1991), 315 pp.
[5] A. A. Likal'ter, Usp. Fiz. Nauk **162**, No. 7, 119 (1992) [Sov. Phys. Usp. **35**, 591 (1992)].
[6] I. T. Yakubov, Usp. Fiz. Nauk **163**, 35 (1993) [Phys. Usp. **36**, 365 (1993)].
[7] A. A. Likal'ter, Zh. Éksp. Teor. Fiz. **106**, 163 (1994) [JETP **79**, 899 (1994)].
[8] A. A. Likal'ter, Teplofiz. Vys. Temp. **32**, 803 (1994).
[9] A. W. DeSilva and H.-J. Kunze, Phys. Rev. E **49**, 4448 (1994).
[10] A. Ng, P. Celliers, G. Xu *et al.*, Phys. Rev. E **52**, 4299 (1995).
[11] A. A. Likal'ter, Zh. Éksp. Teor. Fiz. **107**, 1996 (1995) [JETP **80**, 1105 (1995)].
[12] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 2 vols., 3rd ed. (Pergamon Press, Oxford, 1980).
[13] S. Flügge, *Practical Quantum Mechanics*, 2 Vols. [Springer-Verlag, Berlin, 1971; Mir, Moscow, 1974, 341 pp.].
[14] N. R. Arista, A. Gras-Marti, and R. A. Baragiola, Phys. Rev. **40**, 6873 (1989).
[15] I. K. Kikoin (Ed.), *Tables of Physical Quantities. A Handbook* [in Russian] (Atomizdat, Moscow, 1976), 1008 pp.

FIG. 6. The same as in Fig. 4 for the third ionization potential.

Translated by Paul F. Schippnick

Probe diagnostics of strongly ionized inert-gas plasmas at atmospheric pressure

F. G. Baksht, N. K. Mitrofanov, A. B. Rybakov, and S. M. Shkol'nik

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted March 26, 1997)
Zh. Tekh. Fiz. **68**, 51–55 (June 1998)

The technique of probe measurements (experiment and theory) is applied in a dense, strongly ionized inert-gas plasma at atmospheric pressure. The measurements are performed in a high-current (250–550 A) free-burning argon arc with a thermionic cathode. As a control technique we used spectroscopic measurements. Comparison of calculation with experiment reveals good agreement. The possibility of determining the plasma potential from the measured floating-probe potential is demonstrated. © *1998 American Institute of Physics.*
[S1063-7842(98)01006-X]

INTRODUCTION

Virtues of the probe method of plasma diagnostics include locality, the possibility to carry out measurements in devices from which extraction of light is hindered or completely impossible, etc. In practice, only the probe method allows one to determine the plasma potential—the most important parameter in the study of processes occurring near the electrodes in various discharges. The latter circumstance is the most important stimulus for ongoing studies directed at extending the range of plasma parameters accessible to probe diagnostics.

In a strongly ionized plasma at atmospheric pressure probe measurements have been carried out for some time (see, e.g., Refs. 1 and 2). However, the technique of carrying out such measurements has still not been fully worked out. This is frequently explained by the technical difficulties involved in performing the measurements (high heat flux densities to the probe), but mainly by difficulties of interpretation of the results by virtue of the absence of a consistent theory of current collection by the probe under these conditions. The assertion made in Ref. 1 on the basis of only qualitative arguments that the potential V_f of a floating probe relative to the plasma in a strongly ionized, high-pressure plasma has an absolute value of $|V_f|=(2\pm1)$ V has been widely accepted. However, this estimate, as results of recent calculations have shown,[3,4] is probably incorrect even in order of magnitude, since a calculation gives $|V_f|\geqslant 10$ V.

Since measurements of the floating-probe potential relative to the electrode in atmospheric-pressure arc discharges usually give a value $|\varphi_f|\leqslant 10$ V, in the determination of the cathode potential drop V_k according to Ref. 1 it is possible to obtain a value of V_k more than two times lower than the value that follows from the calculation in Ref. 4. In the determination of the anode drop V_a the choice of the correct value of V_f is no less important, since not only the magnitude but also the sign of V_a depend on it.

The aim of the present work is an experimental check of the theory developed in Ref. 4 and further development of the probe-measurement technique at high ionization levels in an atmospheric-pressure plasma. As the control technique we used spectroscopic measurements.

The experiments were carried out in a freely burning atmospheric-pressure arc in argon, the most frequently used plasma-forming material in plasma devices. Reference 4 presents results of calculations of the probe characteristics in a xenon plasma. The present paper applies the theory developed in Ref. 4 to a calculation of the probe characteristics in an argon plasma.

EXPERIMENTAL SETUP AND MEASUREMENT TECHNIQUE

The experiments were performed in a water-cooled stainless-steel chamber with an inner diameter of 180 mm and a height of 200 mm. The chamber was first pumped down to a pressure of $p\leqslant10^{-3}$ Torr, flushed with argon and then filled with argon to 3–5% above atmospheric pressure. The arc was oriented vertically: a tungsten rod-shaped cathode with diameter $D=2$ mm was located below the arc, and a flat copper water-cooled anode above it. Care was taken not to sharpen the end of the cathode, which had the shape of a hemisphere. The interelectrode gap was 12 mm. Thermionic emission from the cathode was generated by self-heating which was produced by a constant-current (50–70 A) auxiliary (service) arc. The main pulsed discharge source—a generator with low internal resistance, pumping out single rectangular current pulses with $I\leqslant1000$ A and duration up to 5 ms—was connected in parallel with the service arc. The leading edge of the pulse was formed by transitory processes in the discharge and was ~1 ms in duration. Measurements were performed in the steady-state regime 3–4 ms after onset of the pulsed discharge. The experimental setup is described in more detail in Ref. 5.

Spherical probes with diameter $d\equiv2a=0.45$ -0.50 mm were fabricated from a tungsten wire 0.35 mm in diameter. The wire was insulated by Al_2O_3 ceramic 0.75 mm in diameter. The area of the cylindrical part of the probe extending out from the insulator did not exceed 15% of the area of the sphere. With the help of an electromagnet the probe was shot across the arc through the central zone of the arc channel with a speed of $\geqslant1$ m/s. During the motion of

the probe a special device generated time markers which were formed with the help of a comb-type light chopper mounted to the shaft of the probe holder.

With the help of an S9-8 digital storage oscilloscope we recorded the time dependence of the floating-probe potential φ_f relative to the grounded cathode or the ion current I_i when a negative bias of 5–15 V relative to φ_f was applied to the probe. The position of the probe was determined with the help of the time markers. As a result of the unavoidable air gap of the probe holder inside the solenoid of the electromagnet, the accuracy of determination of the position of the probe was ≈ 0.3 mm. Measurements in each of the regimes were carried out in the form of a series of 8–10 oscillograms, after which the results were averaged, discarding outliers whose appearance was connected with the bias of the arc channel. The position of the arc was monitored visually on a screen onto which its magnified image was projected.

Before commencement of the measurements a negative bias of ≈ 20 V relative to φ_f was applied to the probe and the probe was shot through the plasma several times. As the probe passed through the central part of the arc channel, a cathode ''spot'' formed on the surface of the probe, cleaning its surface without causing any noticeable erosion since the current was limited to 10 A. Such a preliminary cleaning prevented the appearance of spots on the probe during ion current measurements which were carried out at somewhat lower (in magnitude) biases. Such cleaning was repeated as needed during the measurements.

To record the arc emission spectrum we used an MDR-3 spectrometer joined to an OSA B&M Spektronik multichannel optical analyzer. The half-width of the instrument function of the spectroscopic setup (for the width of the entrance slit ≤ 0.08 mm) was 0.12 nm. During the measurements the width of the entrance slit of the monochromator was set to 0.15 mm, which ensured undistorted transmission of the actual intensity distribution in the emission spectrum. A 25-nm segment of spectrum was recorded in each exposure.

With the help of a system of crossed mirrors and a two-lens quartz condenser a reduced, horizontally oriented image of the arc was formed in the plane of the entrance slit, which was scanned across the slit by rotating a plane-parallel quartz plate. We estimated the spatial resolution along the discharge axis to be ≤ 0.3 mm. The exposure time (temporal resolution) of the analyzer was varied within the limits 0.2–0.5 ms. The delay of startup of the analyzer was chosen with the intention that the start of recording of the spectrum correspond to the moment the probe passed through the paraxial region of the discharge. The spectral sensitivity of the setup was measured with the help of an SI8-200U reference band-lamp.

The probe and spectroscopic measurements were not performed simultaneously; however, processing of a large quantity of measurements revealed good reproducibility of the arc burning regimes.

THEORY

The operation of the probe in dense plasma was previously analyzed theoretically mainly in application to regimes where collisions of electrons with ions and neutrals do not affect the electron temperature distribution $T_e(r)$ in the probe sheath. In a large number of calculations the value of T_e in the probe sheath was in general not determined but was assigned as a parameter (see, e.g., Refs. 6–9). In fact, here it was assumed that T_e coincides with the electron temperature in the unperturbed plasma. In the present work we calculated for a different situation, typical of rather high pressures, in which as a consequence of intense collisions between the electrons and heavy particles a single temperature $T(r)$ is established in the main part of the probe sheath for both the electrons and the heavy component. This temperature differs from the temperature of the unperturbed plasma T_∞. First of all, this difference is connected with a lowering of the temperature of the heavy particles in the probe sheath from T_∞ to the temperature T_0 of the probe surface. Second, it is connected with a lowering of the electron temperature as a consequence of their motion toward the probe in a retarding electric field in a quasineutral plasma, and also with energy losses by the electrons to ionization and overcoming the retarding potential barrier in the Langmuir sheath of the probe. In the theoretical treatment the probe sheath region was divided up into a series of layers in line with the role of the predominant effects in them (in more detail see Refs. 4 and 10). Under the conditions considered here the following hierarchy of characteristic lengths obtains: $L_0 \ll l_i \ll L_M \ll L_i \ll L_T \ll a$. Here L_0 is the thickness of the Langmuir sheath of the probe, l_i is the mean free path of the ions, L_M is the Maxwellization length of the electrons, L_i is the length of single ionization of the argon atoms by the Maxwellized electrons, L_T is the relaxation length of the temperature of the heavy component to the electron temperature. If the above sequence of inequalities is fulfilled, it is possible to analyze processes in the indicated regions separately, taking into account the presence of narrower layers in the effective boundary conditions. The inner layers for $r \lesssim L_T$ can be treated as planar. Numerical calculations using a technique we developed earlier[4] were performed for a spherical probe of radius $a = 0.25$ mm in an argon plasma at atmospheric pressure.

The condition for realization of the indicated hierarchy of characteristic lengths, in addition to a high enough pressure, includes a comparatively high temperature T_∞ of the plasma. In particular, if the condition $L_i \ll L_T$ is fulfilled, it is necessary that the electron temperature in the ionization region exceed 2 eV (Ref. 4). In this case the temperature of the unperturbed plasma should exceed 2.5 eV. In this case the unperturbed plasma consists only of electrons and doubly charged argon ions Ar^{++}, while in the intermediate region $L_i \ll r < a$ a transition takes place from single to double ionization of the plasma. The extent of the corresponding transition region under the conditions considered here exceed by an order of magnitude the recombination length of the Ar^{++} ions with the electrons. This allows us to use the condition of local thermodynamic equilibrium (LTE) to determine the composition of the argon plasma in the probe sheath everywhere for $r > L_i$.

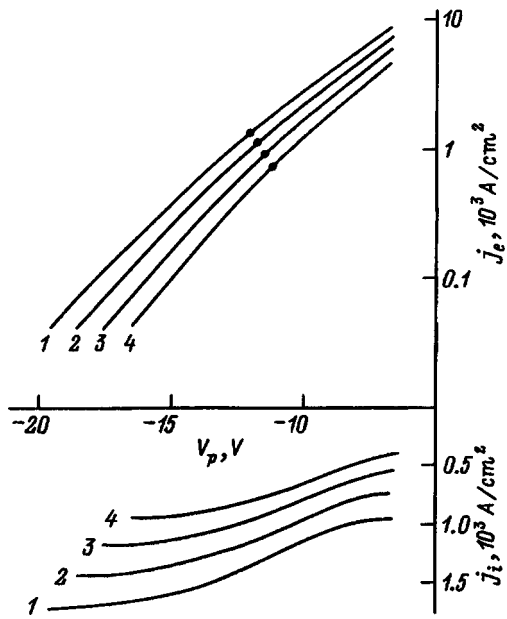Calculated results are shown in Figs. 1 and 2. Figure 1 plots the calculated dependence of the electron current den-

FIG. 1. Density of the electron and ion current to the probe as functions of the probe potential. $T_0 = 0.15$ eV; T_∞, eV: *1*—3.4, *2*—3.2, *3*—3.0, *4*—2.8. The points show the values of the floating-probe potential V_f.

sity j_e and ion current density j_i to the probe on the probe potential V_p relative to the unperturbed plasma. Figure 2 plots the dependence of the floating-probe potential V_f (relative to the unperturbed plasma) on the temperature T_∞.

EXPERIMENTAL RESULTS

As was shown above, the criteria of applicability of the theory are met in a plasma with $T > 2.5$ eV. Such a temperature in a freely burning atmospheric-pressure arc is reached near the cathode. Earlier studies showed[5] that the characteristic dimension of the region of plasma with $T > 2.5$ eV for $I > 250$ A is ≈ 2 mm and grows as the current is increased. By virtue of this fact, measurements were carried out for $I \geqslant 250$ A at a distance $z_0 \approx 1$ mm from the cathode. As calculations have shown,[4] the perturbations introduced by the cold probe in the plasma relax at the distance $L_T \sim 10^{-1}$ mm $\ll z_0$. The size of the current collection region
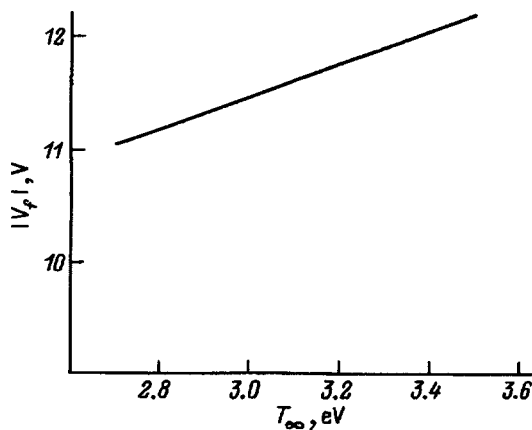

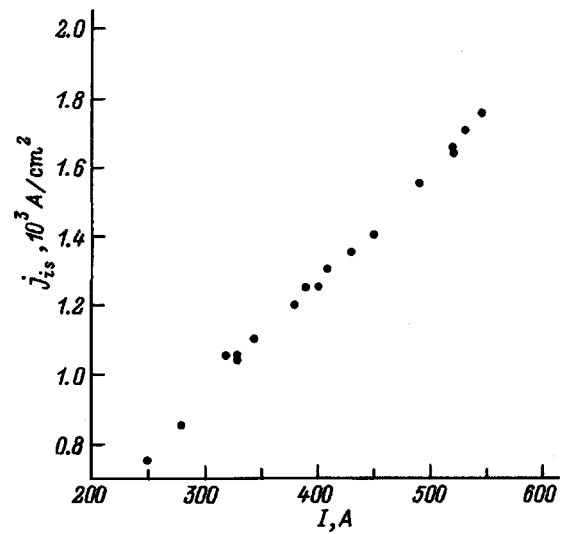FIG. 2. Dependence of the floating-probe potential on the plasma temperature T_∞. $T_0 = 0.15$ eV.

FIG. 3. Saturation value of the ion current density to the probe as a function of the arc current ($z_0 \approx 1$ mm).

is on the order of the probe radius $a = 0.25$ mm and is also quite small in comparison with z_0. In addition, it should be noted that near the cathode at distances $z < D$ there exists a stagnation zone; a cathode jet, whose velocity $v \sim 10^4$ cm/s (Ref. 1), is formed in the region $z \geqslant D$. Therefore probe measurements near the axis of the arc channel at distances $a < z_0 < D$ can be assumed to be correct and can be compared with calculations for an immobile plasma.

The probe measurements were carried out in the following way. For a given discharge current the floating-probe potential φ_f relative to the grounded cathode was determined. Then, different negative biases (5–15 V) relative to φ_f were applied to the probe and the ion current was measured. The dependence of the ion current I_i on time had a wide maximum which corresponded to passage of the probe through the paraxial region of the arc. The maximum value of I_i was reached near the arc axis. In the investigated regimes we observed unmistakable saturation of the ion current with increase of the magnitude of the negative bias of the probe relative to φ_f. The dependence of the saturation ion current density j_{is} measured in this way on the arc current I is plotted in Fig. 3.

The plasma temperature at the distance z_0 from the cathode on the discharge axis was measured as a function of the arc current by the method of relative line intensities without using the Abel inversion. As the results of Ref. 11 show, under the conditions of our measurements the associated error of determination of the temperature does not exceed a few percent. The high temperature of the near-cathode plasma was the reason for using the spectral lines of Ar III for diagnostic purposes. Isolated lines with known probabilities of radiative transitions were chosen[12] which had been used earlier[13,14] in arc plasma thermometry. In the working segment of the emission spectrum 315–340 nm against the background of the recombination–bremsstrahlung continuum we distinctly observed the two triplets of Ar III (six lines in all) $4s'\ ^3D^0 - 4p'\ ^3F$ and $4s\ ^5S^0 - 4p\ ^5P$ with ex-
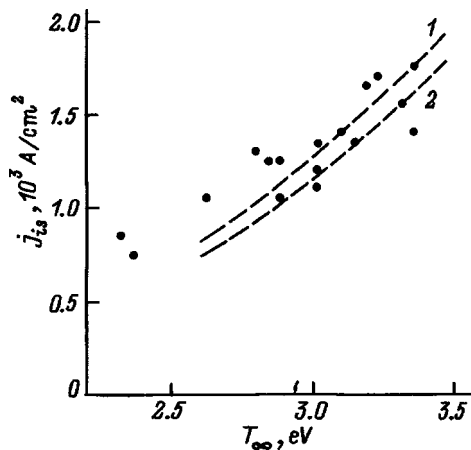
FIG. 4. Dependence of the saturation value of the ion current density to the probe on the plasma temperature. Points—experiment, dashed curves—calculation. T_0, eV: *1*—0.05, *2*—0.25.

citation potentials of the upper levels ≈ 28.1 V and ≈ 25.4 V, respectively. However, for diagnostic purposes we were able to use only the Ar III line pairs with wavelengths (333.61 and 330.19 nm) and (333.61 and 328.58 nm). Measurements on other line pairs gave significant spread and unrealistically exaggerated values of the temperature. This is a result of the impossibility of resolving the implemented spectral lines of Ar III and the near-lying lines of atomic tungsten and tungsten ion and also the lines of carbon ion. The presence of carbon was due to its removal from the cathode pin during arc burning, which was noted earlier.[5] The excitation temperature of the lines measured by this method under the conditions of our experiment coincides with the plasma temperature.[15]

The spectroscopic measurements were used to construct the dependence of the saturation ion current density j_{is} to the probe on the plasma temperature (Fig. 4).

DISCUSSION OF RESULTS

The calculated results plotted in Fig. 1 show that the dependence of the logarithm of the electron current to the probe on the probe potential in a strongly ionized atmospheric-pressure argon plasma is substantially nonlinear. An analogous result was also obtained in the case of xenon.[4] This has to do with the fact that the electron temperature at the surface of the probe is lowered when the potential difference between the probe and the plasma is decreased, i.e., when the electron current is increased. This effect was discovered experimentally several years ago and investigated theoretically in the case of a strongly ionized low-pressure plasma.[16] At a plasma density an order of magnitude lower than in the present work, the effect was significant only when a large electron current was collected by the probe (in comparison with the ion current), but for $I_e \sim I_i$ the probe characteristics turned out to be undistorted and allowed us to determine the electron temperature.

Under the conditions investigated in the present work, that is to say, in a strongly ionized atmospheric-pressure plasma a transitional segment of the probe characteristic is distorted for $I_e \sim I_i$. What is more, the significant decrease of the ion current to the probe in comparison with the saturation ion current at $I_e \sim I_i$ (Fig. 1) prevents us from using the measured probe current–voltage characteristic to directly determine I_e (by extrapolating I_{is}). All this makes the transitional segment of the characteristic unsuitable for diagnostic purposes. Nevertheless, the plasma temperature T_∞ can be determined by measuring the saturation ion current to the probe. Indeed, as calculation shows, j_{is} is a sensitive function of the plasma temperature and depends only very weakly on the surface temperature of the probe (the dashed curves in Fig. 4). This is important since during the measurements the surface temperature of the probe T_0 varies in an uncontrolled way. The maximum value of T_0 depends on the residence time of the probe in the plasma and the probe potential relative to the plasma. Estimates show that in our experiments the surface temperature of the probe could reach 2500 K. The comparison of calculation with the results of experiment shown in Fig. 4 reveals good agreement.

The results obtained confirm that the theory developed in Ref. 4 adequately describes current collection by a floating probe in a strongly ionized atmospheric-pressure plasma. In this work we have demonstrated the possibility of measuring the plasma temperature indirectly from measurements of the saturation ion current, and we have also demonstrated the possibility of determining the plasma temperature from measurements of the potential of the floating probe.

[1] W. Finkelnburg and G. Mecker, *Electric Arcs and Thermal Plasma* (IL, Moscow, 1961), 370 pp.

[2] B. S. Gavryushchenko, R. Ya. Kucherov, A. V. Postogarov *et al.*, Zh. Tekh. Fiz. **45**, 2119 (1975) [Sov. Phys. Tech. Phys. **20**, 1333 (1975)].

[3] E. M. Sklyarova and I. B. Chekmarov, Zh. Tekh. Fiz. **64**(7), 28 (1994) [Tech. Phys. **39**, 649 (1994)].

[4] F. G. Bakit and A. B. Rybakov, Zh. Tekh. Fiz. **67**(12), 16 (1997) [Tech. Phys. **42**, 1385 (1997)].

[5] G. A. Dyuzhev, N. K. Mitrofanov, and S. M. Shkol'nik, Zh. Tekh. Fiz. **67**(1) 35 (1997) [Tech. Phys. **42**, 30 (1997)].

[6] K. N. Ul'yanov, Zh. Tekh. Fiz. **40**, 790 (1970) [Sov. Phys. Tech. Phys. **15**, 613 (1970)].

[7] F. G. Bakit, Zh. Tekh. Fiz. **43**, 214 (1973). [Sov. Phys. Tech. Phys. **18**, 139 (1973)].

[8] F. G. Bakit, G. A. Dyuzhev, N. K. Mitrofanov *et al.*, Zh. Tekh. Fiz. **43**, 2574 (1973) [Sov. Phys. Tech. Phys. **18**, 1617 (1973)].

[9] K. N. Ul'yanov, Teplofiz. Vys. Temp. **16**, 492 (1978).

[10] F. G. Bakit and V. G. Yur'ev, Zh. Tekh. Fiz. **49**, 905 (1979) [Sov. Phys. Tech. Phys. **24**, 535 (1979)].

[11] A. A. Kurskov, E. A. Ershov-Pavlov, and L. V. Chvyaleva, Zh. Prikl. Spektrosk. **45**, 753 (1986).

[12] W. L. Wiese *et al.*, *Atomic Transition Probabilities*, NSRDS-NBS 22, Vol. II, U.S. Dept. of Commerce, Washington, D.C. (1969).

[13] K. Behringer and P. Thoma, J. Quant. Spectrosc. Radiat. Transf. **16**, 605 (1976).

[14] S. Pellerin, B. Pokrzywka, K. Musiol, and J. Chapelle, J. Phys. III (France) **5**, 2029 (1995).

[15] H. W. Drawin, Z. Phys. **228**(2), 99 (1969).

[16] F. G. Bakit, G. A. Dyuzhev, S. M. Shkol'nik, and V. G. Yur'ev, Zh. Tekh. Fiz. **47**, 2280 (1977) [Sov. Phys. Tech. Phys. **22**, 1319 (1977)].

Translated by Paul F. Schippnick

Numerical calculation of the heavy-ion energy spectrum in the cathode sheath of a glow discharge in a gas mixture

V. I. Kristya

Scientific-Research Institute of Materials for Electronics Technology, 248650 Kaluga, Russia
(Submitted April 23, 1997)
Zh. Tekh. Fiz. **68**, 56–59 (June 1998)

A numerical method is developed for solving the equation for the heavy-ion total-energy
distribution function in the cathode sheath of a glow discharge in an inert-gas mixture which
requires much less computer time than the Monte Carlo method. It is shown that it allows
one to calculate with satisfactory accuracy the energy spectrum of the heavy ions bombarding
the cathode in glow-discharge devices. © *1998 American Institute of Physics.*
[S1063-7842(98)01106-4]

In many gas-discharge devices such as gas lasers and displays, the working medium is a mixture of inert gases containing a light gas with a small admixture of a heavy one. Their service life in many cases is a function of the cathode sputtering time, where the main contribution to the sputtering comes from ions of the heavy component.[1,2] To model the evolution of the emission surface of the cathode in such a discharge, it is necessary to know the energy distribution function of the ions of the heavy component of the medium.

The energy distribution function of the ions in the cathode sheath of the discharge has been calculated by analytical methods in a number of papers[3–6] for a discharge in a pure gas. However, as was shown in Ref. 7, the use of such a ion distribution function to calculate the sputtering rate of the cathode in a gas mixture can lead to qualitatively invalid results. On the other hand,[8,9] Monte Carlo simulation of the ion distribution function requires large amounts of computer time, which limits the applicability of this method for modeling cathode sputtering in a glow-discharge plasma in gas mixtures. In Ref. 10, I proposed an equation for the total-energy distribution function of the ions of the heavy component of the mixture and found its analytical solution in the continuous-slowing-down approximation for a heavy ion in a light gas. The expression I obtained there for the ion distribution function allows one with satisfactory accuracy to calculate the cathode sputtering rate in a gas mixture. However, it does not describe a number of properties of the actual distribution function of the ions (in particular, its high-energy tail, due to the stochastic nature of the ion–atom collisions, which is not taken into account in the continuous-slowing-down approximation for an ion in a gas).

In the present paper I develop a method for numerical solution of the equation proposed in Ref. 10 for the distribution function of the heavy ions which allows one to calculate the energy spectrum of the heavy ions without using the continuous-slowing-down approximation and requires much less computer time than the Monte Carlo method.

When an ion moves in a gas of its own species in the cathode sheath of a discharge, the main ion–atom interaction is resonant charge transfer, whose cross section significantly exceeds the elastic collision cross section and the electron ionization cross section of the parent atom. In a mixture containing only a moderate amount of heavy gas, the motion of the heavy ions can be strongly affected by their elastic collisions with atoms of the light gas between charge-transfer events on gas atoms of their own species. Nonresonant charge transfer on atoms of the light gas, on the other hand, can be ignored, since its cross section for inert gases at ion energies below 1 keV is small.[11]

If the mass ratio of the atoms of the two components is large ($M_H/M_L \gg 1$), then the deviation of the trajectories of the heavy ions from the normal to the cathode surface will be small[10] and their motion can be treated as one-dimensional. Thus, if we let the z axis be directed along the normal to the cathode surface, the coordinate $z=0$ correspond to the boundary of the plasma with the cathode sheath, and the coordinate $z=d_c$, to the surface of the cathode, then the heavy-ion distribution function $f(z,\varepsilon)$ in the cathode sheath will satisfy the equation[10]

$$\frac{\partial f}{\partial z} - e \frac{\partial \varphi}{\partial z} \frac{\partial f}{\partial \varepsilon} = \frac{1}{\lambda_c}[\delta(\varepsilon) - f] + \frac{1}{\lambda_e}$$
$$\times \left[\int_\varepsilon^{\varepsilon/(1-\gamma)} \frac{f(z,\varepsilon')d\varepsilon'}{\gamma \varepsilon'} - f \right] \quad (1)$$

with the boundary condition $f(0,\varepsilon) = \delta(\varepsilon - \varepsilon_0)$, where λ_c and λ_e are the heavy-atom resonant charge-transfer length and the elastic collision length of the heavy atoms with the light atoms, $\gamma = 4M_H M_L/(M_H + M_L)^2$, φ is the electric field potential, and ε_0 is the ion energy at the boundary of the cathode sheath.

Integration of Eq. (1) over ε gives the ion flux conservation law

$$\int f(z,\varepsilon)d\varepsilon = 1. \quad (2)$$

Replacing ε by the new variable

$$s = \varepsilon + e\varphi(z), \quad (3)$$

we obtain from Eq. (1) an equation for the function $f(z,s)$, which contains a derivative in only one variable

$$\frac{\partial f}{\partial z} + \left[\frac{1}{\lambda_c} + \frac{1}{\lambda_e}\right]f = \frac{1}{\lambda_c}\delta(s - e\varphi) + \frac{1}{\lambda_e}\int_s^{s_m}\frac{f(z,s')ds'}{\gamma(s'-e\varphi)}, \tag{4}$$

where

$$s_m = \begin{cases} [s - \gamma e\varphi]/(1-\gamma), & s < \varepsilon_0(1-\gamma) + \gamma e\varphi, \\ \varepsilon_0, & s \geqslant \varepsilon_0(1-\gamma) + \gamma e\varphi. \end{cases}$$

Integrating Eq. (4), we find

$$f(z,s) = \int_0^z dz' \ \exp\left[\left(\frac{1}{\lambda_c} + \frac{1}{\lambda_e}\right)(z'-z)\right]$$
$$\times \left[\frac{1}{\lambda_c}\delta(s - e\varphi) + \frac{1}{\lambda_e}\int_s^{s_m}\frac{f(z',s')ds'}{\gamma(s'-e\varphi)}\right]$$
$$+ \delta(s - \varepsilon_0)\exp\left[-\left(\frac{1}{\lambda_c} + \frac{1}{\lambda_e}\right)z\right]. \tag{5}$$

We represent the function $f(z,s)$ in the form

$$f(z,s) = \delta(s - \varepsilon_0)\exp\left[-\left(\frac{1}{\lambda_c} + \frac{1}{\lambda_e}\right)z\right] + h(z,s). \tag{6}$$

The first term in expression (6) is the ion distribution function of the primary ion beam entering the cathode sheath from the discharge, and the second term is the distribution function of the secondary ions formed in the cathode sheath in charge-transfer and elastic collisions of the primary ions. Substituting expression (6) into formula (5), we obtain an equation for $h(z,s)$

$$h(z,s) = \int_0^z dz' \ \exp\left[\left(\frac{1}{\lambda_c} + \frac{1}{\lambda_e}\right)(z'-z)\right]$$
$$\times \left[\frac{1}{\lambda_c}\delta(s - e\varphi) + \frac{1}{\lambda_e}\int_s^{s_m}\right.$$
$$\times \left.\frac{\delta(s' - \varepsilon_0)\exp[-(1/\lambda_c + 1/\lambda_e)z'] + h(z',s')}{\gamma[s' - e\varphi(z')]}ds'\right]. \tag{7}$$

After differentiating with respect to z it is possible to reduce this equation to the form

$$\frac{\partial h}{\partial z} = -\left(\frac{1}{\lambda_c} + \frac{1}{\lambda_e}\right)h(z,s) + \frac{1}{\lambda_c}\delta(s - e\varphi) + \frac{1}{\lambda_e}$$
$$\times \exp\left[-\left(\frac{1}{\lambda_c} + \frac{1}{\lambda_e}\right)z\right]\frac{\theta\{s - [\varepsilon_0 - \gamma(\varepsilon_0 - e\varphi)]\}}{\gamma(\varepsilon_0 - e\varphi)}$$
$$+ \frac{1}{\lambda_e}\int_s^{s_m}\frac{h(z,s')}{\gamma(s' - e\varphi)}ds', \tag{8}$$

where $\theta(x)$ is the Heaviside step function.

The boundary condition for Eq. (8) follows from the boundary condition for the function $f(z,s)$ and has the form $h(0,s) = 0$. The function $h(z,s)$, as follows from Eq. (3), is

nonzero in the interval from $e\varphi(z)$ to ε_0, and the condition of ion flux conservation (2) after substitution of relation (6) takes the form

$$\int_{e\varphi}^{\varepsilon_0}h(z,s)ds = 1 - \exp\left[-\left(\frac{1}{\lambda_c} + \frac{1}{\lambda_e}\right)z\right]. \tag{9}$$

To find a numerical solution to Eq. (8) we can use a method similar to that proposed in Ref. 12 to solve the kinetic equation for the electron distribution function in the cathode sheath of a discharge.

We divide the interval $[0, d_c]$ along the z axis into n segments of length $\Delta z = d_c/n$, whose end-points are equal to $z_i = i\Delta z$, $i = 0,1,\ldots,n$. We divide up the interval $[e\varphi(z_i), \varepsilon_0]$ of variation of s in each cross section $z = z_i$ into m segments of length $\Delta s_i = (\varepsilon_0 - e\varphi(z_i))/m$, i.e., $s_k^i = e\varphi(z_i) + k\Delta s_i$, $k = 0,1,\ldots,m$. In the first cross section, as follows from the boundary condition for Eq. (8), we have $h(z_0, s_k^0) = 0$. To find the values of the function $h(z,s)$ in other cross sections from Eq. (8), we can use the Cauchy–Euler method,[13] which gives

$$h(z_i, s_k^i) = h(z_{i-1}, s_k^i) + h_z(z_{i-1}, s_k^i)\Delta z. \tag{10}$$

The values $h(z_{i-1}, s_k^i)$ are determined by interpolation over the known values $h(z_{i-1}, s_k^i)$ for $s_k^i \geqslant s_0^{i-1}$ and are set equal to zero for $s_k^i < s_0^{i-1}$. The function $h_z(z_{i-1}, s_k^i)$ is a finite-difference approximation of the right-hand side of Eq. (8)

$$h_z(z_{i-1}, s_k^i) = -\left(\frac{1}{\lambda_c} + \frac{1}{\lambda_e}\right)h(z_{i-1}, s_k^i) + \theta(-s_k^i + s_0^{i-1})\alpha_k^i$$
$$+ \frac{1}{\lambda_e}\exp\left[-\left(\frac{1}{\lambda_c} + \frac{1}{\lambda_e}\right)z_{i-1}\right]\frac{\theta(s_k^i + \varepsilon_m^{i-1} - \varepsilon_0)}{\varepsilon_m^{i-1}}$$
$$+ \frac{1}{\lambda_e}\int_{s_k^i}^{\tau_m^i}\frac{h(z_{i-1}, s')}{\gamma[s' - e\varphi(z_{i-1})]}ds', \tag{11}$$

where

$$\varepsilon_m^{i-1} = \gamma[\varepsilon_0 - e\varphi(z_{i-1})],$$

$$\tau_m^i = \begin{cases} [s_k^i - \gamma e\varphi(z_{i-1})]/(1-\gamma), \\ \qquad s_k^i < \varepsilon_0(1-\gamma) + \gamma e\varphi(z_{i-1}), \\ \varepsilon_0, \quad s_k^i \geqslant \varepsilon_0(1-\gamma) + \gamma e\varphi(z_{i-1}). \end{cases}$$

The function α_k^i describes the distribution of the ions that have undergone charge transfer in the segment $[z_{i-1}, z_i]$ in the interval $[s_0^i, s_0^{i-1}]$. The correct choice of its form, as calculations show, has a substantial affect on the calculated ion distribution function. Since $\Delta z \ll \lambda_e$, this function can be found from Eq. (8) in neglect of the elastic scattering of ions at a distance Δz after charge transfer, which gives

$$\alpha_k^i = \frac{\exp(z_{sk}^i/\lambda_c)}{\lambda_c e\varphi'(z_{sk}^i)[\exp(z_{s0}^{i-1}/\lambda_c) - \exp(z_{s0}^i/\lambda_c)]}, \tag{12}$$

where the quantities z_{sk}^i, z_{s0}^i, and z_{s0}^{i-1} are found from the equation $s - e\varphi(z) = 0$ for s equal to s_k^i, s_0^i, and s_0^{i-1}, respectively.
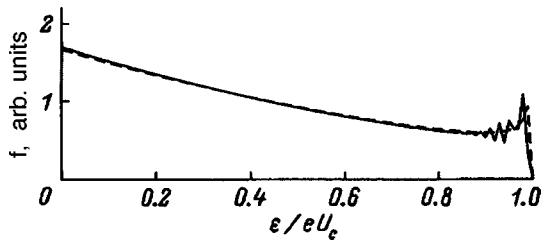
FIG. 1. Ion distribution function at the cathode surface, as calculated by the proposed method (solid line) and the exact solution (13) (dashed line) in the absence of elastic collisions of the heavy ions with the light atoms ($\lambda_c/d_c = 0.313$, $\lambda_e/d_c \gg 1$).

To avoid nonconservation of the ion flux due to discretization errors, at each step in z the function $h(z,s)$ is renormalized on the basis of relation (9). Using relations (6), (9)–(12) successively for $i = 1,2,\dots,n$, it is possible to find the energy spectrum of the heavy ions over the entire extent of the cathode sheath from the boundary with the plasma to the cathode surface.

To estimate the accuracy of the proposed method of solution of Eq. (1), calculations were carried out without taking into account the elastic collisions of the heavy ions with the light atoms ($\lambda_e \gg d_c$). In this case the velocities of the ions are directed along the z axis, and the ion distribution function, if the initial energy ε_0 of the ions at the boundary of the cathode sheath is neglected, is given by[3,10]

$$f(z,\varepsilon) = \frac{1}{e|\varphi'[z_0(\varepsilon)]|}\left[\delta[z_0(\varepsilon)] + \frac{1}{\lambda_c}\right.$$
$$\left.\times\exp\left(\frac{z_0(\varepsilon)}{\lambda_c}\right)\right]\exp\left(-\frac{z}{\lambda_c}\right), \qquad (13)$$

where the dependence $z_0(\varepsilon)$ is given by $\varepsilon = e[\varphi(z_0) - \varphi(z)]$.

The results obtained for the case of a quadratic dependence of $\varphi(z)$ (Refs. 3 and 14) [$\varphi(z) = -U_c(z/d_c)^2$, where U_c is the cathode potential drop] and $m = n = 100$ are plotted in Fig. 1, whence it follows that for the given number of divisions of the spatial and energy intervals of variation of the function $h(z,s)$ the numerical solution coincides quite well with the exact solution. Therefore these values of m and n were used in further calculations of the ion distribution function in the presence of elastic collisions, where an exact solution of the problem is lacking.

Figure 2 plots the ion distribution function for three values of the discharge current density j in a 15:1 helium–neon mixture at a pressure of 6 Torr and temperature of 300 K ($\varepsilon_0 = 4$ eV, with the values of U_c and d_c determined from the Aston model[1]), as found by the given method and as obtained by a two-dimensional Monte Carlo simulation of the motion of the ions in the cathode sheath by the technique described in Ref. 10. It can be seen that the method of calculating the ion distribution function proposed here gives results which are very close to the two-dimensional Monte Carlo results. Computer time requirements using the one-dimensional model are roughly an order of magnitude less, and in contrast to the Monte Carlo method they do not grow with the ratio d_c/λ_c.
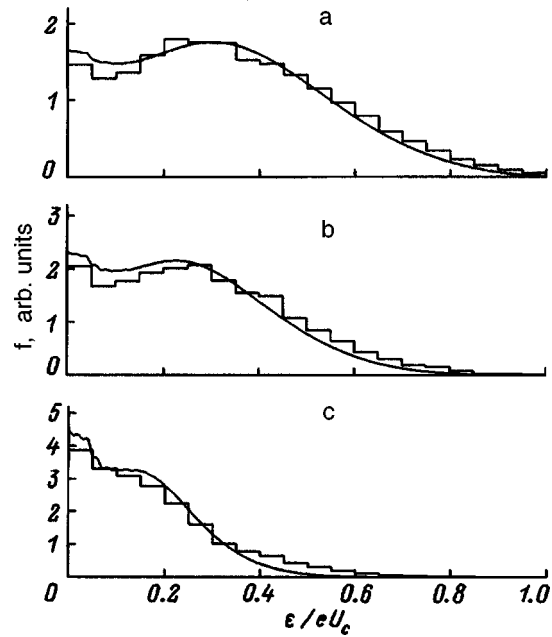

FIG. 2. Distribution function of ions of the heavy component at the cathode surface in a helium–neon mixture, calculated by the proposed method (curves) and by the Monte Carlo method (histograms) for three discharge regimes. j, mA/cm²: a—1.0, b—0.6, c—0.2; U_c, V: a—291, b—272, c—250; d_c, cm: a—0.230, b—0.269, c—0.409; λ_c/d_c: a—0.313, b—0.226, c—0.124; λ_e/d_c: a—0.125, b—0.090, c—0.050.

Figure 3 plots the dependence of the cathode sputtering coefficient averaged over the ion energies

$$R = \int Y(\varepsilon)f(d_c,\varepsilon)d\varepsilon \qquad (14)$$

on the discharge current density, calculated on the basis of the ion distribution function found by different methods [$Y(\varepsilon) = a(\varepsilon - \varepsilon_t)^2$ is the ion cathode sputtering coefficient,[1] ε_t is the threshold sputtering energy, and a is a constant for the given kind of ions and cathode material]. There is good agreement between the results obtained from the one-dimensional and two-dimensional models.

Consequently, the proposed method, based on a numerical solution of the one-dimensional equation for the ion distribution function, allows one to calculate the energy spec-
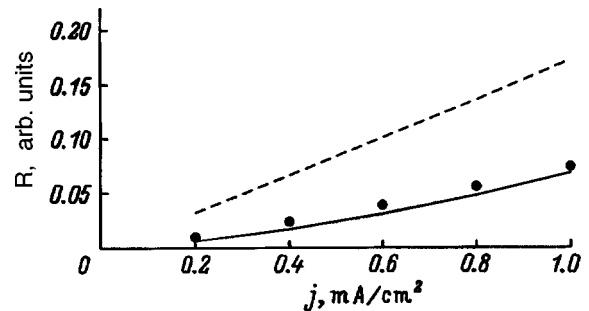

FIG. 3. Dependence of the cathode sputtering coefficient averaged over the energy spectrum of the ions, on the discharge current density in a helium–neon mixture for $\varepsilon_t = 30$ eV, as found by the Monte Carlo method (points), by the proposed method (solid line), and on the basis of the ion distribution function (13) for a single-component gas (dashed line).

trum of the heavy ions in the cathode sheath of a glow discharge in an inert-gas mixture with satisfactory accuracy and without the use of large amounts of computer time. It can be used to create a self-consistent model describing the interaction with the cathode surface for a glow-discharge plasma of complicated composition.

[1] L. H. Hall, J. Appl. Phys. **64**, 2630 (1988).

[2] G. G. Bondarenko, A. P. Korzhavyĭ, V. I. Kristya *et al.*, Metally, No. 5, 54 (1996).

[3] I. Abril, A. Gras-Marti, and J. A. Valles-Abarca, Phys. Rev. A **28**, 3677 (1983).

[4] V. V. Kuchinskiĭ, V. S. Sukhomlinov, and E. G. Sheĭkin, Zh. Tekh. Fiz. **55**, 67 (1985) [Sov. Phys. Tech. Phys. **30**, 38 (1985)].

[5] V. P. Konovalov, J. Bretagne, and G. Gousset, J. Phys. D **25**, 1073 (1992).

[6] O. B. Firsov and V. V. Kuchinskiĭ, Fiz. Plazmy **18**, 114 (1992) [Sov. J. Plasma Phys. **18**, 61 (1992)].

[7] A. P. Korgiaviy and V. I. Kristya, J. Appl. Phys. **70**, 5117 (1991).

[8] B. E. Thompson and H. H. Sawin, J. Appl. Phys. **63**, 2241 (1988).

[9] R. T. Farouki, S. Hamaguchi, and M. Dalvie, Phys. Rev. A **44**, 2664 (1991).

[10] V. I. Kristya, Zh. Tekh. Fiz. **66**(6), 8 (1996) [Tech. Phys. **41**, 525 (1996)].

[11] J. B. Hasted, *Physics of Atomic Collisions* [Butterworths, London, 1964; Mir, Moscow, 1965, 710 pp.].

[12] T. C. Paulick, J. Appl. Phys. **67**, 2774 (1990).

[13] D. Potter, *Computational Physics* [Wiley, New York, 1973; Mir, Moscow, 1975, 392 pp.].

[14] A. P. Korzhavyĭ and V. I. Kristya, Zh. Tekh. Fiz. **63**(2), 200 (1993) [Tech. Phys. **38**, 156 (1993)].

Translated by Paul F. Schippnick

Motion of the cathode spot of a vacuum arc in an external magnetic field

S. A. Barengol'ts, E. A. Litvinov, E. Yu. Sadovskaya, and D. L. Shmelev

Institute of Electrical Physics, Ural Branch of the Russian Academy of Sciences, 620049 Ekaterinburg, Russia
(Submitted May 19, 1997)
Zh. Tekh. Fiz. **68**, 60–64 (June 1998)

The problem of the motion of the cathode spot of a vacuum arc electrical discharge in a magnetic field applied tangential to the cathode surface is considered. The treatment is based on concepts of the nonstationary, cyclical nature of processes occurring in the cathode spot and the key role of return electrons falling out of the near-cathode plasma back onto the cathode. © *1998 American Institute of Physics.* [S1063-7842(98)01206-9]

INTRODUCTION

In the present paper we consider the paradoxical phenomenon of the retrograde motion of the cathode spot of a vacuum arc in a tangential magnetic field.[1] This phenomenon is attracting research interest because no generally accepted explanation has yet appeared. The existence and motion of the cathode spot is treated as an nonstationary, cyclical process of appearance and dying off of emission centers or explosive centers.[2,3] The direction of motion of the spot is the direction in which a new center will preferentially arise in place of the old one. On the basis of these ideas we will attempt to explain the phenomenon of retrograde motion.

MODEL OF RETROGRADE MOTION OF THE CATHODE SPOT

Numerical modeling of the expansion of a plasma jet from the emission center of a cathode spot has shown that there exists in the vicinity of the center a ring current carried by return electrons moving from the plasma to the cathode and closed through the emission zone of the center (Fig. 1).[4] The motion of the return electrons forms current loops which together create a toroidal surface. The symmetry axis of the torus is perpendicular to the cathode surface and passes through the center of the spot. The geometry of this arrangement is shown in Fig. 2.

The process of formation of a new emission center is linked with the return electron current. Where this current is
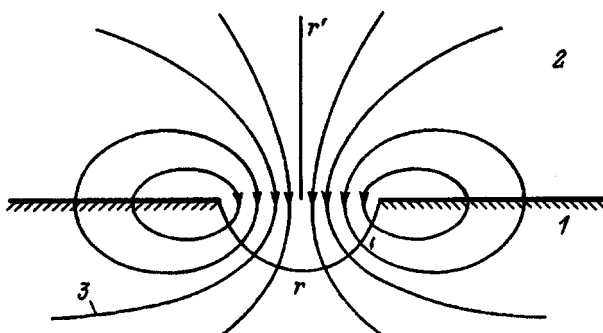

FIG. 1. Geometry of an emission center: *1*—cathode, *2*—plasma, *3*—current lines.

greater, favorable conditions arise for the formation of a new center.[5] The direction of motion of the spot is toward the region where the current of return electrons concentrates.

The magnetic field created by the transport current from the emission center i_1 can be estimated from the formula[6,7]

$$\mathbf{B}_1 = \frac{\mu_0}{4\pi}\frac{2i_1}{r} \tag{1}$$

and is directed according to the ''right-hand rule.'' The magnetic field of the current i_2 of the return electrons is given by

$$\mathbf{B}_2 = \frac{\mu_0}{4\pi}\frac{2i_2}{r}, \tag{2}$$

and this field is nonzero only inside the current torus, where it adds together with the field of the forward current. In the presence of an external magnetic field tangential to the cathode surface, a torque equal to[7]

$$\mathbf{M} = [\mathbf{p}_m \mathbf{B}_\Sigma], \tag{3}$$

will act on each current loop. Here the square brackets denote the vector (cross) product, $\mathbf{B}_\Sigma$ is the total magnetic field, equal to the vector sum of all the fields, and the vector quantity $\mathbf{p}_m$ is the magnetic moment of the current loop carrying a current i

$$\mathbf{p}_m = i\mathbf{S}, \tag{4}$$

where $\mathbf{S}$ is a vector equal in magnitude to the area encompassed by the loop and having the direction prescribed by the right-hand rule.

The torque will rotate the current loop in such a way that the vectors $\mathbf{p}_m$ and $\mathbf{B}_\Sigma$ become parallel and the plane of the loop becomes perpendicular to $\mathbf{B}_\Sigma$. The resulting action on the current torus will have the direction shown in Fig. 2 (**F**), i.e., the loop will tend to ''swing around'' to the ''anti-Ampère'' direction, or the direction of maximum magnetic field. In addition, in a nonuniform magnetic field the current loop is acted on by the force

$$\mathbf{G} = \text{grad}(\mathbf{p}_m \mathbf{B}_\Sigma), \tag{5}$$

which pulls the loop into the region of higher magnetic field. All this leads to a bunching up of the current lines and an
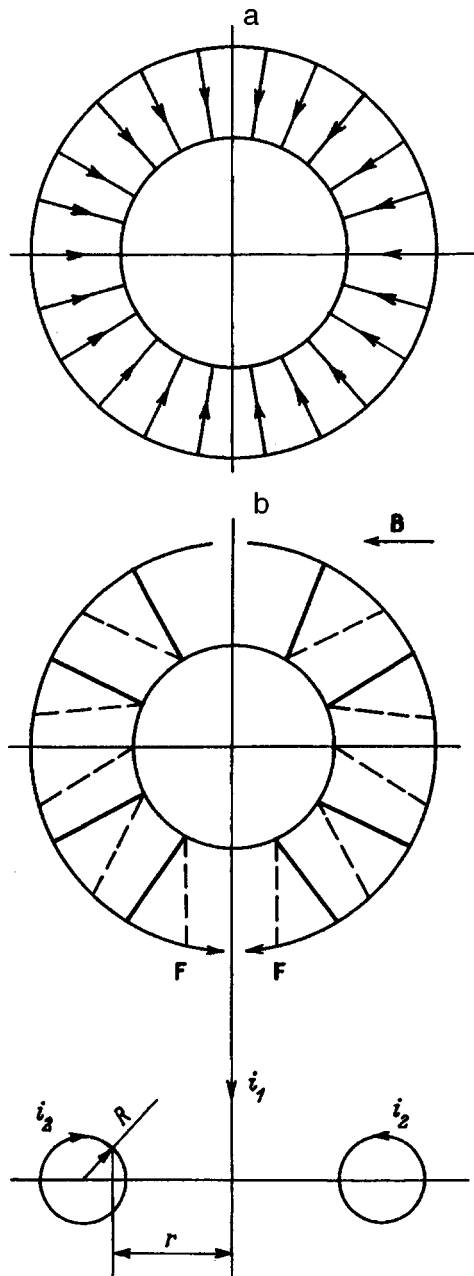
FIG. 2. a—schematic depiction of the arrangement of current contours of return electrons about an emission center, b—view from above and to the side. Dashed lines depict reaction of the contours to an external tangential magnetic field. Arrows on rings show the direction of current. The direction of motion of the electrons is opposite.

increase in the current density of the return electrons from the plasma to the cathode in precisely the retrograde direction.

If the external magnetic field is oriented at an angle with respect to the cathode surface, then the picture of motion of the spot changes. The so-called Robson angle effect arises. This is illustrated in Fig. 3. Figure 4 shows a diagram elucidating this phenomenon. In the given case the current loop not only swings around but also inclines relative to the cathode plane, tending to occupy the position in which its plane becomes perpendicular to the magnetic field vector. Proceed-

ing from the diagram shown in Fig. 4, we can write down the following relations:

$$A_1 A_2 = R \, \cos\left(\frac{\pi}{2} - \theta_B \right),$$ (6)

and

$$\varphi \approx \arctan\left(\frac{R}{R+r} \, \cos\left(\frac{\pi}{2} - \theta_B \right) \right).$$ (7)

Here R is the radius of the current ring and r is the radius of the emission zone. Figure 5 provides a comparison of the picture developed above with experiment. Curves *1* and *2* correspond to experiments with a dirty cathode surface on which the craters are small and relatively far apart.[8,1] In this case, relation (7) must be augmented by the condition $R \gg r$. The corresponding calculated curve (*4*) is also shown.

In experiments with a good vacuum and a well-cleaned cathode surface the size of the craters is significantly greater and they lie one on top of another (curve *3*).[9] In this case it may be assumed that $R \approx r$ (curve *5*).

In addition to this, one more circumstance should be noted. The cathode is separated from the near-cathode plasma by a space-charge sheath in which the charged particles, electrons and ions, move without collisions. To describe the particle motion in the sheath layer we will use the full system of Vlasov equations, with the goal of obtaining the momentum and energy conservation laws of the particles and the field. We write the Vlasov equations together with Maxwell's equations:

$$\frac{\partial f_\alpha}{\partial t} + \mathbf{v} \cdot \nabla f_\alpha + \frac{e_\alpha}{m_\alpha} (\mathbf{E} + [\mathbf{v}\mathbf{H}]) \nabla_v f_\alpha = 0,$$ (8)

$$\nabla \cdot \mathbf{E} = \frac{1}{\varepsilon_0} \sum_\alpha e_\alpha \int f_\alpha d\mathbf{v} = \frac{1}{\varepsilon_0} \sum_\alpha e_\alpha \langle 1 \rangle_\alpha,$$ (9)

$$[\nabla \mathbf{H}] = \varepsilon_0 \frac{\partial E}{\partial t} + \sum_\alpha e_\alpha \int \mathbf{v} f_\alpha d\mathbf{v} = \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} + \sum_\alpha e_\alpha \langle \mathbf{v} \rangle_\alpha,$$ (10)

$$[\nabla \mathbf{E}] = -\frac{\partial \mathbf{B}}{\partial t},$$ (11)
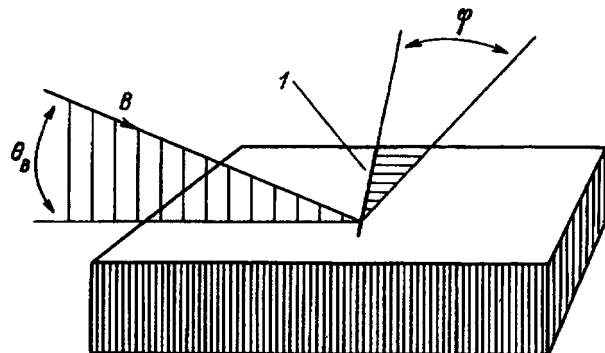
$$\nabla \cdot \mathbf{B} = 0.$$ (12)

FIG. 3. Illustration of the Robson angle effect (*1* is the track of a spot).

Here f_α denotes the distribution function of particles of sort α (electrons and ions from the plasma, electrons from the cathode); e_α and m_α are the charge and mass of the particles; **E** and **H** are the electric and magnetic field vectors; the magnetic field induction $\mathbf{B} = \mu_0\mathbf{H}$, ε_0 and μ_0 are the electric and magnetic constants, $\nabla = \partial/\partial\mathbf{x}$; $\nabla_v = \partial/\partial\mathbf{v}$; the angle brackets denote averaging over the corresponding distribution function.

We multiply Eq. (8) by $m_\alpha\mathbf{v}$, integrate over **v**, and carry out the sum over α. We then have

$$\frac{\partial}{\partial t}\sum_\alpha \langle m_\alpha\mathbf{v}\rangle_\alpha + \nabla\cdot\sum_\alpha \langle m_\alpha\mathbf{vv}\rangle_\alpha$$

$$-\sum_\alpha e_\alpha(\mathbf{E}\langle 1\rangle_\alpha + [\langle\mathbf{v}\rangle_\alpha\mathbf{B}]) = 0. \tag{13}$$

Using Maxwell's equation, we transform the two last terms in Eq. (13)

$$\mathbf{E}\sum_\alpha e_\alpha\langle 1\rangle_\alpha = \varepsilon_0\mathbf{E}\nabla\cdot\mathbf{E},$$

$$\sum_\alpha [\langle e_\alpha\mathbf{v}\rangle_\alpha\mathbf{B}] = [[\nabla\mathbf{H}]\mathbf{B}] - \varepsilon_0\left[\frac{\partial\mathbf{E}}{\partial t}\mathbf{B}\right],$$

$$[\mathbf{E}[\nabla\mathbf{E}]] = \frac{1}{2}\nabla E^2 - \mathbf{E}\cdot\nabla\mathbf{E},$$

$$[[\nabla\mathbf{H}]\mathbf{B}] = -\frac{\mu_0}{2}\nabla H^2 + \mu\mathbf{H}\cdot\nabla\mathbf{H}.$$

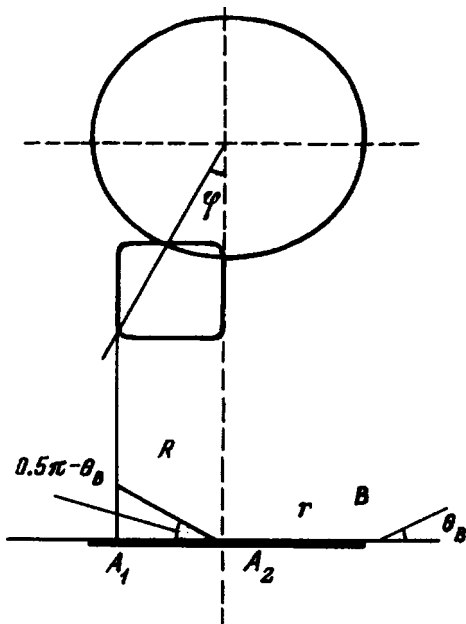We now rewrite Eq. (13) in the following form:

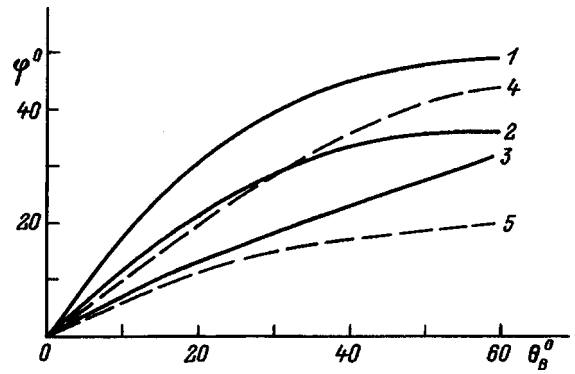FIG. 4. Diagram elucidating the origin of the Robson angle.

FIG. 5. Dependence of the Robson angle on the inclination angle of the external magnetic field relative to the cathode. *1*—Robson's experiment:[8] Al, 0.15 T, 10^{-1} Torr, 28 A; *2*—Kesaev's experiment:[1] Cu, 1 kOe, 10^{-3} Torr, 1–10 A; *3*—Juttner's experiment:[9] Mo, C, 0.37 T, 10^{-9} Torr, 10–300 A; *4*—formula (7) for $R \gg r$; *5*—formula (7) for $R \approx r$.

$$\frac{\partial}{\partial t}\left(\sum_\alpha \langle m_\alpha\mathbf{v}\rangle_\alpha + \varepsilon_0\mu_0[\mathbf{EH}]\right) + \nabla\cdot\sum_\alpha \langle m_\alpha\mathbf{vv}\rangle_\alpha$$

$$-\varepsilon_0\mathbf{E}\nabla\cdot\mathbf{E} + \frac{\mu_0}{2}\nabla H^2 - \mu_0\mathbf{H}\cdot\nabla\mathbf{H}$$

$$+\frac{\varepsilon_0}{2}\nabla E^2 - \varepsilon_0\mathbf{E}\cdot\nabla\mathbf{E} - \mu_0\mathbf{H}\nabla\cdot\mathbf{H} = 0. \tag{14}$$

The last term in Eq. (14) has been added for symmetry [to accommodate Eq. (12)]. According to the rules of vector calculus, we can write

$$\mathbf{A}\nabla\cdot\mathbf{A} - \frac{1}{2}\nabla A^2 + \mathbf{A}\cdot\nabla\mathbf{A} = \nabla\cdot\left(\mathbf{AA} - \frac{1}{2}A^2\hat{U}\right), \tag{15}$$

where the expression **AA** is a tensor (direct product of vectors) and $\hat{U}$ is the identity matrix.

Taking the above vector identity (15) into account, we obtain from Eq. (14) the equation of momentum conservation of the particles and field

$$\frac{\partial}{\partial t}\left\{\sum_\alpha \langle m_\alpha\mathbf{v}\rangle_\alpha + \varepsilon_0\mu_0[\mathbf{EH}]\right\} + \nabla\cdot\left\{\sum_\alpha \langle m_\alpha\mathbf{vv}\rangle_\alpha\right.$$

$$\left. -\left(\varepsilon_0\mathbf{EE} + \mu_0\mathbf{HH} - \frac{\varepsilon_0 E^2 + \mu_0 H^2}{2}\hat{U}\right)\right\} = 0. \tag{16}$$

In an analogous way we obtain the equation of energy conservation. Toward this end we multiply Eq. (1) by $(m_\alpha\mathbf{v}^2)/2$, integrate it over **v**, and carry out the sum over α. We obtain

$$\frac{\partial}{\partial t}\sum_\alpha \left\langle\frac{m_\alpha\mathbf{v}^2}{2}\right\rangle_\alpha + \nabla\cdot\sum_\alpha \left\langle\frac{m_\alpha\mathbf{v}^2}{2}\mathbf{v}\right\rangle_\alpha - \sum_\alpha \langle e_\alpha\mathbf{v}\cdot\mathbf{E}\rangle_\alpha = 0. \tag{17}$$

From Eq. (10)

$$\sum_\alpha \langle e_\alpha\mathbf{v}\cdot\mathbf{E}\rangle_\alpha = [\nabla\mathbf{H}]\cdot\mathbf{E} - \varepsilon_0\frac{\partial\mathbf{E}}{\partial t}\mathbf{E}. \tag{18}$$

Carrying out the vector product in Eq. (18), we have

$$[\nabla\mathbf{H}]\cdot\mathbf{E} = -\nabla\cdot[\mathbf{EH}] - \mu_0\mathbf{H}\frac{\partial\mathbf{H}}{\partial t}. \tag{19}$$

Employing Eqs. (18) and (19), we write the equation of energy conservation of the particles and field in the form

$$\frac{\partial}{\partial t}\left(\sum_\alpha \left\langle \frac{m_\alpha \mathbf{v}^2}{2} \right\rangle_\alpha + \frac{\varepsilon_0 E^2 + \mu_0 H^2}{2} \right)$$

$$+ \nabla \cdot \left(\sum_\alpha \left\langle \frac{m_\alpha \mathbf{v}^2}{2} \mathbf{v} \right\rangle_\alpha + [\mathbf{EH}] \right) = 0. \tag{20}$$

Note the following. The particles, electrons and ions, cross the space-charge sheath over times close to their respective inverse plasma frequencies. At the plasma densities near the center, $n \geqslant 10^{24}$ m^{-3}, these times are very short in comparison with the characteristic lifetime of an explosive emission center $\tau \approx 10^{-8}$ s (Ref. 10).

The dying out of a center is accompanied by a rapid drop of the current and the appearance of induced electric and magnetic fields. The correction to the fields associated with this effect is given by $r/(\tau_1 c)$, where r is the radius of the crater, τ_1 is the characteristic time of current drop, and c is the speed of light.[7] The corrections are substantial if $r/(\tau c) \geqslant 1$. In our case $r \approx 10^{-6}$ m, $\tau \approx 10^{-9}$ s (Ref. 10), and $r/(\tau c) < 10^{-5} \ll 1$, i.e., it is possible to ignore the variation of the fields associated with variation of the current and use only the time-independent forms of Eqs. (16) and (20) and of Maxwell's equations.

The particles in the cathode sheath are not magnetized; however, in order to use the one-dimensional Vlasov equations it is still necessary to show that the tangential component of the electric field, associated with the Ohmic potential drop at the cathode and in the near-cathode plasma due to the arc current, is small. The potential drop at the cathode was taken into account numerically within the framework of the model described in Ref. 11. The calculations show that: a) the potential drop is almost purely Ohmic, and thermoelectric effects are small, b) the potential drop can reach a value $\leqslant 10$ V, i.e., it is comparable with the cathode fall, and c) the potential drop is concentrated within $\leqslant 10^{-6}$ m of the emission zone.

The potential distribution in the near-cathode plasma was considered in Ref. 4. It was shown there that it also has an Ohmic character and that the field is concentrated on a scale of the order of 10^{-6} m.

Thus, we can neglect the tangential (to the cathode) component of the electric field in the cathode sheath and use the one-dimensional time-independent Vlasov equations to calculate the particle characteristics. We integrate the time-independent form of Eq. (16) over the region indicated in Fig. 6. It can be assumed that the particle velocities and the electric field in the sheath have only one component, parallel to the symmetry axis (Fig. 1). The volume integral over the hatched region in Fig. 6 is transformed into an integral over its surface. By virtue of the cylindrical symmetry of the system, integration over the end-faces yields the relation

$$2\pi \int_{r-\Delta r}^{r} r' dr' \left(\sum_\alpha \langle m_\alpha \mathbf{v}^2 \rangle_\alpha - \frac{\varepsilon_0 E^2}{2} \right)\Bigg|_K^P = 0, \tag{21}$$

since the magnetic field strength does not depend on the z coordinate, which runs perpendicular to the cathode. The su-
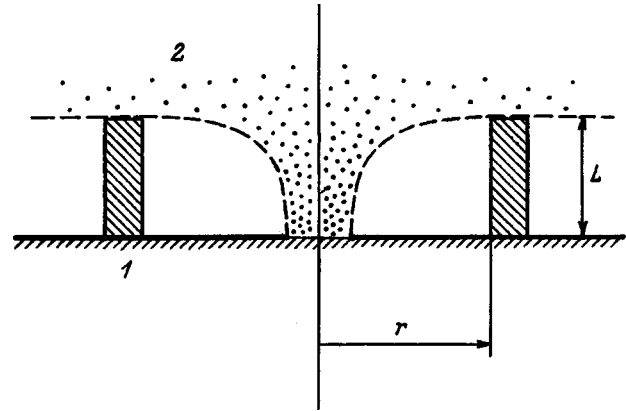

FIG. 6. Diagram of an emission center and the sheath layer separating the cathode (*1*) from the near-cathode plasma (*2*). The hatched regions represent a cross section of the cylindrical cavity over which equations (16) and (20) are to be integrated.

perscript P and subscript K together indicate the flux difference at the boundaries of the sheath with the cathode and with the plasma. Since the quantity Δr is arbitrary, we finally obtain

$$\left(\sum_\alpha \langle m_\alpha \mathbf{v}^2 \rangle_\alpha - \frac{\varepsilon_0 E^2}{2} \right)\Bigg|_K^P = 0. \tag{22}$$

Integration of Eq. (20) over the same region gives the following expression:

$$r\Delta r \left(\sum_\alpha \left\langle \frac{m_\alpha \mathbf{v}^2}{2} \mathbf{v} \right\rangle_\alpha \right)\Bigg|_K^P + r_2$$

$$\times \int_0^L [\mathbf{EH}]_2 dz - r_1 \int_0^L [\mathbf{EH}]_1 dz = 0, \tag{23}$$

where L is the thickness of the sheath, and the subscripts 1 and 2 correspond to the value of the function at the inner and outer radius of the hatched region in Fig. 6.

Taking $\Delta r = r_1 - r_2$ to be sufficiently small, we can write

$$r_2 \int_0^L [\mathbf{EH}]_2 dz - r_1 \int_0^L [\mathbf{EH}]_1 dz \approx \Delta r \int_0^L [\mathbf{EH}] dz.$$

Thus, Eq. (23) can be rewritten as follows:

$$\left(\sum_\alpha \left\langle \frac{m_\alpha \mathbf{v}^2}{2} \mathbf{v} \right\rangle_\alpha \right)\Bigg|_K^P + \frac{1}{r} \int_0^L [\mathbf{EH}] dz = 0. \tag{24}$$

Relation (21) is a consequence of balance of the momentum flux through the end-faces of the cylindrical shell in Fig. 6. It allows one to determine the field strength at the cathode without solving Poisson's equation in the sheath if the particle characteristics are known.

Relation (24) shows that the energy flux transported by the particles normal to the cathode surface increases with increasing energy flux transported along the sheath by the field. Hence the possibility opens up of a new interpretation of retrograde motion of the cathode spot. It is well known that the retrograde motion takes place in the direction of the

maximum magnetic field. In this direction the vector $[\mathbf{EH}]$ is also maximum and, consequently, the normal particle flux toward the cathode has a large value, which in turn leads to a greater probability of appearance of new centers in the direction of the retrograde motion. The additional acceleration of the particles in the sheath has an electrodynamic character.

CONCLUSION

Thus, we can claim reasonable agreement of the experimental data with calculations based on our quite simple model. A second mechanism, associated with the electrodynamic character of the acceleration of the particles in the sheath layer separating the plasma from the cathode probably plays a large role under dirty conditions, when the craters are widely separated from one another.

[1] I. G. Kesaev, *Cathode Processes of an Electric Arc* [in Russian] (Nauka, Moscow, 1968).

[2] G. A. Mesyats and D. I. Proskurovskiǐ, *Pulsed Electrical Discharge* [in Russian] (Nauka, Novosibirsk, 1984).

[3] G. A. Mesyats, Usp. Fiz. Nauk **165**, No. 6, 601 (1995).

[4] E. A. Litvinov, G. A. Mesyats, and A. G. Parfenov, Dokl. Akad. Nauk SSSR **310**, 344 (1990) [Sov. Phys. Dokl. **35**, 47 (1990)].

[5] E. A. Litvinov, G. A. Mesyats, A. G. Parfenov, and A. I. Fedosov, Zh. Tekh. Fiz. **55**, 2270 (1985) [Sov. Phys. Tech. Phys. **30**, 1346 (1985)].

[6] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Vol. 5, *Electricity and Magnetism* (Addison-Wesley, Reading, Mass., 1965; Mir, Moscow, 1977).

[7] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Vol. 6, *Electrodynamics* (Addison-Wesley, Reading, Mass., 1965; Mir, Moscow, 1977).

[8] A. E. Robson, in *Proceedings of the IVth International Conference on Phenomena in Ionized Gases*, Uppsala, 1959, Vol. II b, pp. 346–349.

[9] B. Juttner, Thesis B. Academy of Sciences. Berlin, 1983.

[10] E. A. Litvinov, G. A. Mesyats, and A. G. Parfenov, Dokl. Akad. Nauk SSSR **279**, 864 (1984) [Sov. Phys. Dokl. **29**, 1019 (1984)].

[11] S. A. Barengolts, E. A. Litvinov, A. G. Parfyonov, in *Proceedings of the XIVth International Symposium on Discharge and Electrical Insulation in Vacuum*, Santa Fe, New Mexico (1990), pp. 185–186.

Translated by Paul F. Schippnick

Onset of rotating stall in induction magnetohydrodynamic flows

Yu. A. Polovko, E. P. Romanova, and É. A. Tropp

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted June 3, 1997)
Zh. Tekh. Fiz. **68**, 65–70 (June 1998)

The possible onset of unsteady induction magnetohydrodynamic (MHD) flows in wide channels or in channels of annular cross section is discovered on the basis of a proposed two-dimensional mathematical model. Such secondary flows have the character of rotating stall, which was previously known in high-pressure axial compressors. The existing experimental data confirm the possibility of observing this phenomenon, which can be interpreted as a new type of symmetry loss. It is shown for certain relations between the parameters that the system has a lower margin of stability against disturbances of the rotating-stall type than against steady-state disturbances. In particular, a loss of stability of a steady uniform flow can occur on the descending portion of the external characteristic of the machine. © *1998 American Institute of Physics.* [S1063-7842(98)01306-3]

INTRODUCTION

The loss of stability of uniform induction flows in magnetohydrodynamic (MHD) channels and the properties of the secondary flows that arise have hitherto been studied mainly on the basis of the so-called one-dimensional jet flow model, which was first used for this purpose in Ref. 1. The main results were described in detail in Refs. 2–5. At the same time, transverse velocities, which lead to the development of helical motion, were discovered in experiments devised to study the structure of the flows in channels of induction MHD machines.[6,7] As was stressed in Ref. 8, the reason for the observation of the transverse velocities could not be established unequivocally. This is because there was asymmetry of the primary field with respect to the azimuth (the transverse coordinate) in the experiments performed due to the engineering constraints. The question of whether the transverse motions are caused by this asymmetry or appear as a result of the instability of the symmetric problem remained open, and an investigation of the stability of at least two-dimensional motion was needed to resolve it. Such an attempt was apparently undertaken for the first time in Ref. 9, but the status of the results obtained is not entirely clear, since the stability of a uniform flow was studied in that work only with respect to a narrow class of disturbances, whose wavelength was equal to or a rational fraction of the wavelength of the external traveling field.

The study of the properties of steady secondary flows in Refs. 4 and 5 demonstrated their remarkable similarity to the secondary flows in high-pressure axial compressors, which appear as a result of so-called "rotating stall." [10] This similarity encompasses both external integral characteristics (the presence of an extended horizontal segment on them and operation in an "ideal pressure source" regime[5]) and the flow structure (the existence of internal boundary layers separating a "cell," in which the stream has a velocity differing strongly from the flow nucleus and is possibly even oppositely directed). Such similarity points out the similarity

between the properties of the mathematical models of their different physical phenomena. At the same time, the secondary flows in compressors are unsteady, have a helical character, and the velocity cell formed moves across the width of the channel. The observed similarity between the phenomena provides some basis to assume that flows of such structure can also exist in induction MHD channels. In the present work it is shown on the basis of a proposed two-dimensional model in the long-wavelength approximation that instability of the rotating-stall type can be observed in induction MHD flows. A linear analysis, which permits drawing several conclusions regarding the character of its appearance as a function of the of the parameters of the problem, is performed. The previously studied, steady secondary flows are also described by the two-dimensional system used as special cases of possible solutions for definite relations between the parameters.

MATHEMATICAL MODEL

The flow of the conducting fluid in an induction MHD channel, whose scheme has been presented, for example, in Ref. 4, is investigated. Here and below the notation and directions of the coordinate axes correspond to those adopted in Refs. 4 and 5. We shall use the so-called plane-parallel field model (Ref. 11, p. 155), assuming that there is only one magnetic field component normal to the channel wall,

$$\mathbf{B}=(0,0,\tilde{\mathbf{b}}(x,y,t)\exp[i(\alpha x-\Omega t)]),$$

$$\tilde{\mathbf{b}}(x,y,t)=b_a+ib_r, \tag{1}$$

which corresponds to the fundamental mode of the external electromagnetic field traveling along the channel axis $B_0 = B_m \sin(\alpha x - \Omega t)$. This approximation is natural for channels of small height with walls having a high magnetic permeability. Unlike the so-called jet model, despite the axial symmetry of the machine, we allow the possibility of the existence of an azimuthal velocity component, i.e., we set $\mathbf{V}=(u,v,0)$. Substituting (1) into the induction equation

$$\Delta \mathbf{B} - \mu\sigma\frac{\partial \mathbf{B}}{dt} + \mu\sigma\ \text{curl}(\mathbf{V}\times\mathbf{B}) = -\text{curl}\ j_0, \qquad (2)$$

(δ is the Laplacian operator), we obtain for the relative value of the complex amplitude of the magnetic induction $\mathbf{b} = \widetilde{\mathbf{b}}/B_m$

$$[\mathbf{A}_1 + \mathbf{A}_2]\mathbf{b} = i,$$

$$\mathbf{A}_1 = \frac{\partial^2}{\partial y^2} - \varepsilon\frac{\partial}{\partial t} - 1 + i\varepsilon(1-u),$$

$$\mathbf{A}_2 = \frac{\partial^2}{\partial x^2} + 2i\frac{\partial}{\partial x} - \varepsilon u\frac{\partial}{\partial x} - \varepsilon v\frac{\partial}{\partial y}, \qquad (3)$$

where $\varepsilon = \mu\sigma\Omega$ is the magnetic Reynolds number, $\mathbf{A}_1$ is the operator corresponding to the jet flow model, and $\mathbf{A}_2$ is an additional operator which is conditional on the presence of an azimuthal component of the velocity $\mathbf{V}$ and the possibility of that the field amplitude can vary along the x axis.

We note that in (2) j_0 is the extrinsic current density

$$\text{curl}\ j_0 = -\frac{\partial^2 B_0}{\partial x^2} = -\alpha^2 B_m\ i\ \exp[i(\alpha x - \Omega t)].$$

The equations of motion are averaged with respect to the period and wavelength of the external traveling electromagnetic field. After application of the averaging operator

$$\frac{1}{T}\int_0^T \bullet dt\frac{1}{2\tau}\int_0^{2\tau}\bullet dx,$$

where $T = 2\pi/\Omega$ and $\tau = \pi/\alpha$, the equations of motion for the principal velocity component, which varies slowly in comparison to the variation of the external field, take the form

$$z\left(\frac{\partial u}{\partial \tau} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}\right)$$

$$= -\frac{\partial p}{\partial x} + \frac{j^2}{\varepsilon}\left(b_a - b_a\frac{\partial b_a}{\partial x} - b_r\frac{\partial b_r}{\partial x}\right) - u\sqrt{v^2 + u^2},$$

$$z\left(\frac{\partial v}{\partial \tau} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y}\right)$$

$$= -\frac{\partial p}{\partial y} - \frac{j^2}{\varepsilon}\left(b_a\frac{\partial b_a}{\partial y} + b_r\frac{\partial b_r}{\partial y}\right) - v\sqrt{v^2 + u^2},$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0, \qquad (4)$$

where j is the relative current density due to the voltage supplied and z is the hydraulic inductance of the pump.

Here, as in the jet model, only the turbulent flow at the wall [the last terms of the first two equations in (4)] is taken into account in view of the small height of the channel.

Equations (3) and (4) are supplemented by the periodicity conditions

$$\mathbf{b}\big|_{x=0} = \mathbf{b}\big|_{x=L}, \quad \mathbf{b}\big|_{y=0} = \mathbf{b}\big|_{y=S},$$

$$\frac{\partial\mathbf{b}}{\partial x}\bigg|_{x=0} = \frac{\partial\mathbf{b}}{\partial x}\bigg|_{x=L}, \quad \frac{\partial\mathbf{b}}{\partial y}\bigg|_{y=0} = \frac{\partial\mathbf{b}}{\partial y}\bigg|_{y=S},$$

$$\mathbf{v}\big|_{x=0} = \mathbf{v}\big|_{x=L}, \quad \mathbf{v}\big|_{y=0} = \mathbf{v}\big|_{y=S}, \qquad (5)$$

which can be interpreted either as the conditions for a cylindrical induction MHD channel closed into a torus or as the natural boundary conditions for wide and long channels. Here L and S are the relative length and width of the channel, respectively. The velocity of the external traveling field $[v] = \Omega/\alpha$ was taken as the velocity scale in (3)–(5), and $[t] = 1/\Omega$ and $[b] = B_m$ were taken for the time and the magnetic induction, respectively. We note that the applicability of the model (3)–(5) just formulated is restricted by the averaging procedure used, which requires weak variation of $\mathbf{b}$ and $\mathbf{v}$ over lengths of the order of 2τ and times of the order of T, i.e., the problem is solved in the ''long-wavelength'' approximation.

LINEAR ANALYSIS. SYSTEM OF EQUATIONS SPECIFYING NEUTRAL SURFACES

The problem (3)–(5) admits a solution corresponding to a steady flow that is uniform with respect to both coordinates:

$$u = u_0, \quad v_0 = 0, \quad b_{a0} = \frac{\varepsilon(1-u_0)}{\varepsilon^2(1-u_0)^2 + 1},$$

$$b_{r0} = -\frac{1}{\varepsilon^2(1-u_0)^2 + 1},$$

$$\left(\frac{\partial p}{\partial x}\right)_0 = \frac{j^2(1-u_0)}{\varepsilon^2(1-u_0)^2 + 1} - u_0|u_0|. \qquad (6)$$

We seek solutions of the form

$$\mathbf{U} = \mathbf{U}_0 + \delta\mathbf{U} = \mathbf{U}_0 + \mathbf{C}\ \exp[i(2\pi my/S + 2\pi nx/L - \omega t)], \qquad (7)$$

where

$$\mathbf{U} = \begin{pmatrix} u \\ v \\ b_a \\ b_r \\ p \end{pmatrix}, \quad \mathbf{U}_0 = \begin{pmatrix} u_0 \\ v_0 \\ b_{a0} \\ b_{r0} \\ p_0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix},$$

which branches off from the uniform solution (6). After linearizing the problem (3)–(5) in the vicinity of (6) and plugging in (7), we obtain a system of linear equations, whose determinant equals

$$\Delta = |D_{ik}|,$$

where

$$D_{11} = i\varepsilon\omega - k_1^2 - k_2^2 - 1 - i\varepsilon k_2 u_0,$$

$$D_{12} = -\varepsilon(1-u_0) - 2ik_2,$$

$$D_{13} = \varepsilon b_{r0}, \quad D_{14} = 0, \quad D_{15} = 0,$$

$$D_{21} = \varepsilon(1-u_0) + 2ik_2,$$

$$D_{22} = i\varepsilon\omega - k_1^2 - k_2^2 - 1 - i\varepsilon k_2 u_0,$$

$$D_{23} = -\varepsilon b_{a0}, \quad D_{24} = 0, \quad D_{25} = 0,$$

$$D_{31}=-\frac{j^2}{\varepsilon}(1-ib_{a0}k_2),\quad D_{32}=\frac{ij^2}{\varepsilon}b_{r0}k_2,$$

$$D_{33}=zu_0k_2i+2|u_0|-i\omega z,\quad D_{34}=0,\quad D_{35}=ik_2,$$

$$D_{41}=\frac{ij^2}{\varepsilon}k_1b_{a0},\quad D_{42}=\frac{ij^2}{\varepsilon}b_{r0}k_1,\quad D_{43}=0,$$

$$D_{44}=izu_0k_2+|u_0|-i\omega z,\quad D_{45}=ik_1,$$

$$D_{51}=0,\quad D_{52}=0,\quad D_{53}=ik_2,\quad D_{54}=ik_1,\quad D_{55}=0.$$

Here we have introduced the following new notation: $k_1=2\pi m/S$ and $k_2=2\pi n/L$ are wave numbers, which express the relative density of the waves appearing along the width and length of the channel, respectively. The condition for the existence of nontrivial solutions leads to a system of two nonlinear equations

$$\text{Re}(\Delta)=0,\quad \text{Im}(\Delta)=0,$$

from which the branch point u_0 and the stall rotation frequency ω are determined. The expanded form of this system is

$$R_3|u_0|(R_1+k_1^2)-zR_1R_2R_4-k_1^2j^2[b_{a0}\varepsilon(1-u_0)$$

$$+b_{r0}(R_1+1)]=0,$$

$$R_4|u_0|(R_1+k_1^2)+zR_1R_2R_3-k_1^2j^2[2b_{a0}k_2+\varepsilon b_{r0}R_2]=0,$$
$$(8)$$

where

$$R_1=k_1^2+k_2^2,\quad R_2=k_2u_0-\omega,$$

$$R_3=(R_1+1)^2+\varepsilon^2(1-u_0)^2-\varepsilon^2R_2^2-4k_2^2,$$

$$R_4=2\varepsilon[R_2(R_1+1)+2(1-u_0)k_2].$$

ANALYSIS OF THE SYSTEM SPECIFYING THE BRANCH POINT

As we have already noted, the mathematical model adopted is applicable only in the case of $k_2\ll1$ and $\omega\ll\Omega$ because of the procedure used to average the equations of motion. In this context the form of the asymptotic formula for ω and u_0 at small values of k_2 is of interest. After substituting the series

$$\omega=\omega_0+\omega_1k_2+O(k_2),\quad u_0=u_0^1+u_0^{(2)}k_2^2+O(k_2^2)$$

into (8) and equating the terms with identical powers of k_2, in a first approximation we obtain

$$\omega_0=0,\quad 2|u_0^1|G(G+K^2-1)-j^2(G-K-1)=0,\quad(9)$$

where $K=1+k_1^2$, $G=1+\varepsilon^2s^2$, and $s=1-u_0^1$.

Expression (9) is the familiar equation for determining the branch points of stationary solutions.[1] Thus, the previously investigated steady flows comprise a subset of the solutions of the two-dimensional model, whose branch points are determined from (8) when $k_2=0$. Equating the terms of order k_2 in the second equation and of order k_2^2 in the first equation gives the following expressions:

$$\omega_1=\frac{Gu_0^1[4\varepsilon(Ku_0^1+2s)+z(K^2+\varepsilon^2s^2)]-\varepsilon j^2(2s-u_0^1)}{G[4\varepsilon u_0^1K+z(K^2+\varepsilon^2s^2)]+\varepsilon j^2},$$

$$u_0^{(2)}=-\frac{G[(K^2+\varepsilon^2s^2)u_0^1+2k_1^2u_0^1(2K-\varepsilon^2W^2-4)-2z\varepsilon k_1^2W(KW+2s)]+k_1^2j^2}{2k_1^2[(K^2+\varepsilon^2s^2)(G-2u_0^1\varepsilon^2s)-2Gu_0^1\varepsilon^2s+j^2\varepsilon^2s]},$$

where $W=u_0^1-\omega_1$. It is seen that the phase velocity for the propagation of long waves along the channel $c=\omega/k_2=\omega_1$ does not depend on the wavelength.

The case of large currents due to the voltage supplied $(j\to\infty)$ is also of practical interest. The asymptotic expression then has the simple form

$$u_0=1-\frac{\sqrt{k_1^2+k_2^2+1}}{\varepsilon}+\delta(j),\quad \omega=k_2(3u_0-2)+\delta(j),$$

where $\delta(j)\to0$ and $j\to\infty$.

The system (8) was solved numerically by Newton's method in the general case. The calculations were performed mainly for the pump operating regime of an MHD machine $0<u_0<1$. The problem of flow in a channel closed into a torus is apparently artificial and is not of practical interest; therefore, we at once turn to an analysis of flow in an infinite tube of annular cross section. This corresponds to the continuous variation of k_2 and the discrete set of values $k_1=2\pi n/L=\sqrt{\varkappa n}$, $n=1,2,3...$. Figure 1 shows the neutral curves, i.e., plots of the dependence of u_0 at the branch point

on k_2^2 for various values of z and the mode number n. The nonmonotonic character of the $u_0(k_2^2)$ curves appearing for fairly large values of z is of fundamental importance for the analysis. It means that the region for the existence of flows of the rotating-stall type can be wider than the region for the existence of steady secondary flows, for which $k_2=0$. Unsteady flows should be observed experimentally for such pa-

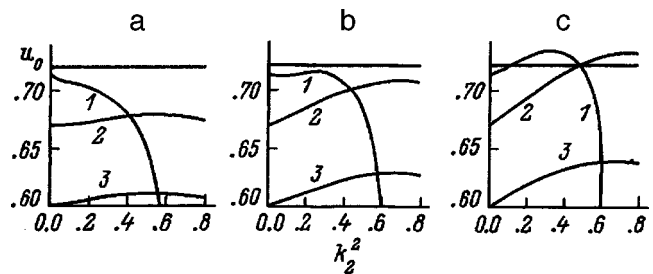
FIG. 1. Dependence of u_0 on the square of the wave number k_2^2 for $\varepsilon=4$, $j^2=40$, $\sqrt{\varkappa}=0.1$, and $z=5$ (a), 10 (b), and 20 (c) in the pump regime for various azimuthal mode numbers $n=1$ (*1*), 2 (*2*), and 3 (*3*).
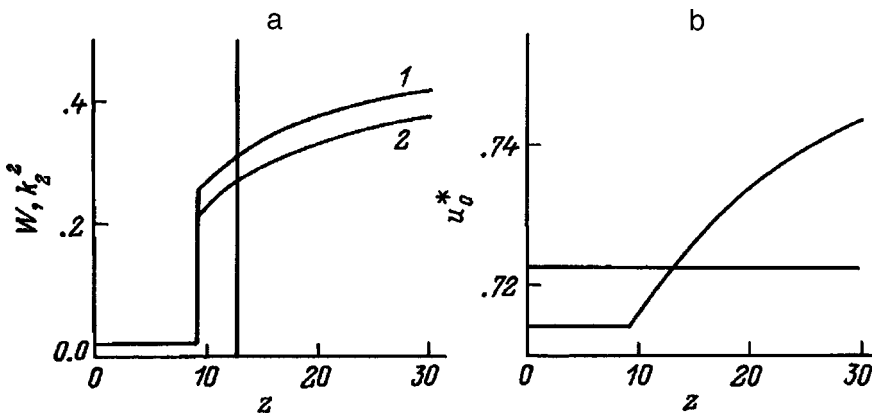
FIG. 2. Critical value of the angular frequency ω (*1*a), the square of the wave number k_2^2 (*2*a), and u_0^* (b) as functions of the hydraulic inductance z for $k_1 = 0.1$, $\varepsilon = 4$, and $j^2 = 40$. The plot was constructed for the region of maximum pressure values in the pump regime.

rameters, since the stability reserve of the system toward such disturbances is minimal.

Figure 2 shows the dependences of the critical values $u^* = \max(u_0)$, ω, and k_2^2 on z. The vertical line in Fig. 2a and the horizontal line in Fig. 2b correspond to the highest value of u_0 possible for the steady uniform flow (6) in the pump regime. The intersection of these lines by the $u^*(z)$, $\omega(z)$, and $k_2^2(z)$ curves attests to the possibility of a loss of stability on the descending branches of the external characteristic. Such behavior is impossible for steady secondary flows.[2–4] It is known that branching external characteristics which are smoother than those predicted by the one-dimensional theory are observed experimentally. They do not exhibit a characteristic peak in the vicinity of the branch point. The possibility of the branching of a static external characteristic on the descending portion can lead to smoothing of that peak. This becomes conclusively clear after a nonlinear analysis and complete numerical determination of the external characteristics of the unsteady secondary flows.

The limiting case of flow between parallel plates, where both wave numbers k_1 and k_2 vary continuously, is also interesting. Figure 3 shows the neutral surface $u_0(k_1^2, k_2^2)$ corresponding to the zeroth damping decrement of the disturbances for this case as an example. It is seen that it has a maximum corresponding to the critical point for the loss of stability for the parameters taken. The instability appears on the descending portion of the external characteristic, but the fairly large values of k_2^2 (0.4–0.7) at the critical point force us to question the applicability of the model used for studying the stability of flow between plates, at least for quantitative calculations. A rigorous analysis calls for studying the stability of a system which has not been averaged with respect to time and the wavelength. This leads to the need to solve the far more complex eigenvalue problem for equations with variable coefficients.

We note that axial symmetry leads to equivalence of the stall rotation directions, i.e., small disturbances with either the azimuthal wave number k_1 or $-k_1$ are always possible. However, as is seen directly from system (8), the sign of k_2 does not vary, i.e., the direction of motion of a wave traveling along the x axis always coincides with the direction of motion of the external field.

EVOLUTION OF FINITE DISTURBANCES

The results obtained demonstrate the possibility, in principle, of the appearance of secondary unsteady induction MHD flows of the rotating-stall type. However, at least two important questions can be solved only as a result of the consideration of finite, but small disturbances. It is not clear, first, whether the instability studied can lead to the twisting of an axisymmetric flow and, second whether the waves (7) appearing can quench one another and lead to a wave which is standing in the azimuthal direction, rather than to a "traveling cell." In fact, as follows from a linear analysis, after the loss of stability at the critical point, the main part of a disturbance of the uniform solution will have the form

$$\delta \mathbf{U} = \eta j_1 \mathbf{U}_1 \, \exp[i(k_2 x + k_1 y - \omega t)] + \eta j_2 \mathbf{U}_1$$

$$\times \exp[i(k_2 x - k_1 y - \omega t)] + \text{c.c.}(O(\varepsilon)), \tag{10}$$

where c.c. denotes the complex-conjugate terms.

The amplitudes j_1 and j_2 are determined from the conditions of solvability of the equations for the next terms in the expansion of $\delta \mathbf{U}$ into a series in the supercriticality $\eta^2 = u_0^* - u_0$. These conditions from a so-called system of branching equations, which has the same form as in the case
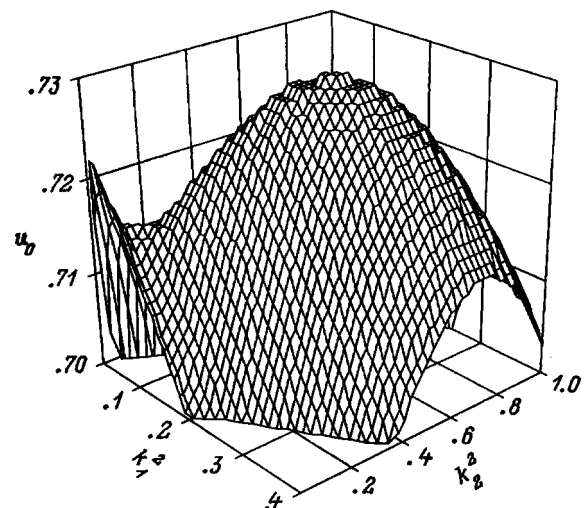

FIG. 3. The neutral surface $u_0(k_1^2, k_2^2)$ for $z = 11.3$, $\varepsilon = 4$, and $j^2 = 40$.

of the loss of stability of a hot jet.[12] As follows from the basic principles of symmetry and is indicated by the preliminary analysis performed, the system of branching equations has solutions of the form

$$j_1 = 0, \quad |j|_2^2 = j^2, \tag{11}$$

$$j_2 = 0, \quad |j|_1^2 = j^2, \tag{12}$$

which correspond to the generation of rotating stall with cells moving in opposite directions, and the solution

$$j_1^2 = j_2^2 = \beta^2, \tag{13}$$

which corresponds to a wave which is standing in the azimuthal direction and traveling in the longitudinal direction. The helical waves corresponding to solutions (11) and (12) lead to deformation and twisting of the mean flow already in the next approximation with respect to ε using the Reynolds directions $(\mathbf{u}_1 \cdot \nabla)\mathbf{u}_1$ and nonlinear friction in (4). An analysis of the stability of these solutions near the neutral curve reveals which of the two regimes, i.e., (11) and (12) with twisting or (13) without twisting, is realized. The stability of the solutions (11)–(13) will depend on the relation between j and β, which, in turn, can be defined in terms of the parameters of the original system (3) and (4). This analysis is our next subject of investigation.

We note that the steady flows generated by the evolution of the perturbations (7) for $k_2 = 0$ and their external characteristics were constructed completely in Ref. 2 and that their stability was demonstrated in Ref. 3 within the one-dimensional jet model. However, it is not known whether they are stable in the two-dimensional model (3)–(5). Displacement of the branch points of the nonstationary solutions to the right, into the region of negative slope of the static external characteristic for a uniform flow, leads to the fact that some stationary solutions of (4) can be totally absent over a fairly broad range of variation of the external load after the loss of stability of the uniform flow.

CONCLUSIONS

In summary, we formulate the main conclusions following from the analysis performed.

1. Twisting helical flows of the traveling-wave type can appear along with the previously studied static secondary flows in induction MHD channels.

2. The appearance of secondary flows of such a kind has several similarities to so-called rotating stall, which has previously been described for high-pressure compressors.

3. At fairly high values of the hydrodynamic inductance, the region for the existence of such twisting flows can be wider than that of steady secondary flows.

4. At fairly high values of the twisting hydrodynamic inductance, flows can also appear on the descending branches of static external characteristics in the pump and generator regimes, but this is ruled out for steady secondary flows.

5. In all likelihood, the analysis performed accounts for the previously observed twisting of the flows in the channels of cylindrical induction MHD machines and the smoothing of the external characteristics of the secondary flows in comparison to the flows predicted by the one-dimensional jet model.

6. The further study of rotating stall in induction MHD flows will involve a nonlinear analysis in the vicinity of the neutral curve and numerical construction of nonstationary solutions and the corresponding external characteristics far from the branch point. In addition, a refinement of the proposed mathematical model, which would eliminate the constraints on the density and velocity of the traveling waves caused by the procedure used to average the equations of motion, is possible.

[1] A. K. Gaĭlitis and O. A. Lielausis, Magn. Gidrodin. No. **1**, 106 (1975).

[2] Yu. A. Polovko and É. A. Tropp, Magn. Gidrodin. No. **4**, 106 (1986).

[3] Yu. A. Polovko, Magn. Gidrodin. No. **3**, 81 (1989).

[4] Yu. A. Polovko, E. P. Romanova, and É. A. Tropp, Zh. Tekh. Fiz. **66**(4), 36 (1996) [Tech. Phys. **41**, 315 (1996)].

[5] Yu. A. Polovko, E. P. Romanova, and É. A. Tropp, Zh. Tekh. Fiz. **67**(6), 5 (1997) [Tech. Phys. **42**, 591 (1997)].

[6] A. O. Klavinya, O. A. Lielausis, and V. V. Riekstin'sh, ''Nonuniform flow in the channel of a cylindrical pump,'' in *Abstracts of the 8th Riga Conference on Magnetohydrodynamics* [in Russian], Riga (1975), Vol. 2, pp. 79–81.

[7] I. R. Kirillov, A. P. Ogorodnikov, and V. P. Ostapenko, Magn. Gidrodin. No. **2**, 107 (1980).

[8] R. A. Valdmane, Ya. Ya. Valdmanis, and L. Ya. Ulmanis, Magn. Gidrodin. No. **1**, 103 (1986).

[9] S. Yu. Reutskiĭ, Magn. Gidrodin. No. 1, 121 (1987).

[10] E. M. Greitzer, Trans. Am. Soc. Mech. Eng. **103**, 193 (1980).

[11] A. I. Vol'dek, *Induction Magnetohydrodynamic Machines with a Liquid-Metal Working Fluid* [in Russian], Énergiya, Leningrad (1970), 271 pp.

[12] E. M. Zhdanova, in *Stability and Turbulence* [in Russian], M. A. Gol'dshtik and V. N. Shtern (Eds.), Novosibirsk (1985), pp. 71–81.

Translated by P. Shelnitz

Stability of the normal phase in a bounded current-carrying superconducting film

A. S. Rudyĭ

Yaroslavl State University, 150000 Yaroslavl, Russia
(Submitted July 30, 1996; resubmitted June 27, 1997)
Zh. Tekh. Fiz. **68**, 71–77 (June 1998)

The two-phase equilibrium states of a current-carrying thin superconducting film in the case of convective heat transfer on the free surface are considered, and their stability is investigated in a first approximation. It is shown that of the two equilibrium states, the state with the normal-phase region of larger size is stable. In the limiting case of an infinitely long film, the stable two-phase equilibrium state tends to a spatially uniform normal state, and the unstable state remains localized. In a definite range of values of the system parameters, the relaxation time of such a formation can be fairly long, and it should be regarded as a quasistable equilibrium state. © *1998 American Institute of Physics.* [S1063-7842(98)01406-8]

INTRODUCTION

One of the problems associated with the development and use of cryoelectronic devices is thermal destruction of the superconducting state, which is accompanied by the formation of a normal phase in the superconducting region. In Ref. 1 such nonuniform equilibrium states were examined in reference to one-dimensional and planar structures as the most widely encountered elements in cryoelectronics. In particular, the following was demonstrated in Ref. 1, where an idealized model of a thin superconducting film carrying an alternating current and immersed in a cooling medium was investigated. At values of the Stekly parameter σ exceeding the critical value σ_c, along with the uniform superconducting state there are nonuniform equilibrium states, in which the central part of the film is in the normal state. Because of the so-called external nonlinearity of the system (the discontinuity of the parameters and the source density on the phase boundary), at $\sigma > \sigma_c$ there are two such double-fronted nonuniform equilibrium states.

In situations where the existence of some element of a cryoelectronic device in a nonuniform state is a necessary condition for its operation, as, for example, in Franzen's bolometer,[2] the problem of the stability of a localized normal state arises, which was not investigated in Ref. 1. The present work is devoted to an analysis of the stability of the stationary solutions obtained in Ref. 1 against symmetric perturbations of the temperature field and an investigation of the asymptotic behavior of the two equilibrium states.

DYNAMICAL MODEL OF THE SYSTEM

We consider the thin superconducting film carrying an alternating current in Fig. 1, whose central portion is in the normal state. Let the film be immersed in a thermostat filled with a liquid or a gas, and let the entire system, with the exception of the film, have a temperature below the critical value and be in thermal equilibrium. We examine the case in which the nonuniform temperature field of the film is one-dimensional. We introduce the following notation for the deviation of the temperature of the superconducting film from the thermostat temperature: $T_s(x,t)$ is the deviation of the temperature in the superconducting region; $T_n(x,t)$ is the deviation of the temperature in the region of the normal phase. It is clear that the stationary distribution of the temperature is symmetric relative to the plane passing through the middle of the film $x = \delta$. Let the system deviate from the equilibrium state in the initial moment so that the temperature distribution remains symmetric, as before. In this case the boundary-value problem for determining the temperature field of the superconducting (S) and normal (N) phases has the form
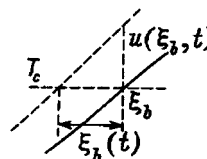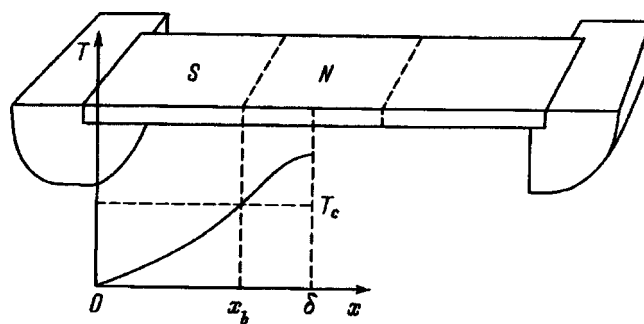

FIG. 1. Superconducting film in the circuit of an ac source with convective heat transfer on the free surface. The deviation of the film temperature in the two-phase state from the thermostat temperature is shown. Inset—displacement of the phase boundary upon perturbation of the temperature field.

$$c_{Vs}\dot{T}_s = \lambda_s T_s'' - 2\frac{\alpha}{h}T_s,$$

$$c_{Vn}\dot{T}_n = \lambda_n T_n'' + \rho_0(1+\beta T_n)\overline{j^2} - 2\frac{\alpha}{h}T_n,$$

$$T_s(0,t)=0, \quad T_n'(x,t)|_{x=\delta}=0,$$

$$T_s[x_b(t),t]=T_n[x_b(t),t]=T_c,$$

$$\lambda_s T_s'(x,t)|_{x=x_b(t)} = \lambda_n T_n'(x,t)|_{x=x_b(t)}. \quad (1)$$

Here c_{Vs}, c_{Vn}, λ_s, and λ_n are the heat capacities per unit volume and the thermal conductivities of the superconducting and normal phases, respectively; ρ_0 is the resistance of the normal phase; β is the temperature coefficient of resistance; $\overline{j^2}$ is the period-averaged value of the square of the current density; α is the heat transfer coefficient, h is the thickness of the film, x_b is the coordinate of the phase boundary, and T_c is the critical temperature. The last terms on the right-hand sides of Eqs. (1) take into account the heat loss due to convective heat transfer on the surface. In addition, in view of the small value of h, we assume that there is no thermal resistance or temperature gradient along a normal to the surface.

Going over to the normalized variables $\Theta = T/T_c$ and $\xi = x/\delta$, we bring the problem (1) into the dimensionless form

$$\frac{\delta^2}{a_s}\dot{\Theta}_s(\xi,t) = \Theta_s''(\xi,t) - 2\mathrm{Bi}_s\Theta_s(\xi,t),$$

$$\frac{\delta^2}{a_n}\dot{\Theta}_n(\xi,t) = \Theta_n''(\xi,t) - \kappa\Theta_n(\xi,t) + K,$$

$$\Theta_s(0,t)=0, \quad \Theta_n'(\xi,t)|_{\xi=1}=0,$$

$$\Theta_s[\xi_b(t),t]=\Theta_n[\xi_b(t),t]=1,$$

$$\lambda_s\Theta_s'[\xi,t]|_{\xi=\xi_b(t)} = \lambda_n\Theta_n'[\xi,t]|_{\xi=\xi_b(t)}. \quad (2)$$

Here $\mathrm{Bi}=\alpha\delta^2/h\lambda$ is the product of the intrinsic Biot number $\alpha\delta/\lambda$ and the similarity criterion of parametric form δ/h, and $\kappa=2\mathrm{Bi}_n - \beta\rho_0\overline{j^2}\delta^2/\lambda_n$ and $K=\rho_0\overline{j^2}\delta^2/\lambda_n T_c$ are auxiliary parameters. We seek the solution of the problem (2) in the form of a sum of the stationary and nonstationary solutions

$$\Theta(\xi,t)=\bar{\Theta}(\xi)+u(\xi,t), \quad \xi_b(t)=\bar{\xi}_b+\xi_b(t), \quad (3)$$

where $u(\xi,t)$ is a small symmetric perturbation of the equilibrium state:

$$u(\bar{\xi}_b,t)/\bar{\Theta}(\bar{\xi}_b) \ll 1. \quad (4)$$

It is not difficult to show (Fig. 1) that the displacement of the coordinate of the phase boundary $\xi_b(t)$ for such a perturbation of the temperature field will also be small. Then the continuity conditions of the problem (2), which are written preliminarily as

$$\Theta_s[\bar{\xi}_b+\xi_b(t)]+u_s[\bar{\xi}_b+\xi_b(t),t]$$

$$=\bar{\Theta}_n[\bar{\xi}_b+\xi_b(t)]+u_s[\bar{\xi}_b+\xi_b(t),t]=1,$$

$$\lambda_s\{\bar{\Theta}_s'[\bar{\xi}_b+\xi(t)]+u_s'[\bar{\xi}_b+\xi(t),t]\}$$

$$=\lambda_n\{\bar{\Theta}_n'[\bar{\xi}_b+\xi(t)]+u_s'[\bar{\xi}_b+\xi(t),t]\}, \quad (5)$$

can be expanded in a Taylor series in the vicinity of the point $\bar{\xi}_b$ with respect to the variable $\xi_b(t)$.[3] After discarding the terms which are nonlinear with respect to the perturbations, we obtain a linear approximation of the continuity conditions:

$$\bar{\Theta}_s(\bar{\xi}_b)+\bar{\Theta}_s'(\bar{\xi}_b)\xi_b(t)+u_s(\bar{\xi}_b,t)$$

$$=\bar{\Theta}_n(\bar{\xi}_b)+\bar{\Theta}_n'(\bar{\xi}_b)\xi_b(t)+u_n(\bar{\xi}_b,t)=1,$$

$$\lambda_s[\bar{\Theta}_s'(\bar{\xi}_b)+\bar{\Theta}_s''(\bar{\xi}_b)\xi_b(t)+u_s'(\bar{\xi}_b,t)]$$

$$=\lambda_n[\bar{\Theta}_n'(\bar{\xi}_b)+\bar{\Theta}_n''(\bar{\xi}_b)\xi_b(t)u_n'(\bar{\xi}_b,t)]. \quad (6)$$

Separating the stationary and nonstationary components in the problem (2), with consideration of (6) we have

$$\bar{\Theta}_s''(\xi)=-2\mathrm{Bi}_s\bar{\Theta}_s(\xi)=0, \quad \bar{\Theta}_n''(\xi)-\kappa\bar{\Theta}_n(\xi)+K=0,$$

$$\bar{\Theta}_s(0)=0, \quad \bar{\Theta}_n'(\xi)|_{\xi=1}=0,$$

$$\bar{\Theta}_s(\bar{\xi}_b)=\bar{\Theta}_n(\bar{\xi}_b)=1, \quad \lambda_s\bar{\Theta}_s'(\xi)|_{\xi=\bar{\xi}_b}=\lambda_n\bar{\Theta}_n'(\xi)|_{\xi=\bar{\xi}_b}; \quad (7)$$

$$\frac{\delta^2}{a_s}\dot{u}_s(\xi,t)=u_s''(\xi,t)-2\mathrm{Bi}_s u_s(\xi,t),$$

$$\frac{\delta^2}{a_n}\dot{u}_n(\xi,t)=u_n''(\xi,t)-\kappa u_n(\xi,t),$$

$$u_s(0,t)=0, \quad u_n'(\xi,t)|_{\xi=1}=0,$$

$$u_s(\bar{\xi}_b,t)+\bar{\Theta}_s'(\bar{\xi}_b)\xi_b(t)=u_n(\bar{\xi}_b,t)+\bar{\Theta}_n'(\bar{\xi}_b)\xi_b(t)=0,$$

$$\lambda_s[u_s'(\bar{\xi}_b,t)+\bar{\Theta}_s''(\bar{\xi}_b)\xi_b(t)]=\lambda_n[u_n'(\bar{\xi}_b,t)+\bar{\Theta}_n''(\bar{\xi}_b)\xi_b(t)]. \quad (8)$$

The stationary problem (7) was considered in Ref. 1, where it was shown that the system has only the uniform solution $\bar{\Theta}(\xi)=0$, which corresponds to the superconducting state of the film, for values of $\sigma=K/\kappa$ that are smaller than the critical value σ_c. When $\sigma \gg \sigma_c$, the problem (7) has nonuniform solutions of the form

$$\bar{\Theta}_s(\xi)=\frac{\sinh\sqrt{2\mathrm{Bi}_s}\xi}{\sinh\sqrt{2\mathrm{Bi}_s}\bar{\xi}_b^{(m)}}; \quad \xi<\bar{\xi}_b^{(m)},$$

$$\bar{\Theta}_n(\xi)=\sigma+(1-\sigma)\frac{\cosh\sqrt{\kappa}(1-\xi)}{\cosh\sqrt{\kappa}(1-\bar{\xi}_b^{(m)})}; \quad \xi>\bar{\xi}_b^{(m)}, \quad (9)$$

where $\sigma=K/\kappa$ has the same meaning as the Stekly parameter[1] $\sigma_0=\overline{j^2}\rho_0 h/2\alpha T_c$ for a classical superconductor, to which it is related by the equality $\sigma=\sigma_0/(1-\beta T_c\sigma_0)$.

The coordinate of the phase boundary $\bar{\xi}_0^{(m)}$ is defined as the root of the equation

$$\coth \ \sqrt{2\mathrm{Bi}_n \frac{\sigma_0}{\sigma}} \ (1-\overline{\xi}_b)\coth \ \sqrt{2\mathrm{Bi}_s}\overline{\xi}_b = (\sigma-1) \ \sqrt{\frac{\lambda_n}{\lambda_s}\frac{\sigma_0}{\sigma}}. \tag{10}$$

For planar structures based on classical superconductors the number of equilibrium states m in the supercritical region ($\sigma > \sigma_c$) is equal to two. According to Ref. 1, more than two stationary states are permissible for high-temperature superconductors. It should be stipulated that we are dealing with multifronted two-phase states, whose analysis requires appropriate formulation of the boundary-value problem.

INVESTIGATION OF THE STABILITY OF NONUNIFORM STATIONARY STATES

To determine the stability of the two-phase states in a first approximation we seek solutions of the problem (8). The last two boundary conditions of this problem assign the continuity of the nonstationary components of the temperature and the heat flux on the phase boundary. Let us consider these conditions in greater detail.

Substitution of the solution sought in the form $u(\xi,t) = V(\xi)\exp(\nu t)$ into the continuity condition for the temperature leads to the following law for the motion of the phase boundary:

$$\xi_b(t) = -\frac{V_s(\overline{\xi}_b)}{\overline{\Theta}'_s(\overline{\xi}_b)}\exp(\nu t),$$

$$\xi_b(t) = -\frac{V_n(\overline{\xi}_b)}{\overline{\Theta}'_n(\overline{\xi}_b)}\exp(\nu t). \tag{11}$$

Eliminating $\xi_b(t)$ from (11) and taking into account that

$$\frac{\overline{\Theta}'_s(\overline{\xi}_b)}{\overline{\Theta}'_n(\overline{\xi}_b)} = \frac{\lambda_n}{\lambda_s}, \tag{12}$$

we transform the first of the continuity conditions into the simpler condition

$$\frac{V_s(\overline{\xi}_b)}{V_n(\overline{\xi}_b)} = \frac{\lambda_n}{\lambda_s}. \tag{13}$$

We write the continuity conditions for the heat flux with consideration of (11) as

$$\left[\frac{V'_s(\xi)}{V_s(\xi)} - \frac{\overline{\Theta}''_s(\xi)}{\overline{\Theta}'_s(\xi)}\right]_{\xi=\overline{\xi}_b} = \left[\frac{V'_n(\xi)}{V_n(\xi)} - \frac{\overline{\Theta}''_n(\xi)}{\overline{\Theta}'_n(\xi)}\right]_{\xi=\overline{\xi}_b}. \tag{14}$$

Separating the variables in (8) and replacing the continuity conditions by (13) and (14), we arrive at the Sturm–Liouville problem

$$V''_s(\xi) = \mu_s^2 V_s(\xi), \quad V''_n(\xi) = \mu_n^2 V_n(\xi),$$

$$V_s(0) = 0; \quad V'(\xi)|_{\xi=1} = 0, \quad \frac{V_s(\overline{\xi}_b)}{V_n(\overline{\xi}_b)} = \frac{\lambda_n}{\lambda_s},$$

$$\left[\frac{V'_s(\xi)}{V_s(\xi)} - \frac{\overline{\Theta}''_s(\xi)}{\overline{\Theta}'_s(\xi)}\right]_{\xi=\overline{\xi}_b} = \left[\frac{V'_n(\xi)}{V_n(\xi)} - \frac{\overline{\Theta}''_n(\xi)}{\overline{\Theta}'_n(\xi)}\right]_{\xi=\overline{\xi}_b}, \tag{15}$$

where

$$\mu_s^2 = \frac{\delta^2}{a_s}\nu + 2\mathrm{Bi}, \quad \mu_n^2 = \frac{\delta^2}{a_n}\nu + \kappa.$$

The solutions which satisfy the first three boundary conditions have the form

$$V_s(\xi) = C \sinh \mu_s\xi, \tag{16}$$

$$V_n(\xi) = C\frac{\lambda_s}{\lambda_n}\frac{\sinh \mu_s\overline{\xi}_b}{\cosh \mu_n(1-\overline{\xi}_b)} \cosh \mu_n(1-\xi), \tag{17}$$

where C is a constant.

Substituting (9), (16), and (17) into the last boundary condition, we obtain an equation for the eigenvalues of the operators (15)

$$\mu_s \coth \mu_s\overline{\xi}_b + \mu_n \tanh \mu_n(1-\overline{\xi}_b)$$
$$= \sqrt{2\mathrm{Bi}_s} \tanh \sqrt{2\mathrm{Bi}_s}\overline{\xi}_b + \sqrt{\kappa} \coth \sqrt{\kappa}(1-\overline{\xi}_b). \tag{18}$$

Thus, to determine the stability of the stationary solutions of (9), we must find the roots of the characteristic equation (18) for the corresponding values of $\overline{\xi}_b$.

FURTHER IDEALIZATION OF THE MATHEMATICAL MODEL

The analysis of Eq. (18) in its general form is a fairly formidable problem; therefore, we shall confine ourselves to an investigation of the two-phase state of a superconductor with a low temperature coefficient of resistance in a small vicinity of the transition point. In this case the thermal conductivities of the superconducting and normal phases, as well as the Biot numbers, can clearly be assumed to be equal: $\lambda_s = \lambda_n$, $\mathrm{Bi}_s = \mathrm{Bi}_n \equiv \mathrm{Bi}$. In view of the small value of $\rho_0\beta$, it is permissible to take $\kappa = 2\mathrm{Bi}$ and set $\mu_n = \mu_s \equiv \mu$ and to thereby significantly simplify Eq. (18):

$$\mu[\coth \mu\overline{\xi}_b + \tanh \mu(1-\overline{\xi}_b)] = A, \tag{19}$$

where

$$A = \sqrt{2\mathrm{Bi}} \ [\tanh \sqrt{2\mathrm{Bi}}\overline{\xi}_b + \coth \sqrt{2\mathrm{Bi}}(1-\overline{\xi}_b)]. \tag{20}$$

Now the proof of the stability of the stationary solutions reduces to calculating μ for assigned values of $\overline{\xi}_b$ and Bi and determining the sign of the parameter

$$\nu = \frac{a}{\delta^2}(\mu^2 - 2\mathrm{Bi}). \tag{21}$$

The left-hand side of Eq. (20) is a function of the complex variable $\mu = \mu' + i\mu''$, while the right-hand side of the equation, which is an implicit function of the control parameter $A = A[\overline{\xi}_b(\sigma)]$, does not depend on μ. Separating the real and imaginary parts in (20) and eliminating A, we arrive at the equation
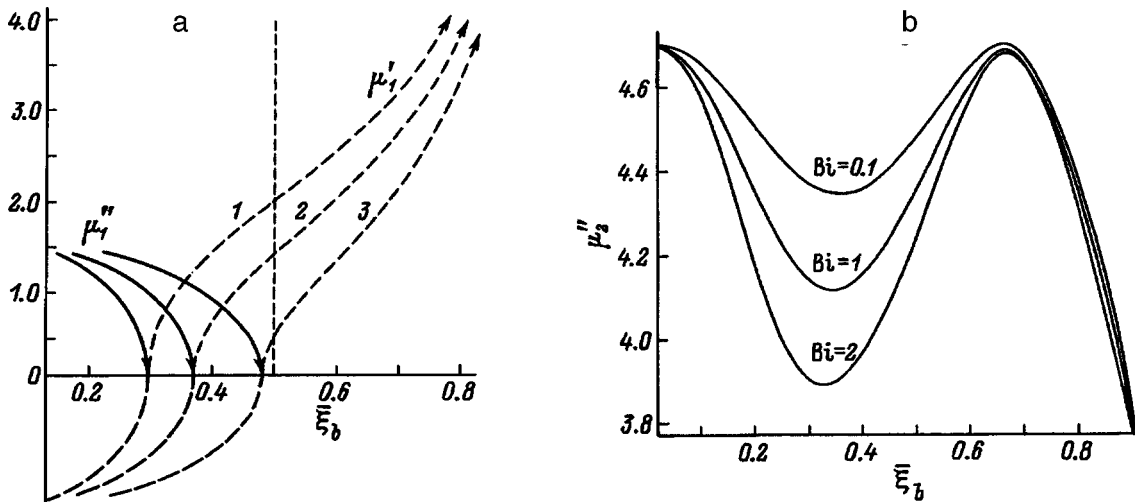
FIG. 2. Dependence of the roots of the characteristic equation on the coordinate of the phase boundary: a—the solid line in the upper half plane corresponds to the position of the first root of Eq. (23) on the imaginary axis, and the dotted line corresponds to the position on the real axis. Biot number: *1*—2, *2*—1, *3*—0.1; b—second root of Eq. (23).

$$\frac{\mu'' \cosh \mu' \cos \mu'' + \mu' \sinh \mu' \sin \mu''}{\mu' \cosh \mu' \cos \mu'' - \mu'' \sinh \mu' \sin \mu''}$$

$$= \frac{\cosh \mu' \sin \mu'' + \cosh \mu'(2\bar{\xi}_b - 1)\sin \mu''(2\bar{\xi}_b - 1)}{\sinh \mu' \cos \mu'' + \sinh \mu'(2\bar{\xi}_b - 1)\cos \mu''(2\bar{\xi}_b - 1)}.$$

$$(22)$$

The results of a numerical search for roots of Eq. (22) indicate that the equation does not have any complex roots and that all the points of the spectrum lie either on the real axis or on the imaginary axis. Alternately setting $\mu = i\mu''$ and $\mu = \mu'$ in (21), we arrive at the following equations:

$$\frac{2\mu'' \cos \mu''}{\sin \mu'' + \sin(2\bar{\xi}_b - 1)\mu''} = \sqrt{2\mathrm{Bi}}(\tanh \sqrt{2\mathrm{Bi}}\bar{\xi}_b$$

$$+ \coth \sqrt{2\mathrm{Bi}}(1 - \bar{\xi}_b)),$$

$$\frac{2\mu' \cosh \mu''}{\sinh \mu' + \sinh(2\bar{\xi}_b - 1)\mu'} = \sqrt{2\mathrm{Bi}}[\tanh \sqrt{2\mathrm{Bi}}\bar{\xi}_b$$

$$+ \coth \sqrt{2\mathrm{Bi}}(1 - \bar{\xi}_b)].$$

$$(23)$$

The first of Eqs. (23) has an infinite number of roots μ_k'', two of which are shown in Figs. 2a and 2b. It is seen from Fig. 2a that the first pair of complex-conjugate roots $\pm i\mu_1''$ tends to zero as $\bar{\xi}_b$ increases. At the value of the Stekly parameter for which $\bar{\xi}_b = \xi_c$, where ξ_c is the root of the equation

$$\sqrt{2\mathrm{Bi}}\bar{\xi}_b[\tanh \sqrt{2\mathrm{Bi}}\bar{\xi}_b + \coth \sqrt{2\mathrm{Bi}}(1 - \bar{\xi}_b)] = 1, \qquad (24)$$

the first pair of roots moves over to the real-number axis. The negative root moves over to the positive semiaxis and vice versa. Thus, the second of Eqs. (23) has a single pair of roots.

The stability of the equilibrium states (9) is determined by the further course of the $\mu'(\bar{\xi}_b)$ curve. As is shown in Fig. 3, the plot of $\bar{\xi}_b(\sigma_0)$ is symmetric relative to $\xi = 0.5$; therefore, in one of the equilibrium states, for example, $m = 1$, we always have $\bar{\xi}_b < 0.5$, while in the other equilibrium state $\bar{\xi}_b > 0.5$. It is easy to see that upon passage of a real root through the point $\xi = 0.5$, i.e., upon passage from one equilibrium state to the other, ν changes sign for any Bi. For this reason, in the second of Eqs. (23) we must set $\bar{\xi}_b = 0.5$ and solve it numerically with respect to μ^2. The plot of $\mu^2 = \mu^2(2\mathrm{Bi})$ in Fig. 4 shows that, regardless of the Biot number, reversal of the sign of ν_1 always occurs at $\bar{\xi}_b = 0.5$ and, therefore, the second equilibrium state is always unstable.
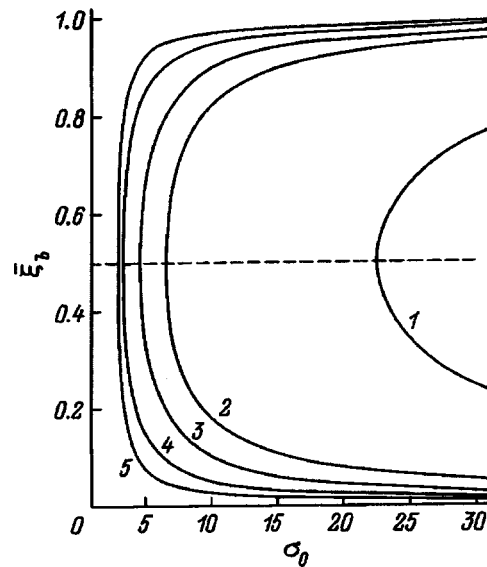

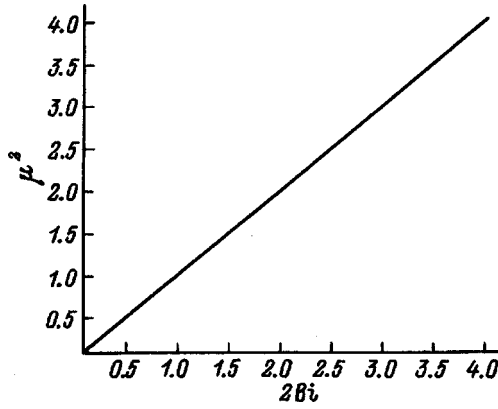FIG. 3. Dependence of the coordinate of the phase boundary on the Stekly parameter. Bi: *1*—0.1, *2*—0.5, *3*—1, *4*—3, *5*—7.

FIG. 4. Plot of $\mu^2(2\mathrm{Bi})$ at the branch point of the stationary solutions for $\sigma=\sigma_c$.

In the range $\bar{\xi}_b<\bar{\xi}_c$, where all the eigenvalues of the operator (16) are imaginary $[\mu_k=i\mu_k''(\bar{\xi}_b)]$, the solutions (17) and (18) of the boundary-value problem (8) have the form

$$\Theta_s(\xi,t)=\sum_{k=1}^{\infty} C_k \sin(\mu_k''\xi)\exp(\nu_k t),$$

$$\Theta_n(\xi,t)=\sum_{k=1}^{\infty} C_k \frac{\lambda_s}{\lambda_n} \frac{\sin(\mu_k''\bar{\xi}_b)}{\cos \mu_k''(1-\bar{\xi}_b)} \cos \mu_k''(1-\xi)\exp(\nu_k t),$$

(25)

where $\nu_k=-a_s/\delta^2(\mu_k''^2+2\mathrm{Bi})$ is negative and the equilibrium states are stable.

When $\bar{\xi}_b>\bar{\xi}_c$, the first eigenvalue of the operator (16) becomes real $[\mu_1=\mu_1'(\bar{\xi}_b)]$, and the solutions (25) transform into

$$\Theta_s(\xi,t)=C \sinh(\mu_1'\xi)\exp(\nu_1 t)$$
$$+\sum_{k=2}^{\infty} C_k \sin(\mu_k''\xi)\exp(\nu_k t),$$

$$\Theta_n(\xi,t)=C \frac{\lambda_s}{\lambda_n} \frac{\sinh(\mu_1'\bar{\xi}_b)}{\cosh \mu_1'(1-\bar{\xi}_b)} \cosh \mu_1'(1-\xi)$$

$$\times\exp(\nu_1 t)+\sum_{k=2}^{\infty} C_k \frac{\lambda_s}{\lambda_n} \frac{\sin(\mu_k''\bar{\xi}_b)}{\cos \mu_k''(1-\bar{\xi}_b)}$$

$$\times\cos \mu_k''(1-\xi)\exp(\nu_k t).$$

(26)

When $\bar{\xi}_b>0.5$, the sign of ν changes, and the solution (26) becomes exponentially unstable.

INVESTIGATION OF THE ASYMPTOTIC BEHAVIOR OF THE STATIONARY SOLUTIONS

Let us conclude with a discussion of the physical nature of an unstable equilibrium state. For reasons of convenience we move the origin of coordinates to the center of the film. Then the stationary solutions (9) transform into

$$\bar{\Theta}_s(\xi)=\frac{\cosh \sqrt{2\mathrm{Bi}_s}(1-\xi)}{\sinh \sqrt{2\mathrm{Bi}_s}(1-\bar{\xi}_b^{(m)})},$$

$$\bar{\Theta}_n(\xi)=\sigma+(1-\sigma) \frac{\cosh \sqrt{\kappa}\xi}{\cosh \sqrt{\kappa}\bar{\xi}_0^{(m)}},$$

(27)

and the condition (10) takes the form

$$\coth \sqrt{2\mathrm{Bi}_n\frac{\sigma_0}{\sigma}} \bar{\xi}_b \coth \sqrt{2\mathrm{Bi}_s}(1-\bar{\xi}_b)=(\sigma-1) \sqrt{\frac{\lambda_n}{\lambda_s}\frac{\sigma_0}{\sigma}}.$$

(28)

Assuming that $\bar{\xi}_b^{(1)}<\bar{\xi}_b^{(2)}$, we consider the $m=2$ state. As the length of the film tends to infinity, i.e., when the system becomes degenerate with respect to δ, the coordinate of the free boundary $\bar{\xi}_b^{(2)}$ also tends to infinity, and the temperature of the normal phase (27) tends to a spatially uniform distribution:

$$\bar{\Theta}_n(\xi)=\sigma.$$

(29)

It follows from (29) that the solution (27) corresponds to a state whose inhomogeneity is caused only by the proximity of the boundary. When the latter is moved away to infinity, the system tends to the normal state with a uniform temperature distribution.

In the case of $m=1$, the limiting transition $\bar{x}_b\to\infty$ leads to the expressions

$$\bar{\Theta}_s(x)=\exp \sqrt{\frac{2\alpha}{h\lambda_s}} (\bar{x}_b^{(1)}-x),$$

$$\bar{\Theta}_n(x)=\sigma+(1-\sigma) \frac{\cosh \sqrt{2\alpha\sigma_0/h\lambda_n\sigma}x}{\cosh \sqrt{2\alpha\sigma_0/h\lambda_n\sigma}\bar{x}_b^{(1)}},$$

(30)

where

$$\bar{x}_b^{(1)}= \sqrt{\frac{h\lambda_n}{2\alpha}\frac{\sigma}{\sigma_0}} \operatorname{arctanh}(\sigma-1) \sqrt{\frac{\lambda_n}{\lambda_s}\frac{\sigma}{\sigma_0}}.$$
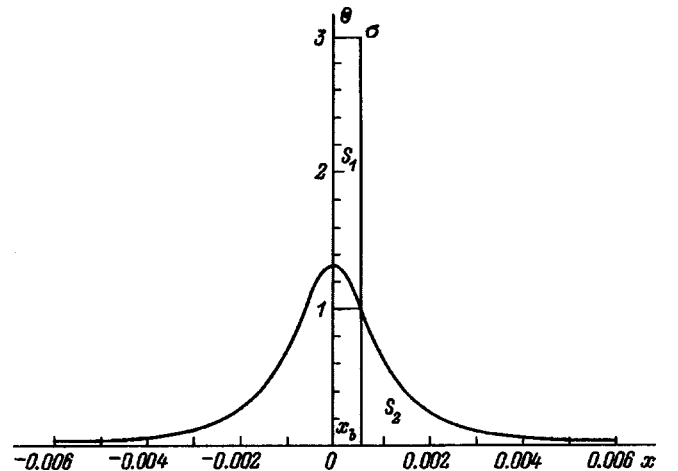
(31)

FIG. 5. Locally nonuniform equilibrium state of an infinite thin superconducting film. The temperature field corresponds to the system parameters $\beta=0$, $\sigma=3$, and $2\alpha/h\lambda_n=10^6$. The additional plots illustrate the law of equal areas.

The same expressions were obtained in Ref. 4 as self-similar solutions describing a localized, spatially nonuniform equilibrium state of an infinitely long superconducting film (Fig. 5).

It is noteworthy that relation (31), like (10), follows directly from the energy conservation law. It is not difficult to obtain these relations by equating the total quantity of heat evolved and the heat loss in the system. For a film of infinite length, relation (31) allows the following simple geometric interpretation. We find the areas S_1 and S_2 of the hatched objects in Fig. 5:

$$S_1 = \int_0^{\bar{x}_b} \bar{\Theta}_n(x)\,dx = (\sigma-1)\sqrt{\frac{h\lambda_n}{2\alpha}\frac{\sigma}{\sigma_0}}\tanh\sqrt{\frac{2\alpha}{h\lambda_n}\frac{\sigma_0}{\sigma}}\,\bar{x}_b,$$

$$S_2 = \int_{\bar{x}_b}^{\infty} \bar{\Theta}_s(x)\,dx = \sqrt{\frac{h\lambda_s}{2\alpha}}. \tag{32}$$

Equating S_1 and S_2, we obtain (31) again, whence it follows that $S_1 = S_2$. The latter relation is a specific addition to the law of equal areas.[5]

Self-sustained localized nonuniform formations in dissipative systems are customarily called autosolitons. In Ref. 4 the equilibrium state defined by (30) and (31) was interpreted as an autosoliton normal phase, since the question of its stability had not been investigated at that time. The same state was obtained above as an asymptotic form of an unstable solution when the system became degenerate; therefore, it can be expected that it also remains unstable in the case of a film of infinite length.

Let us investigate the asymptotic behavior of the characteristic equation defined by (19) and (20). When the origin of coordinates is displaced to $\xi = 1$, Eqs. (19) and (20), which were written for the right-hand half plane ($\xi > 0$), transform into

$$\bar{x}_b\sqrt{\frac{c_{vs}}{\lambda_s}\nu+\frac{2\alpha}{\lambda_s h}}\coth\sqrt{\frac{c_{vs}}{\lambda_s}\nu+\frac{2\alpha}{\lambda_s h}}\,(\delta-\bar{x}_b)$$

$$-\bar{x}_b\sqrt{\frac{2\alpha}{\lambda_s h}}\tanh\sqrt{\frac{2\alpha}{\lambda_s h}}\,(\delta-\bar{x}_b)$$

$$=-\bar{x}_b\sqrt{\frac{c_{vn}}{\lambda_n}\nu+\frac{2\alpha}{\lambda_n h}}\tanh\sqrt{\frac{c_{vn}}{\lambda_n}\nu+\frac{2\alpha}{\lambda_n h}}\,\bar{x}_b$$

$$+\bar{x}_b\sqrt{\frac{2\alpha}{\lambda_n h}}\coth\sqrt{\frac{2\alpha}{\lambda_n h}}\,\bar{x}_b. \tag{33}$$

We assume that the thermal conductivities of the two phases are equal ($\lambda_s = \lambda_n = \lambda$) and that the specific heat undergoes a jump upon passage into the superconducting state ($c_{vs} = c_{vn} + \Delta c$). Introducing the notation

$$z = \bar{x}_b\sqrt{\frac{c_{vn}}{\lambda}\nu+\frac{2\alpha}{\lambda h}}, \qquad y = \bar{x}_b\sqrt{\frac{2\alpha}{\lambda h}}$$

and allowing the film length to tend to infinity ($\delta \to \infty$), we obtain

$$z\left[\tanh z \pm \sqrt{1+\frac{\Delta c}{c_{vn}}\left(1-\frac{y^2}{z^2}\right)}\right]-y(\coth y \pm 1)=0, \tag{34}$$

where the plus sign corresponds to the right-hand phase boundary.

Equation (34), like (22), does not have complex roots, and only one root exists for real z, it being such that $z > y$. Hence it follows that the parameter

$$\nu = \frac{\lambda}{c_{vn}\bar{x}_b^2}(z^2-y^2)$$

is always positive and that the solution defined by (30) and (31) is unstable.

This result means that the localized nonuniform equilibrium state considered in Ref. 4, is not a stationary autosoliton. It cannot appear spontaneously, and the artificially created temperature distribution defined by (30) and (31) is unstable. On the other hand, it is clear that at large values of the variables the implicitly assigned function (34) tends to the explicit form $y = z$. Setting $z - y = \varepsilon$, where $\varepsilon \ll 1$, it is not difficult to obtain the following estimate:

$$\varepsilon \approx \frac{4}{2+\Delta c/c_{nv}}z\,\exp(-2z),$$

$$\nu = \frac{\lambda}{\bar{x}_b^2}\frac{8}{2c_n+\Delta c}z^2\,\exp(-2z), \tag{35}$$

according to which the characteristic exponent of the solution (26) is practically equal to zero when $y \gg 1$. Such a quasistable nonuniform formation can remain long-lived even for comparatively small sizes of the normal region ($\bar{x}_b \sim 10^{-3}$ m).

CONCLUSIONS

The results of the analysis performed allow us to state that of the nonuniform equilibrium states found in Ref. 2 for a thin current-carrying superconducting film, the state which corresponds to a normal-phase region of larger size is stable. The fact that the stability was investigated only with respect to symmetric perturbations is not significant in the present case, since of the two equilibrium states only one is generally unstable.

In the equilibrium states with a normal phase of smaller dimensions the conditions on the film boundary weakly influence the heat balance, and the nonuniform solution remains localized upon passage to a film of infinite length. This solution is also unstable in an infinitely long film, but the relaxation time to the locally uniform equilibrium state can be fairly long when the parameters of the system have certain values. The instability of a given equilibrium state means that spontaneous localization of the normal phase (the formation of a stationary autosoliton) is impossible in the system under consideration. Thin films based on new superconductors may be more promising in this sense. As was shown in Ref. 2, more than two equilibrium states are possible in films with a large temperature coefficient of resistance; therefore, it would be interesting to investigate the asymptotic behavior of the corresponding stable solutions.

[1])The dimensionless quantity σ_0, which serves as a measure of the ratio of the characteristic flux of heat evolved $\overline{j^2}\rho_0 dV$ to the heat removal rate $\alpha(T_c - T_0)dS$ in superconductors with a transport current, is called the Stekly parameter. In particular, for thin films based on classical superconductors $\sigma_0 = \overline{j^2}\rho_0 h/2\alpha(T_c - T_0)$. As was shown in Ref. 1, the ratio $\sigma = K/\kappa$ can serve as the Stekly parameter for materials with a large temperature coefficient of resistance.

[1] A. S. Rudyĭ, Pis'ma Zh. Tekh. Fiz. **22**(9), 85 (1996) [Tech. Phys. Lett. **22**, 382 (1996)].

[2] W. Franzen, J. Opt. Soc. Am. **53**, 596 (1963).

[3] A. S. Rudy and A. Yu. Kolesov, Nonlinear Analysis **1** (1997).

[4] A. S. Rudyĭ, Pis'ma Zh. Tekh. Fiz. **22**(20), 62 (1996) [Tech. Phys. Lett. **22**, 848 (1996)].

[5] A. V. Gurevich, R. G. Mints, and A. L. Rakhmanov, *The Physics of Composite Superconductors*, CRC Press, Boca Raton, Florida (1995) [Russian original, Nauka, Moscow (1987), 240 pp.].

Translated by P. Shelnitz

Collinear light scattering on dipole-exchange spin waves in inhomogeneous ferromagnetic films

L. V. Lutsev

Domen Scientific-Research Institute, 196084 St. Petersburg, Russia
(Submitted June 19, 1995; resubmitted February 9, 1998)
Zh. Tekh. Fiz. **68**, 78–84 (June 1998)

A theory is developed for the collinear TE–TM scattering of optical waveguide modes on dipole-exchange spin waves in perpendicularly magnetized ferromagnetic films that are inhomogeneous across their thickness. It is found in homogeneous ferromagnetic films and in films with small deviations from homogeneity that the TE–TM scattering on higher spin-wave modes is strongest when the synchronism conditions for the transverse phases and for the longitudinal and transverse wave vectors are satisfied. When the thickness of the planar optical waveguide does not match the thickness of the ferromagnetic film, the phase synchronism condition is violated with the resultant appearance of an oscillating type of dependence of the TE–TM scattering on the spin-wave mode number. The scattering of light on spin-wave modes in films with a magnetization gradient is investigated in the presence of turning points for the magnetostatic potential. It is found that the existence of a turning point in the region of the antinode for the optical modes leads to an increase in the scattering amplitude. The formation of inhomogeneous magnetooptical structures and superlattices based on $(Lu,Y,Bi)_3(Fe,Ga)_5O_{12}$ is discussed. © *1998 American Institute of Physics.*
[S1063-7842(98)01506-2]

INTRODUCTION

The interaction of light with spin waves in iron garnet film has been studied intently in recent years. This interaction can be utilized for both practical and research purposes: for ultrahigh-frequency optical modulators and for studying the spin-wave processes occurring in iron garnet thin films. The results of investigations of the noncollinear interaction of optical waveguide modes with spin waves and TE–TM mode conversion were presented in Refs. 1 and 2. The features of the TE↔TM conversion of optical modes upon collinear scattering were studied in Refs. 3–7 by both theoretical and experimental methods. A theoretical analysis of the diffraction of optical modes on surface and bulk spin waves for an arbitrary angle of incidence of the optical mode was conducted in Ref. 8. The purpose of the present work is to take into account the exchange interaction accompanying the scattering of optical waveguide modes on spin waves in inhomogeneous ferromagnetic films. The exchange interaction must be taken into account, if the inhomogeneous ferromagnetic film has a layer with a turning point for the magnetostatic potential of the spin wave. The magnitude of the variable magnetic moment is greater in this layer than in other layers, leading, in turn, to an increase in the scattering of the optical waveguide modes.

This paper is divided into three parts. The first two parts describe the properties of dipole-exchange spin waves and present the dispersion relations and eigenfunctions of optical waveguide modes. In the third part the optical-mode coupling equations are derived, and the conditions for achieving maximum TE–TM scattering are analyzed for various film structures with homogeneous and inhomogeneous ferromag-

netic layers. The formation of inhomogeneous magnetooptical structures and superlattices based on $(Lu,Y,Bi)_3(Fe,Ga)_5O_{12}$ is discussed.

1. DIPOLE-EXCHANGE SPIN WAVES

Let us consider a perpendicularly magnetized ferromagnetic planar structure of thickness d with magnetic and dielectric parameters that are inhomogeneous across its thickness (Fig. 1). The $0z$ axis is perpendicular, and the $0x$ and $0y$ axes are parallel to the film surface. We assume that the spin-wave and optical modes propagate along the $0x$ axis.

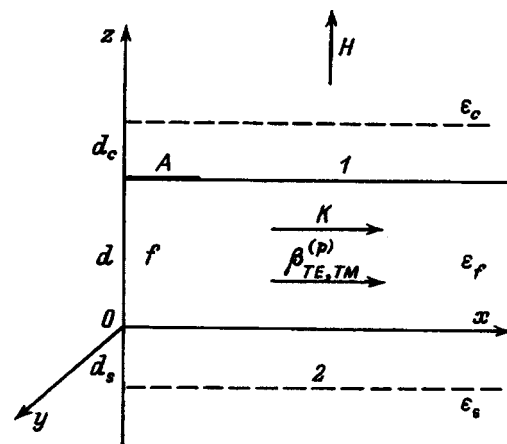
FIG. 1. Geometry of a planar structure for collinear light scattering on spin waves: *f*—ferromagnetic layer; *1, 2*—cladding and transitional nonferromagnetic layer with the thicknesses d_c and d_s; *A*—antenna used to excite the spin waves.

Spin waves are described in the magnetostatic approximation by the magnetostatic potential $\varphi(x,z,t)$, which is found from the equations[9]

$$\frac{\partial \mathbf{M}}{\partial t} = -\gamma \left[M \cdot \frac{\delta \mathcal{H}}{\delta \mathbf{M}} \right],$$

$$\mathrm{div}(\mathbf{h} + 4\pi\mathbf{m}) = 0, \tag{1}$$

where

$$\mathcal{H} = \int \left[-\mathbf{M}(\mathbf{H}+\mathbf{h}) + 2\pi(\mathbf{M},\mathbf{n})^2 - \frac{1}{2}\beta_a(\mathbf{M},\mathbf{n})^2 \right.$$
$$\left. - \frac{1}{2}\alpha \frac{\partial \mathbf{M}}{\partial r_i}\frac{\partial \mathbf{M}}{\partial r_i} \right] dV \tag{2}$$

is the effective classical dipole-exchange Hamiltonian, $\mathbf{M} = \mathbf{M}_0 + \mathbf{m}(x,z,t)$ is the magnetic moment density ($\mathbf{M}_0$ does not depend on x and t, and $|\mathbf{m}(x,z,t)| \ll M_0$), γ is the gyromagnetic ratio, $\mathbf{H} \| 0z$ is the external constant magnetic field, $\mathbf{h}(x,z,t) = -\nabla\varphi(x,z,t)$ and $\varphi(x,z,t)$ are the variable magnetic field and the magnetostatic potential of the spin wave, α is the exchange coupling constant, β_a specifies the uniaxial anisotropy field $\mathbf{H}_a = \beta_a(\mathbf{M},\mathbf{n})\mathbf{n}$ with the $\mathbf{n}$ axis, which is perpendicular to the film surface, the term $2\pi(\mathbf{M},\mathbf{n})^2$ in (2) describes the energy of the demagnetizing magnetic field of the film, and $\partial/\partial r_i$ is an abbreviated form for writing the derivatives $\partial/\partial x$, $\partial/\partial y$, and $\partial/\partial z$.

The system of equations (1) was investigated in Ref. 10 in a linear approximation with respect to $\mathbf{m}(x,z,t)$ for the case of films that are inhomogeneous across their thickness with the magnetic parameters $\gamma(z)$, $\alpha(z)$, $M_0(z)$, and $H_a(z)$. The calculation of TE–TM scattering on a spin wave requires finding the distribution of $\mathbf{m}$ across the thickness of the ferromagnetic structure. We shall consider ferromagnetic films that are homogeneous and weakly inhomogeneous across their thickness, as well as ferromagnetic films with a magnetization gradient across their thickness. In the latter case it is assumed that the magnetostatic potential φ of the spin wave has a turning point within the film.

A. Ferromagnetic films that are homogeneous and weakly inhomogeneous across their thickness. We assume that the spin-wave frequency ω is far from the ferromagnetic resonance frequencies of any layer of the ferromagnetic film. In this case the potential φ does not have turning points within the film, and the magnetic susceptibility tensor $\chi_{ik}(\omega)$ does not have singular points. Ferromagnetic structures that are weakly inhomogeneous across their thickness are understood to be structures with weak deviations $\Delta\chi_{ik}$ from the mean values $\bar{\chi}_{ik}$ (Ref. 10). The small value of $\Delta\chi_{ik}/\bar{\chi}_{ik}$ makes it possible to use perturbation theory, where the first approximations for calculating the dispersion relations and the magnetostatic potential are the dispersion relations and the potential φ of a homogeneous film with the mean parameters $\bar{\chi}_{ik}$, $\bar{\gamma}$, $\bar{\alpha}$, $\bar{M}_0$, and $\bar{H}_a$. The expansion parameters are the $\Delta\chi_{ik}/\bar{\chi}_{ik}$, which, in turn, are determined by the relative deviations $\Delta\gamma/\bar{\gamma}$, $\Delta\alpha/\bar{\alpha}$, $\Delta M_0/\overline{M_0}$, and $\Delta H_a/\bar{H}_a$.

The spin-wave eigenfunctions of a homogeneous film are

$$\varphi(x,z,t) = (2\pi)^{-1}\varphi_n(z)\exp(ikx + i\omega_n t), \tag{3}$$

where

$$\varphi_n(z) = P_n \begin{cases} \cos[k_z^{(n)}(z-d/2) \\ \quad + (n-1)\pi/2] & (0 \leq z \leq d), \\ (-1)^{n-1}\dfrac{k_z^{(n)}}{k_0^{(n)}}\exp[|k|(d-z)] & (z>d), \\ \dfrac{k_z^{(n)}}{k_0^{(n)}}\exp(|k|z) & (z<0), \end{cases} \tag{4}$$

$n = 1,2,3,\dots$ is the mode number, P_n is the normalization parameter, $k_0^{(n)2} = k^2 + k_z^{(n)2}$, k is the wave vector, and $k_z^{(n)}$ is defined by the relation

$$2\cot k_z^{(n)}d = \frac{k_z^{(n)}}{|k|} - \frac{|k|}{k_z^{(n)}}. \tag{5}$$

The dispersion relation for the nth mode has the form

$$\omega_n^2 = (\Omega + \gamma\alpha M_0 k_0^{(n)2})(\Omega + \gamma\alpha M_0 k_0^{(n)2}$$
$$+ \gamma 4\pi M_0 k^2/k_0^{(n)2}), \tag{6}$$

where $\Omega = \gamma(H - 4\pi M_0 + H_a)$.

Taking into account the values of the group velocity $v_g^{(1)}$ of the first mode and $k_z^{(n)}$ for $|k| \ll \pi/d$ (Ref. 10), it follows from the dispersion relation (6) that the dispersion curves of the first and higher modes cross when

$$k_n = \alpha\gamma M_0 k_z^{(n)2}/v_g^{(1)} = \alpha\pi(n-1)^2/d^3. \tag{7}$$

The second equality in (1) gives the relation between the variation of the magnetic moment density $m_x(z)$ and the magnetostatic potential $\varphi_n(z)$ of the nth mode:

$$m_x(z) = \frac{ik_0^{(n)2}}{4\pi|k|}\varphi_n(z). \tag{8}$$

For a weakly inhomogeneous ferromagnetic film, in a first approximation with respect to the degree of deviation of the magnetic parameters from the parameters of the homogeneous film structure the dispersion relation of a spin wave is specified by the expression

$$\omega_n = \langle n|\Omega(z) + \gamma(z)M_0(z)(\alpha(z)k_0^{(n)2} + 2\pi k^2/k_0^{(n)2})|n\rangle, \tag{9}$$

where $\langle n|f(z)|l\rangle \equiv \int_0^d \varphi_n^*(z)f(z)\varphi_l(z)dz$ with the normalization factors $P_j = (d/2 + |k|/k_0^{(j)2})^{-1/2}$ $(j=n,l)$.

To determine the coupling coefficient of TE and TM modes for scattering on a spin wave it is convenient to relate the normalization of the eigenfunctions $\varphi_n(z)$ (4) to the energy $U^{(n)}$ of the spin wave per unit area of the film. According to Ref. 11, the energy of a spin wave $U^{(n)}$ is related to the magnon number density N_n:

$$U^{(n)} = \int_0^d N_n\hbar\omega_n dz = \int_0^d \frac{\bar{m}_x^2\omega_n}{2\gamma M_0}dz \tag{10}$$

(the bar denotes a time average).

With consideration of (3), (4), and (8), from (10) we find the normalization parameter

$$P_n = \frac{(4\pi)^2 |k|}{k_0^{(n)2}} \sqrt{\frac{2\gamma M_0 U^{(n)}}{\omega_n (d + 2|k|/k_0^{(n)2})}}. \tag{11}$$

B. Ferromagnetic films with a magnetization gradient across their thickness. Let us consider a ferromagnetic film with linear variation of the magnetization $4\pi M_0 = 4\pi \bar{M}_0 - \mu(z - d/2)$ and the frequency range in which there is a turning point for the potential φ_n within the ferromagnetic film structure. According to Ref. 10, the distribution of $m_\pm = m_x \pm i m_y$ across the thickness is described by the expression

$$m_\pm(z) = \frac{i|k|}{w(0)} \left[v_2^\pm(z) \int_0^z \frac{\varphi(\xi)}{\alpha(\xi)} v_1^\pm(\xi) d\xi \right.$$
$$\left. + v_1^\pm(z) \int_z^d \frac{\varphi(\xi)}{\alpha(\xi)} v_2^\pm(\xi) d\xi \right], \tag{12}$$

where $w(0)$ is a Wronskian, $v_1^\pm(z) = \mathrm{Ai}[(\mu/\alpha\bar{M}_0)^{1/3}(z - z_0^\pm)]$ and $v_2^\pm(z) = \mathrm{Bi}[(\mu/\alpha\bar{M}_0)^{1/3}(z - z_0^\pm)]$ are Airy functions, and $z_0^\pm = [\mp\omega/\gamma - \alpha k^2 \bar{M}_0 - H + 4\pi\bar{M}_0 - H_a(0)]/\mu$.

The substitution of $m_\pm(m_x, m_y)$ into the second equality in (1) gives an integrodifferential equation in $\varphi(z)$. This equation was solved numerically. The dispersion relation $\omega(k)$ was found from the requirement that $\varphi(z)$ and $\partial\varphi(z)/\partial z$ be continuous on the boundary of the ferromagnetic film. Outside the film $\varphi(z) \sim \exp(-|kz|)$.

When $k \to 0$, we can employ a simplified formula specifying ω_n (Refs. 12 and 13):

$$\omega_n = \min \Omega(z) + \gamma [2\pi^2 \mu^2 \alpha \bar{M}_0 (n - 1/2)^2]^{1/3}, \tag{13}$$

where

$$\Omega(z) = \gamma(z)[H - 4\pi M_0(z) + H_a(z)].$$

In this approximation the turning point z_r is given by the relation

$$z_r = [2\pi^2 \alpha \bar{M}_0 (n - 1/2)^2/\mu]^{1/3}. \tag{14}$$

Normalization of the eigenfunctions $\varphi_n(z)$ per unit of the energy density of the spin wave $U^{(n)}$, which is needed to compare the TE–TM scattering amplitudes in different ferromagnetic structures, was found numerically using Eqs. (10) and (12).

2. OPTICAL WAVEGUIDE MODES

The equations describing TE and TM optical waveguide modes and the effects of TE↔TM conversion are obtained from Maxwell's equations.[14] We consider the case in which the diagonal component of the dielectric constant $\varepsilon_0(z)$ is a function only of z and is much greater than the off-diagonal components. The off-diagonal components take into account the gyrotropic effects, i.e., the dependence on $\mathbf{m}(x, z, t)$:[15]

$$\varepsilon_0(z) \gg \varepsilon_{lj} = i g e_{ljk} m_k(t)(l \neq j), \tag{15}$$

where $g = F_f \lambda \sqrt{\varepsilon_f}/\pi M_0$, F_f is the Faraday coefficient, λ is the wavelength of the light in a vacuum, ε_f is the mean value of $\varepsilon_0(z)$ in the ferromagnetic film, e_{ljk} is a totally antisymmetric tensor, and $l, j, k = \{x, y, z\}$.

The equations for the TE and TM modes are derived in analogy to the equations for the TE and TM modes in a planar structure in Refs. 14 and 15 by taking the Fourier transform with respect to t and retaining the first terms of the approximation with respect to $\varepsilon_{lj}/\varepsilon_0$ $(l \neq j)$. A TE mode is completely characterized by the field component E_y:

$$\left(\frac{\partial^2}{\partial z^2} + \frac{\partial^2}{\partial x^2} + \frac{\varepsilon_0(z)\omega_{TE}^2}{c^2} \right) E_y + \frac{i\omega_{TE}^2}{c \varepsilon_0(z)\omega_{TM}}$$
$$\times \left(\varepsilon_{yx} * \frac{\partial H_y}{\partial z} - \varepsilon_{yz} * \frac{\partial H_y}{\partial x} \right) = 0. \tag{16}$$

A TM mode is completely characterized by H_y:

$$\varepsilon_0(z) \frac{\partial}{\partial z} \left(\frac{1}{\varepsilon_0(z)} * \frac{\partial H_y}{\partial z} \right) + \frac{\partial^2 H_y}{\partial x^2} + \frac{\varepsilon_0(z)\omega_{TM}^2}{c^2} H_y + \varepsilon_0(z)$$
$$\times \left[\frac{\partial}{\partial z} \left(\frac{\varepsilon_{xz}}{\varepsilon_0(z)^2} * \frac{\partial H_y}{\partial x} \right) + \frac{\partial}{\partial x} \left(\frac{\varepsilon_{zx}}{\varepsilon_0(z)^2} * \frac{\partial H_y}{\partial z} \right) \right]$$
$$+ \frac{i\omega_{TM}\varepsilon_0(z)}{c} \left[\frac{\partial}{\partial z} \left(\frac{\varepsilon_{xy}}{\varepsilon_0(z)} * E_y \right) - \frac{\partial}{\partial x} \left(\frac{\varepsilon_{zy}}{\varepsilon_0(z)} * E_y \right) \right] = 0. \tag{17}$$

In (16) and (17) the symbol $*$ denotes the convolution

$$(u * w)(\omega) = (2\pi)^{-1/2} \int u(\omega - \omega_1) w(\omega_1) d\omega_1.$$

The terms containing the convolution describe effects with a change in frequency upon scattering. If ω_{TE} and ω_{TM} are much greater than the spin-wave frequency ω_n in (6), (9), or (13), the convolution is replaced by the product. The terms with H_y in (16) and with E_y in (17) lead to TE↔TM mode conversion. The fourth term with H_y in (17) describes the modulation of the TM mode by the spin wave and will not be taken into account below. Because of the condition (15), the terms with H_y in (16) and with E_y in (17) can be regarded as perturbations.

For films with constant values of $\varepsilon_0(z)$ (ε_c is the value in the cladding, ε_f is the value in the film, and ε_s is the value in the substrate) the eigenfunctions of the unperturbed equations (16) and (17) are orthogonal to one another and have the form[14]

$$\Psi^{(p)}(x, y) = (2\pi)^{-1/2} \psi^{(p)}(z) \exp(i\beta^{(p)} x),$$

$$\psi^{(p)}(z) = P^{(p)}$$
$$\times \begin{cases} (1 + a_c^{(p)2})^{-1/2} \exp[-\gamma_c^{(p)}(z - d)] & (z > d), \\ \cos(k_f^{(p)} z - \theta_s^{(p)}) & (0 < z < d), \\ (1 + a_s^{(p)2})^{-1/2} \exp(\gamma_s^{(p)} z) & (z < 0), \end{cases} \tag{18}$$

where $p = 0, 1, 2, \ldots$ is the mode number, $k_f^{(p)2} = -\beta^{(p)2} + \varepsilon_f \omega^2/c^2$, $\gamma_{s,c}^{(p)2} = \beta^{(p)2} - \varepsilon_{s,c} \omega^2/c^2$, and $\theta_{s,c}^{(p)} = \arctan a_{s,c}^{(p)}$. For a TE mode

$$a_{s,c}^{(p)} = \gamma_{s,c}^{(p)}/k_f^{(p)}, \quad P^{(p)} = \left(\frac{2}{d + \gamma_c^{(p)-1} + \gamma_s^{(p)-1}} \right)^{1/2}.$$

For a TM mode

$$a_{s,c}^{(p)} = \frac{\gamma_{s,c}^{(p)}\varepsilon_f}{k_f^{(p)}\varepsilon_{s,c}},$$

$$P^{(p)} = \sqrt{2}\left[d + \frac{k_f^{(p)} + \gamma_s^{(p)}a_s^{(p)}}{k_f^{(p)}\gamma_s^{(p)}(1+a_s^{(p)2})} + \frac{k_f^{(p)} + \gamma_c^{(p)}a_c^{(p)}}{k_f^{(p)}\gamma_c^{(p)}(1+a_c^{(p)2})}\right]^{-1/2}.$$

The subscripts f, s, and c indicate that the quantity refers to the film, the substrate, and the cladding. In the phase variables $\theta_{s,c}^{(p)}$ the dispersion relations take the simple form[14]

$$k_f^{(p)}d - \theta_s^{(p)} - \theta_c^{(p)} = \pi p \quad (p=0,1,2,\ldots). \tag{19}$$

The functions (18) together with the radiative modes form a complete orthonormal system. They will be used to calculate the coupling coefficient of TE and TM modes and the conditions for synchronism in an inhomogeneous film in first-order perturbation theory.

3. TE–TM SCATTERING OF OPTICAL MODES ON SPIN WAVES

Variation of the magnetic moment density $\mathbf{m}(x,z,t) = \{m_x,m_y,0\}$ of a spin wave leads to the interconversion of TE and TM modes. We represent the electromagnetic field in a planar waveguide in the form of the superposition of a pair of nearby TE and TM modes:

$$\Psi_B^{(p)}(x,z) = F(x)\Psi_{\text{TE}}^{(p)}(x,z) + G(x)\Psi_{\text{TM}}^{(p)}(x,z), \tag{20}$$

where $\Psi_{\text{TE}}^{(p)}$ and $\Psi_{\text{TM}}^{(p)}$ are the functions (18).

We take into account that a) the derivatives of the amplitudes $\partial F(x)/\partial x$ and $\partial G(x)/\partial x$ are terms that are small in first order, i.e., $\beta_{\text{TE}}^{(p)-1}F^{-1}\partial F/\partial x \ll 1$ and $\beta_{\text{TM}}^{(p)-1}G^{-1}\partial G/\partial x \ll 1$, that b) under the condition $\omega_{\text{TE}},\omega_{\text{TM}} \gg \omega_n$ the convolution in (16) and (17) is replaced by the product, and that c) the dependence of ε_{lj} $(l \neq j)$ on $\mathbf{m}$ has the form (15) and is proportional to $\exp(ikx+i\omega_n t)$. In these approximations we obtain the coupling equations

$$\frac{\partial F(x)}{\partial x} = \frac{i\omega_{\text{TE}}^2\beta_{\text{TM}}^{(p)}\langle E_y^{(p)}|\varepsilon_{yz}|H_y^{(p)}\rangle}{2c\varepsilon_f\omega_{\text{TM}}\beta_{\text{TE}}^{(p)}}\exp(-i\Delta x)G(x),$$

$$\frac{\partial G(x)}{\partial x} = \frac{i\omega_{\text{TM}}(\beta_{\text{TE}}^{(p)}-k)\langle H_y^{(p)}|\varepsilon_{zy}|E_y^{(p)}\rangle}{2c\beta_{\text{TM}}^{(p)}}\exp(i\Delta x)F(x), \tag{21}$$

where

$$\langle E_y^{(p)}|\varepsilon_{yz}|H_y^{(p)}\rangle = \langle H_y^{(p)}|\varepsilon_{zy}|E_y^{(p)*}\rangle$$

$$= \int_0^d \psi_{\text{TE}}^{(p)*}(z)\varepsilon_{yz}(z)\psi_{\text{TM}}^{(p)}(z)dz,$$

$$\Delta = \beta_{\text{TE}}^{(p)} - \beta_{\text{TM}}^{(p)} - k, \quad \varepsilon_{yz} = igm_x.$$

When the synchronism condition $\Delta = 0$ is satisfied, the coupling coefficient of the TE and TM modes reaches a maximum and is given by the expression

$$\nu^{(p)2} = \frac{\omega_{\text{TE}}^2(\beta_{\text{TE}}^{(p)}-k)|\langle E_y^{(p)}|\varepsilon_{yz}|H_y^{(p)}\rangle|^2}{4c^2\varepsilon_f\beta_{\text{TE}}^{(p)}}. \tag{22}$$

The coupling coefficient $\nu^{(p)}$ determines the period of the spatial oscillations along the $0x$ axis for TE$\leftrightarrow$TM conversion:

$$L^{(p)} = \frac{2\pi}{\nu^{(p)}}. \tag{23}$$

We shall now examine some special cases of TE–TM scattering on spin waves in ferromagnetic films that are homogeneous and weakly inhomogeneous across their thickness and in films with a magnetization gradient.

A. Ferromagnetic films that are homogeneous and weakly inhomogeneous across their thickness. Substituting the eigenfunctions (4) and (18) with consideration of (8), (11), and (15) into (22), we obtain the coupling coefficient of a pair of TE and TM modes upon scattering on the nth spin-wave mode

$$\nu_n^{(p)} = P_{\text{TE}}^{(p)}P_{\text{TM}}^{(p)}F_f\left[\frac{(\beta_{\text{TE}}^{(p)}-k)\gamma U^{(n)}}{2\beta_{\text{TE}}^{(p)}\omega_n(d+2|k|/k_0^{(n)2})M_0}\right]^{1/2}$$

$$\times\left|\sum_{j,l=1}^2 \frac{\sin(K_{jl}^{(n,p)}d-\Xi_{j,l}^{(n,p)})+\sin\Xi_{j,l}^{(n,p)}}{K_{jl}^{(n,p)}}\right|, \tag{24}$$

where $K_{jl}^{(n,p)} = k_z^{(n)} + \eta^{(j)}k_{f,\text{TE}}^{(p)} + \eta^{(l)}K_{f,\text{TM}}^{(p)}$,

$\Xi_{jl}^{(n,p)} = (k_z^{(n)}d-\pi(n-1))/2 + \eta^{(j)}\theta_{s,\text{TE}}^{(p)} + \eta^{(l)}\theta_{s,\text{TM}}^{(p)}$,

$\eta^{(1)} = 1, \quad \eta^{(2)} = -1$.

An analysis of the relation obtained (24) shows that the scattering will be strongest when the conditions

$$K_{je}^{(n,p)} = 0 \quad (j,l=1,2), \tag{25}$$

$$\Xi_{jl}^{(n,p)} = 2\pi r \quad (j,l=1,2, \ r=0,\pm 1,\pm 2,\ldots), \tag{26}$$

which can be called the conditions for synchronism of the transverse wave vectors and phase synchronism, are satisfied. The physical meaning of these conditions can be explained in the following manner. Let us consider two adjacent layers with antinodes for the spin-wave mode. Since $\mathbf{m}$ has opposite values in these layers, the rotation of the polarization plane is determined by the difference between the Faraday effects in these layers. The maximum total rotation of the polarization plane occurs in the case in which there is an antinode of the optical mode in one of the layers and there is a node in the other layer, as is reflected in the conditions (25) and (26).

To illustrate the importance of the fulfillment of relations (25) and (26) for obtaining maximum TE$\rightarrow$TM conversion, we performed numerical calculations for a YIG/GGG structure and $k=k_n$. Fulfillment of the condition $\Delta = \beta_{\text{TE}}^{(p)} - \beta_{\text{TM}}^{(p)} - k = 0$, which can be called the synchronism condition for longitudinal wave vectors, was achieved by introducing layer 2 (Fig. 1) with linear variation of the dielectric constant according to the law $\varepsilon(z) = \varepsilon_f + (\varepsilon_f - \varepsilon_s)z/d_s$ $(z \in [0,-d_s])$ within the layer (other ways of achieving the condition $\Delta = 0$ were described in Ref. 16). The eigenfunctions of the optical modes of such a film structure were found from (16)
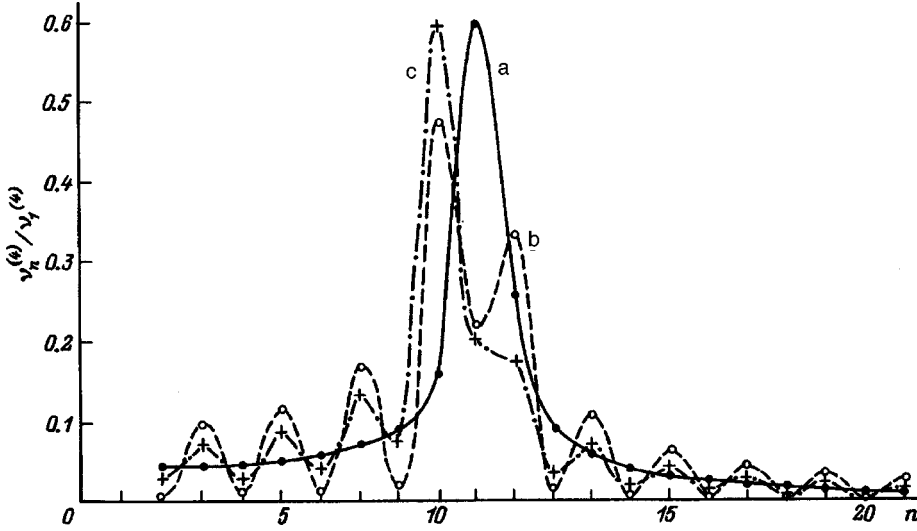
FIG. 2. Ratio between the coupling coefficients $\nu_n^{(4)}/\nu_1^{(4)}$ for TE$_4\to$TM$_4$ scattering in a homogeneous ferromagnetic layer (a YIG/GGG structure) as a function of the spin-wave mode number for $k=k_n$ [Eq. (7)]. d_c, μm: a—0, b—0.5, c—1.0.

and (17) using perturbation theory in the form of a series in powers of the deviation from a homogeneous film. In a first approximation the eigenfunctions have the form (18) with altered values of $\beta^{(p)}$

$$\text{TE:} \beta^{(p)2} = \beta_0^{(p)2} + (2\pi/\lambda)^2 \langle E_y^{(p)}|(\varepsilon(z)-\varepsilon_0)|E_y^{(p)}\rangle,$$

$$\text{TM:} \beta^{(p)2} = \beta_0^{(p)2} + (2\pi/\lambda)^2 \langle H_y^{(p)}|(\varepsilon(z)-\varepsilon_0)|H_y^{(p)}\rangle$$
$$- \left\langle H_y^{(p)} \left| \frac{\partial \varepsilon(z)}{\varepsilon(z)\partial z} \frac{\partial}{\partial z} \right| H_y^{(p)} \right\rangle. \qquad (27)$$

Figure 2 presents the ratios between coupling coefficients $\nu_n^{(4)}/\nu_1^{(4)}$ for TE$_4\to$TM$_4$ scattering as a function of the number n of the spin-wave mode for various values of the cladding thickness d_c. The mean values for YIG were used: $4\pi M_0 = 1750$ Oe, $H_a = 0$, $\alpha = 4\pi \cdot 3.2 \times 10^{-12}$ cm^2, $\gamma = 2\pi \times 2.83$ MHz/Oe for $d = 10$ μm, $\omega_n/2\pi = 9$ GHz, $\lambda = 1.15$ μm, $n_c = \sqrt{\varepsilon_c} = 1.0$, $n_f = \sqrt{\varepsilon_f} = 2.220$, and $n_s = \sqrt{\varepsilon_s} = 1.945$. The transition layer 2 (Fig. 1), which was employed to achieve the condition $\Delta = 0$, had a thickness $d_s \cong 0.6$ μm. The presence of this layer required recalculation of the values of $\nu_n^{(4)}$, $K_{jl}^{(n,4)}$, and $\Xi_{jl}^{(n,4)}$ with altered values of $\beta^{(4)}$ according to Eqs. (18), (24), and (27). The cladding 1 had a dielectric constant identical to the dielectric constant of the YIG layer ($\sqrt{\varepsilon_f} = 2.220$) and was nonmagnetic. Thus, the thickness of the spin-wave waveguide did not coincide with the thickness of the optical waveguide. This led to the additions $k_f^{(p)}d_c$ to the $\Xi_{jl}^{(n,p)}$, to violation of the phase synchronism conditions (26), and, thus, to the oscillating character of the dependence of the TE–TM scattering on the spin-wave mode number. The synchronism condition (25) $k_z^{(n)} - k_{f,\text{TE}}^{(4)} - k_{f,\text{TM}}^{(4)} = 0$ was satisfied for $n = 9 - 10$.

Formula (24) for the coupling coefficient $\nu_n^{(p)}$ was obtained for a homogeneous ferromagnetic layer. The employment of this formula for weakly inhomogeneous ferromagnetic structures in the absence of turning points for the magnetostatic potential is permissible in first-order perturba-

tion theory. The expansion parameter here is $\Delta\chi_{ik}/\bar{\chi}_{ik}$, which, in turn, is determined by the relative deviations $\Delta\gamma/\bar{\gamma}$, $\Delta\alpha/\bar{\alpha}$, $\Delta M_0/\bar{M}_0$, and $\Delta H_a/\bar{H}_a$.

B. Ferromagnetic films with a magnetization gradient across their thickness. The coupling coefficient of TE and TM modes upon scattering on the nth spin-wave mode in a film with the magnetization gradient $4\pi M_0 = 4\pi\bar{M}_0 - \mu(z - d/2)$ was found from (22) after plugging in the eigenfunctions (18) with consideration of the distribution of $m_\pm$ [Eqs. (12) and (15)]. A numerical calculation was performed for a YIG/GGG structure. The external magnetic field (for $\omega_n/2\pi = 9$ GHz) was selected so that there would be a turning point for the magnetostatic potential within the ferromagnetic film. Figure 3 presents the ratios of the coupling coefficient $\nu_n^{(4)}$ for TE$_4\to$TM$_4$ scattering in a film with a gradient to the coupling coefficient $\nu_{1\text{hom}}^{(4)}$ for TE$_4\to$TM$_4$ scattering in a homogeneous ferromagnetic film as a function of the spin-wave mode number n of the for $k \to 0$. The values $4\pi\bar{M}_0 = 1750$ Oe and $\mu = 10$ Oe/μm were used. All the remaining parameters were the same as for the homogeneous ferromagnetic structure. The condition $\Delta = 0$ was achieved just as in the case of the homogeneous ferromagnetic film by introducing intermediate layer 2 (Fig. 1). An analysis of the distribution of $m_\pm$ (12) reveals that $m_\pm$ has its greatest amplitude in the vicinity of the turning point z_r (14). If the antinodes of the TE and TM optical modes are located in the layer with the turning point z_r, the TE–TM scattering is strongest. In Fig. 3 this is observed for the spin-wave modes with $n = 2$, 7–8, and 14–16. Variation of the cladding thickness d_c leads to displacement of the antinodes relative to z_r. It is seen from a comparison of the plots in Figs. 2 and 3 that if there is a turning point for the magnetostatic potential $\varphi(z)$, the amplitude of the TE–TM scattering in inhomogeneous ferromagnetic films can take a larger value than in homogeneous films. It can be concluded on this basis that the most promising magnetooptical materials are those in which TE–TM scattering is possible on a periodic distribution of several layers with turning points for $\varphi(z)$. They can be magnetic superlattices, i.e., film structures with spatially pe-
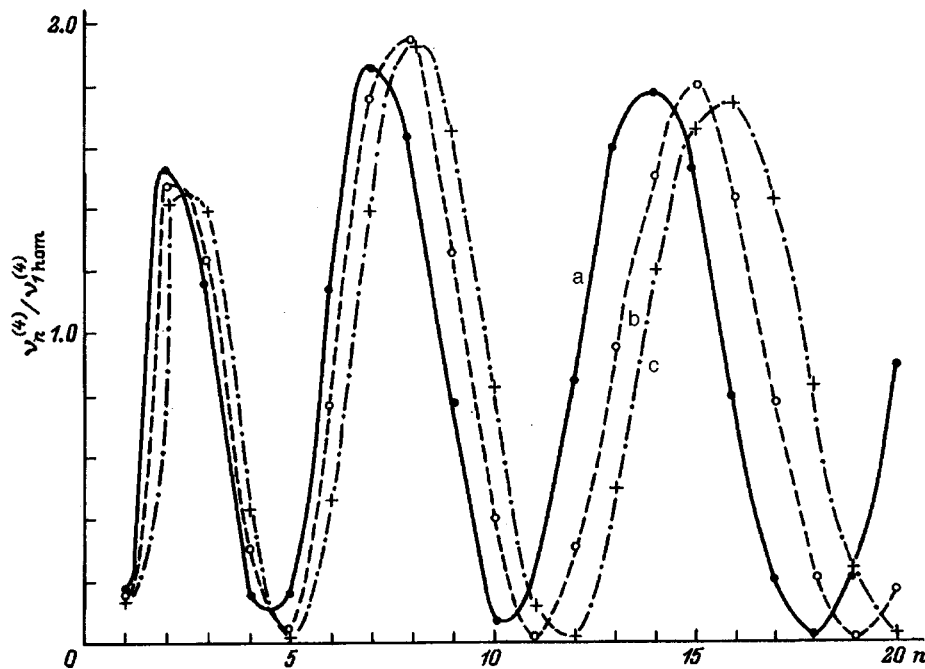
FIG. 3. Ratio between the coupling coefficients $\nu_n^{(4)}/\nu_{1\,\mathrm{hom}}^{(4)}$ for $TE_4 \rightarrow TM_4$ scattering in an inhomogeneous ferromagnetic layer as a function of the spin-wave mode number for $k \rightarrow 0$. The values of d_c are the same as in Fig. 2.

riodic deviations of the magnetic parameters from constant values across the thickness, with $d \gg \sqrt{L}$ and a fairly large number of periods. It was noted in Ref. 10 that superlattice structures can be obtained on the basis of multicomponent iron garnets. Iron garnets with (Lu^{3+},Y^{3+}),(Lu^{3+},Y^{3+},Bi^{3+}), or (Lu^{3+},Y^{3+},La^{3+}) in dodecahedral lattice sites were proposed. Periodic variation of the growth conditions during epitaxial growth leads to variations in the entry of these ions into the dodecahedral sites and to variations in the entry of Pb^{2+} and Pb^{4+}. This, in turn, leads to variations in the uniaxial growth anisotropy.[12] More detailed investigations were performed for the iron garnet $(Lu,Y,Bi)_3(Fe,Ga)_5O_{12}$ in Ref. 17. Trial runs of the epitaxial growth of films showed that the formation of these iron garnet structures during a single production cycle can be achieved by varying the production parameters (the growth temperature, the rotation rate of the substrate, etc.).

CONCLUSIONS

The following conclusions can be drawn as a result of the theoretical analysis performed.

a) In homogeneous ferromagnetic films and in films with slight deviations from homogeneity TE–TM scattering on higher spin-wave modes is strongest when the synchronism conditions for the transverse phases and the longitudinal and transverse wave vectors are satisfied. When the thicknesses of the planar optical waveguide and the ferromagnetic film do not match, the phase synchronism condition is violated with the resultant appearance of an oscillating type of dependence of the TE–TM scattering on the spin-wave mode number.

b) A layer with a turning point for the magnetooptical potential in a film with a magnetization gradient across its thickness makes the largest contribution to TE–TM scattering. The juxtaposition of this layer, whose position depends on the spin-wave mode number, to the region of the antinode

of the optical modes increases the amplitude of the TE–TM scattering in comparison to the scattering in homogeneous films.

[1] D. Young and C. S. Tsai, Appl. Phys. Lett. **53**, 1696 (1988).

[2] C. S. Tsai and D. Young, Appl. Phys. Lett. **54**, 196 (1989).

[3] A. D. Fisher, J. N. Lee, E. S. Gaynor, and A. B. Tveten, Appl. Phys. Lett. **41**, 779 (1982).

[4] Yu. V. Gulyaev, I. A. Ignat'ev, V. G. Plekhanov, and A. F. Popkov, Radiotekh. Elektron. **30**, 1522 (1985).

[5] O. G. Rutkin, N. G. Kovshikov, A. A. Stashkevich *et al.*, Pis'ma Zh. Tekh. Fiz. **11**, 933 (1985) [Sov. Tech. Phys. Lett. **11**, 386 (1985)].

[6] N. Bilaniuk and D. D. Stancil, J. Appl. Phys. **67**, 508 (1990).

[7] V. V. Matyushev and A. A. Stashkevich, J. Appl. Phys. **69**, 5972 (1991).

[8] A. A. Solomko, Yu. A. Gaĭdaĭ, A. V. Dovzhenko *et al.*, Opt. Spektrosk. **66**, 190 (1989) [Opt. Spectrosc. (USSR) **66**, 110 (1989)].

[9] A. I. Akhiezer, V. G. Bar'yakhtar, and S. V. Peletminskii, *Spin Waves*, North-Holland, Amsterdam (1968).

[10] L. V. Lutsev, Zh. Tekh. Fiz. **65**(2), 41 (1995) [Tech. Phys. **40**, 139 (1995)].

[11] A. G. Gurevich and G. A. Melkov, *Magnetization Oscillations and Waves*, CRC Press, Boca Raton (1996) [Russian original, Nauka, Moscow (1994), 464 pp.].

[12] L. V. Lutsev, V. O. Shcherbakova, and G. Ya. Fedorova, Fiz. Tverd. Tela. (St. Petersburg), **35**, 2208 (1993) [Phys. Solid State **35**, 1098 (1993)].

[13] L. V. Lutsev, I. L. Berezin, and Yu. M. Yakovlev, Elektron. Tekh. Ser. Elektron. Sverkhvys. Chastot., No. 5(419), pp. 5–8 (1989).

[14] H. Kogelnik, in *Guided-Wave Optoelectronics*, T. Tamir (Ed.), Springer-Verlag, Berlin–New York (1988); Mir, Moscow (1991), pp. 18–131.

[15] A. K. Zvezdin and V. A. Kotov, *Magnetooptics of Thin Films* [in Russian], Nauka, Moscow (1988), 192 pp.

[16] A. M. Prokhorov, G. A. Smolenskiĭ, and A. N. Ageev, Usp. Fiz. Nauk. **143**, 33 (1984) [Usp. Fiz. Nauk **27**, 339 (1984)].

[17] L. V. Lutsev, V. L. Ivashintsova, Yu. M. Yakovlev *et al.*, in *First Joint Conference on Magnetoelectronics* [in Russian], Moscow (1995), pp. 105–106.

Translated by P. Shelnitz

Postgrowth residual stresses in polycrystalline zinc selenide

L. K. Andrianova, I. I. Afanas'ev, and A. A. Dunaev

S. I. Vavilov State Optical Institute All-Russia Science Center, 199034 St. Petersburg, Russia
(Submitted August 30, 1996; resubmitted December 8, 1997)
Zh. Tekh. Fiz. **68**, 85–90 (June 1998)

The residual stresses in samples of polycrystalline ZnSe are studied by measuring the
photoelasticity in the visible part of the spectrum with transillumination parallel and perpendicular
to the growth axis. The thermal and growth components of the birefringence, which exhibit
different types of distributions among samples, are investigated. It is established that the thermal
component has a nearly equilibrium distribution, while the growth component has an
asymmetric distribution, which reflects individual features in the growth of each specific sample.
© *1998 American Institute of Physics.* [S1063-7842(98)01606-7]

The residual stresses in industrial metals, which comprise a widely encountered class of polycrystalline metals, are customarily divided into stresses of the first and second kinds.[1] The former include macrostresses, which affect the volume of a body as a whole, and the latter include microstresses, which act within the grains of a material and on their boundaries. Unlike metals, polycrystalline ZnSe is transparent in the visible region of the spectrum. Under the influence of mechanical stresses and strains, the spherical optical indicatrix $(X_i/n^2)=1$ is slightly distorted and takes the form of a uniaxial or biaxial ellipsoid, depending on the character of the stressed (strained) state. This permits the application of the ordinary tools and methods of photomechanical analysis to it. Optical effects of the action of stresses of both kinds can be observed in the field of vision of a polariscope. In the present work only stresses of the first kind are investigated.

The application of the methods of photomechanics to zinc selenide is significantly simpler than in the case of other photoelastic crystalline materials. First, like AgCl or KRS, it has a fairly high optical sensitivity,[2,3] but, unlike those materials, it is less plastic and, therefore, behaves essentially like an elastic medium, and the application of the law of photoelasticity, which does not take into account the time factor when a force is applied, is more correct for just such media. Second, it can be regarded as a proven fact that in any case polycrystalline ZnSe is an isotropic photoelastic material within the sensitivity range of polariscope-polarimeters.[3] This facilitates application of the photoelasticity law to it without consideration of the crystallographic coordinates of the medium, which must be taken into account in the case of single crystals.[4] Thus, it can be regarded as a simple and effective sensor of the piezobirefringence and stresses that appear or are manifested in various stages of a production process.

The original disks of polycrystalline ZnSe, with a diameter of 150–400 mm and a thickness of 15–40 mm, were grown by vacuum desublimation.[5] An analysis of the microstructure and x-ray diffraction investigations provide evidence of the anisotropy of the external form of the grains and their textured character, which become stronger as the condensate thickness increases.[6,7] The disks were optically treated in both planes and investigated in polariscope-polarimeters with a field of vision from 150 to 300 mm without magnification using a diffuse source of white light. A KSP-7 polariscope with a magnification of 8–10× and a Sénarmont compensator for an optical wavelength of 546.1 nm, which was isolated by a filter, was employed for more exact measurements. Templates cut from disks so that in a section one pair of their sides would be equal to the thickness of the disk and the length would be equal to the radius of the disk or a chord, depending on the cutting geometry, were investigated in a similar manner. The disks and templates were marked to show which planes belonged to the substrate and the growth surface.

A right-hand system of axis, in which the growth axis coincides with the direction of the Z coordinate, was used to describe the geometry of the piezobirefringence and the stresses (Fig. 1). In such a system the field of values of the birefringence $N_\parallel = \Delta n_r - \Delta n_\alpha$, where Δn_r and Δn_α are small increments of the refractive index along a radius of the disk and in the direction perpendicular to it, can be measured when the disks are transilluminated along Z. In an axisymmetric distribution of the birefringence, the central part of the circular disks is free of birefringence, since $\Delta n_r = \Delta n_\alpha$. The birefringence in directions perpendicular to Z cannot be measured in an intact disk, since there are no immersion liquids with a refractive index close to 2.60; therefore, such measurements were performed on templates. When the templates are transilluminated in directions perpendicular to Z, the field of values of the birefringence $N_\perp = \Delta n_z - \Delta n_\alpha$, where Δn_z is the increment of the refractive index along the Z axis and Δn_α is the increment of the refractive index along a direction perpendicular to Z measured in the perpendicular direction, can be determined. In the central parts of the axisymmetric field $N_\parallel$ and $N_\perp$ depend on the influence of the thermal stresses caused by the axial temperature gradient controlling the vapor condensation process.

It is known from the theory of photoelasticity of optically and mechanically isotropic bodies that differences between normal stresses can be calculated from measured values of the birefringence, if the photoelastic constant of the material is known. Using the notation $t_\parallel(r,\alpha)=\sigma_r-\sigma_\alpha$, we
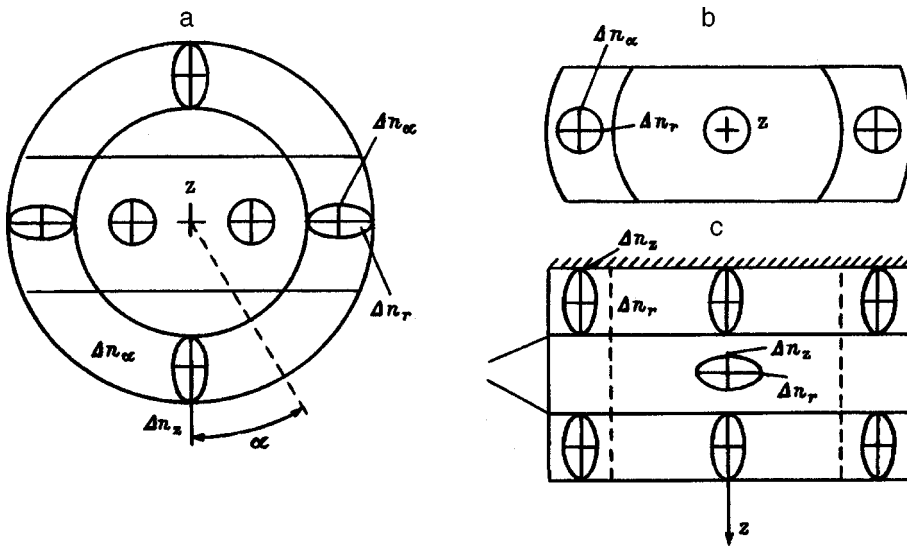
FIG. 1. Asymmetric arrangement of ellipses in sections of the optical indicatrix in planes of the samples investigated far from the edge part when disks are measured along the Z axis and $\Delta n_r - \Delta n_\alpha > 0$ (a), when a template is measured along the Z axis and $\Delta n_r - \Delta n_\alpha = 0$ (b), and when a template is measured perpendicularly to the Z axis (c): *1*—neutral lines with zero birefringence.

can write the relation between the stresses and the birefringence in the form of the formula

$$t_\| = B N_\| \cos 2\varphi_\|, \tag{1}$$

where $B = 2/n^3 (\pi_{11} - \pi_{12})$ is the photoelastic constant of the material and $\varphi_\|$ is the azimuth of the principal directions of the indicatrix and of the stresses σ_r and σ_α.

When the distribution of the optical anisotropy and the stresses in the original disk is axisymmetric, the principal mechanical and optical directions are oriented radially; therefore, $\varphi_\| = 0$. Otherwise, the principal stresses are not radial, and this situation leads to the appearance of tangential stresses, $\tau_\|$, which can be defined by the formula

$$\tau_\| = B N_\| \sin 2\varphi_\| / 2. \tag{2}$$

Knowing $t_\|$ and $\tau_\|$, we can also define the difference between the principal stresses $t = \sqrt{t_\|^2 + 4\tau_\|^2}$, the azimuth of the principal optical and mechanical directions being defined by an angle $\varphi_\| \neq 0$ directly during the polarization measurements. In addition,

$$\tan 2\varphi_\| = 2\tau_\| / t_\|. \tag{3}$$

RESULTS

1. Residual birefringence and stresses. The magnitude and distribution of the optical anisotropy clearly depend on the magnitude of the axial temperature gradient, which assigns the growth rate. If the thickness of the layer grown is small, the amount of heat removed from the cylindrical surface of the disk is also small. However, above a certain thickness, at which the cylindrical surface acquires a sufficient area, the latter becomes a source of heat losses, which cause the appearance of a radial temperature gradient. Transilluminating the disk along the Z axis, we discover axisymmetric fields of $N_\|$ (Fig. 1a). When there is no radial temperature gradient, there are likewise no peripheral zones with birefringence. In this case the disk does not contain piezobirefringence of the first kind, but it has birefringence, which is noticeable within the grain structure of the material (piezo-

birefringence of the second kind). A similar phenomenon is observed when templates are transilluminated in the same direction (Fig. 1b).

Regardless of whether a distribution of birefringence is observed in the disks when they are observed along the Z axis or not, the piezobirefringence $N_\perp$ is always observed when they are observed in the directions perpendicular to Z, i.e., in templates of the radial and chord type (Fig. 1c). The pattern of this birefringence in the form of indicatrices and their orientation contains an odd number of zones with birefringence of opposite sign. If $N_\| = 0$ over the entire plane of a disk, and $N_\perp \neq 0$, as is always observed, the observed birefringence is associated with the formation of uniaxial optical indicatrices of opposite sign (Fig. 1c). Such indicatrices clearly form under the action of an axial temperature gradient $\Delta T_\|$ and in the absence of a radial gradient $\Delta T_\perp$. If a radial gradient still appears for a number of technological reasons, it leads to the distortion of circular sections of the uniaxial indicatrices (Fig. 1c). Biaxial indicatrices, whose circular sections can be at different angles to the observation directions taken in the present work, form. The sections of such indicatrices in the plane of the sample display ellipses in peripheral zones (Fig. 1a). A comparison of this figure and Fig. 1c reveals the similarity between them, which is confined to the fact that both patterns are caused by the removal of thermal energy from the surface of the growing preform. With respect to the temperature gradients $\Delta T_\|$ and $\Delta T_\perp$, the indicatrices have identical signs. However, the conditions for mechanical equilibrium of a disk and a template differ significantly. In a disk the thermal stresses created on the edge by the condition $\Delta T_\perp \neq 0$ are balanced by the stresses of the central circular zone, in which the thermal strains are equal in all directions lying in the plane of the disk. This leads to the formation of a uniaxial indicatrix at the center of the sample. When $\Delta T_\| \neq 0$ and the sample is observed in a direction perpendicular to the Z axis, the central zone of a template contains birefringence of opposite sign (Fig. 1c), and its thickness is divided into three zones with birefringence of opposite sign. Thus, the distribution of $N_\|$ measured

a

37 mm

-50 0 50
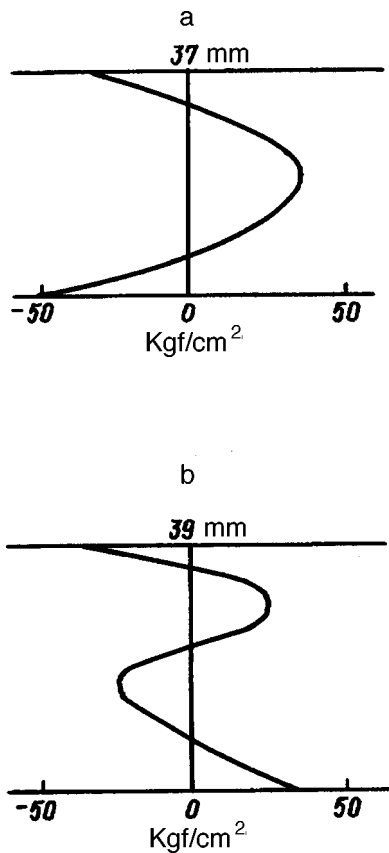
Kgf/cm^2

b

39 mm

-50 0 50

Kgf/cm^2

FIG. 2. Principal types of postgrowth optical anisotropy observed in the direction perpendicular to the growth axis: a—equilibrium diagram of residual thermal stresses of the quenching type (the symmetric type), b—equilibrium diagram of residual thermal stresses with growth stresses due to an axial temperature gradient (the asymmetric type).

a

30 mm

-50 0 50

Kgf/cm^2

b

25 mm

-50 0 50

Kgf/cm^2

FIG. 3. Influence of mechanical treatment on residual thermal stresses of the symmetric type (stress diagrams): a—original state, b—after reduction of the template thickness.

along the Z axis can be represented with motion along the diameter by an $N_\parallel - 0 - N_\parallel$ pattern, and $N_\perp$ can be represented with motion along the Z axis by an $N_\perp - (-N_\perp) - N_\perp$ pattern. Knowing the photoelastic properties of the material,[2,3] its behavior upon heating in the field of vision of the crossed polariscope in the temperature range 20–200 °C (Ref. 8), and relations (1)–(3), we can make the transition from indicatrices to stresses using the classical model of a simple birefringent plate for this transition. The diagram of stresses caused by the radial temperature gradient during the growth and cooling of the sample consists of peripheral compressive stresses, which are balanced at the center by stresses of opposite sign that do not produce birefringence. Cooling from the growth temperatures 800–1000 °C was followed by plastic deformation,[9] as a result of which the sign of the residual stresses was the reverse of the sign observed during growth and cooling under the action of $\Delta T_\perp$.[1] Residual peripheral compressive stresses, which are balanced by uniform in-plane stresses that are equal to one another, but cannot be determined by the polarization method, appear. The magnitude of these stresses is small and amounts to about -30 kgf/cm^2. These stresses vanish after the templates are cut in accordance with the scheme in Fig. 1b.

Figures 2a and 2b present two types of diagrams of the residual stresses $t_\perp = \sigma_z - \sigma_\alpha$, which were observed on several tens of templates. We call the diagram in Fig. 2a sym-
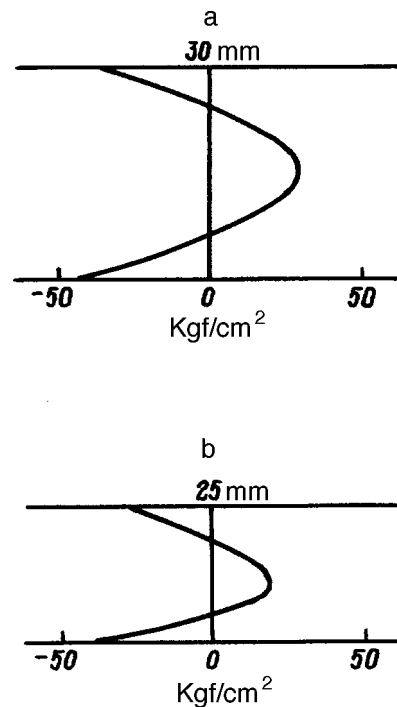
metric, and the diagram in Fig. 2b asymmetric. According to Ref. 8, the diagram in Fig. 2a should correspond to heating. During cooling, the signs of the stresses are reversed due to the plasticity under the high-temperature conditions. Therefore, this diagram could have formed only after plastic deformation, apparently of the grain-boundary type,[9] which would lead, precisely as in the case in Fig. 2, to reversal of the signs of the residual stresses, whose structure reflects the action of the axial temperature gradient $\Delta T_\parallel$ as a whole to within the signs. The diagram in Fig. 2b is also frequently encountered and has a more complicated type of distribution of $N_\perp$ and $t_\perp$ than in the preceding diagram. They are, as it were, two diagrams of opposite sign, which are combined with one another so that the opposite sides of a template have residual stresses of opposite sign. On the substrate side the upper part of the diagram has a form similar to a diagram of the symmetric type. This diagram, however, smoothly joins the analogous diagram of opposite sign in the lower part of the template. The tensile stresses on the lower edge of the template provide evidence that it was, as it were, additionally heated at a low temperature, which is such that even upon further cooling the diagram of opposite sign does not manage to form because of the low plasticity of the material at those temperatures. The material behaved like a hard spring, but this was not previously detected in Ref. 9, since the thermal strains were not frozen.

2. Influence of mechanical treatment on $N_\perp$ and $t_\perp$. The birefringence $N_\parallel$ in disks has been investigated fairly thoroughly, since it is often monitored during the fabrication of optical elements from glass and crystals. The behavior of $N_\perp$, which is observed in templates, is not so well known. The influence of the mechanical treatment of templates
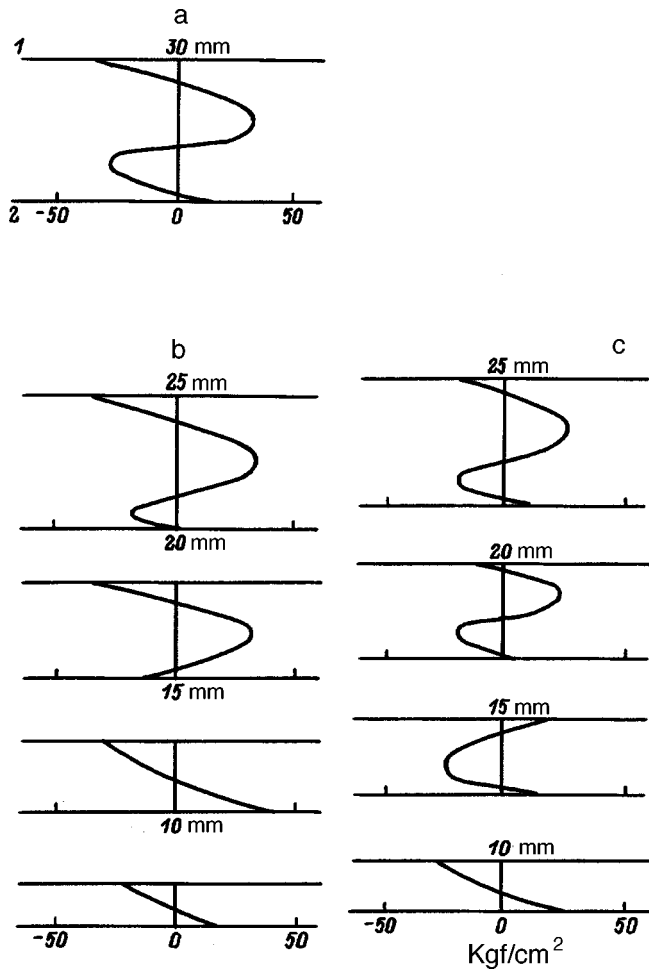
FIG. 4. Influence of mechanical treatment on residual thermal stresses of the asymmetric type (stress diagrams): a—original state, b, c—after reduction of the template thickness; *1*—substrate, *2*—growth surface.

across their thickness is shown in Fig. 3a. After a total reduction of the thickness of a template on both sides by 16%, the original symmetric diagram (Fig. 3a) exhibited slight changes in its form in the part adjacent to the substrate, providing some basis to regard the contact with the substrate as an additional factor which increases the stresses in this part. The opposite side underwent significantly smaller changes. For this reason, it would be useful to investigate the influence of mechanical treatment of the plane of a template adjacent to the substrate in greater detail, rather than the variation of $N_\perp$ and $t_\perp$ in templates with diagrams of the asymmetric type. Figure 4 shows the variation of the stress diagram when 16, 33, 50, and 66% of the thickness is unilaterally ground off. After the thickness is reduced by 33%, the diagram remains asymmetric. After 50% is ground off, the diagram becomes symmetric, but has a sign which is opposite to that of Fig. 3a, and after further treatment (66%), it transforms into a diagram which is similar to a diagram for four-point bending, where the tensile stresses from the growth surface increase significantly and compressive stresses appear on the surface on the substrate side.

3. Influence of isothermal postgrowth annealing. A series of templates from a single original disk was subjected to isothermal annealing at constant temperatures from 400 to
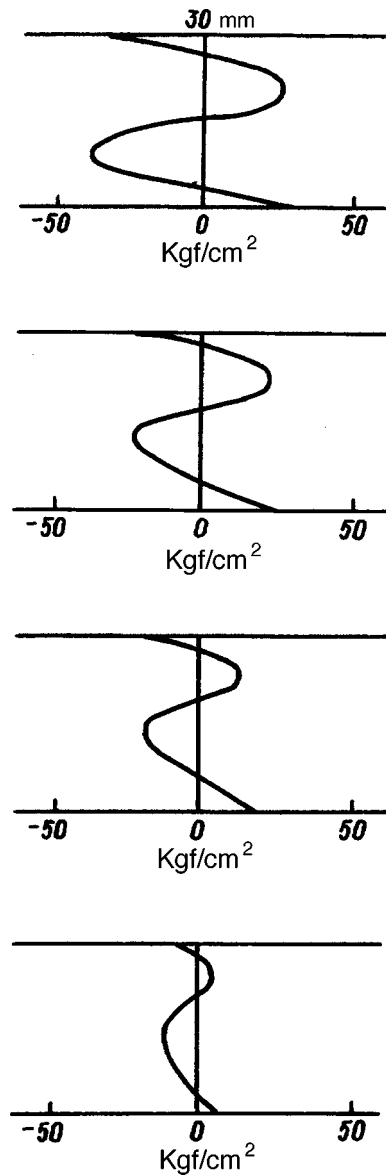

FIG. 5. Influence of isothermal annealing on residual thermal stresses.

1000 °C with holding for 5 h. Each template was annealed once. The annealing temperature variation step was 100 °C. All the templates had the same original birefringence pattern down to the grain structure. The band patterns were photographed before and after annealing, and stress diagrams were constructed from them according to the method adopted. The templates were held in a gradient-free portion of the furnace, and the cooling to room temperature after each anneal was not forced. It was found that annealing at temperatures from 400 to 700 °C did not cause changes in the diagrams of the residual birefringence $N_\perp$ and the stresses $t_\perp$. Annealing at temperatures from 800 to 1000 °C with the same cooling procedure lowers $N_\perp$ and $t_\perp$. The decrease in the residual stresses at 800 °C amounts to 25%, while annealing at 1000 °C lowers them by 80–90%, i.e., increases the optical homogeneity of the material. Figure 5 shows the character of the decrease in the level of the residual stresses as a function of the annealing temperature for a diagram of the asymmetric type.

4. Influence of mechanical treatments. The four-point bending of templates was employed to study the influence of the grain inhomogeneity across the thickness and the residual stresses on the mechanical behavior of the material in the region corresponding to its elasticity at room temperature. The bending was performed so that the stresses appearing as a result would stress the surface facing the growth front, the surface facing the substrate, and the surface parallel to the *Z* axis upon transillumination of the sample in the same direction. The loading was carried out in the field of vision of the polariscope by a lever press with a graduated load. The mechanical moment step during the loading was 0.25 kgf. An analysis of the band pattern for the loading cases indicated demonstrates their completely identical nature, with the exception of the local edge deviations caused by individual features of the grain structure. The higher was the mechanical moment, the more similar the band patterns became. We observed a similar response of the material upon the diametric compression of disks with a diameter of 30 mm and a thickness of 4 mm, which were cut parallel, perpendicularly, and at a 45° angle to the growth axis.[9] This is evidence that, as a whole, the material is homogeneous and isotropic according to its photomechanical behavior and that its grain inhomogeneity, texture, growth features, and residual stresses do not create any macroscopic mechanical features such as reinforcement, anisotropy, and the like in the material.

CONCLUSIONS

1. The induced optical anisotropy in polycrystalline zinc selenide grown by a desublimation technology has two components: a growth component and a thermal component.

2. The thermal component of the birefringence and stresses is observed when disks or templates are transilluminated parallel to the axis of the temperature gradient providing for the growth process. This component has a nearly equilibrium distribution.

3. The growth component of the birefringence and the stresses is most clearly expressed in templates transilluminated perpendicularly to the growth axis. It very often has an asymmetric distribution, which is far from equilibrium, and a tendency toward stratification, which reflects individual features in the growth of each concrete sample.

[1] Ya. B. Fridman, *Mechanical Properties of Metals* [in Russian], Oborongiz, Moscow (1952), 554 pp.

[2] K. K. Dubenskiĭ, A. A. Kaplyanskiĭ, and N. G. Lozovskaya, Fiz. Tverd. Tela (Leningrad) **8**, 2068 (1966) [Sov. Phys. Solid State **8**, 1644 (1967)].

[3] L. K. Andrianova, I. I. Afanas'ev, A. A. Demidenko *et al.*, Opt. Mekh. Promst. **57**(10), 36 (1990) [Sov. J. Opt. Technol. **57**, 615 (1990)].

[4] I. I. Afanas'ev and N. V. Ionina, Opt. Spektrosk. **67**, 319 (1989) [Opt. Spectrosc. (USSR) **67**, 184 (1989)].

[5] I. A. Maksimova, I. A. Mironov, and V. N. Pavlova, ''Method for obtaining polycrystalline blocks of zinc and cadmium chalcogenides for optical ceramics'' [in Russian], Inventor's Certificate No. 844609.

[6] A. A. Demidenko, A. A. Dunaev, S. N. Kolesnikova, and I. A. Mironov, Vysokochist. Veshchestva, No. 1, pp. 103–109 (1991).

[7] G. V. Anan'eva, A. A. Dunaev, and T. I. Merkulyaeva, Vysokochist. Veshchestva, No. 4, pp. 114–119 (1995).

[8] L. K. Andrianova, I. I. Afanas'ev, A. A. Dunaev *et al.*, Zh. Tekh. Fiz. **62**(8), 102 (1992) [Sov. Phys. Tech. Phys. **37**, 708 (1992)].

[9] I. I. Afanas'ev, L. K. Andrianova, and A. A. Demidenko, Opt. Mekh. Promst. **57**(8), 42 (1990) [Sov. J. Opt. Technol. **57**, 490 (1990)].

Translated by P. Shelnitz

Analysis of structural defects in boron-implanted silicon single crystals on the basis of the results of double- and triple-crystal x-ray diffractometry

A. P. Petrakov, N. A. Tikhonov, and S. V. Shilov

Syktyvkar State University, 167001 Syktyvkar, Russia
(Submitted May 6, 1996; resubmitted December 19, 1997)
Zh. Tekh. Fiz. **68**, 91–96 (June 1998)

The structural defects in Si single crystals are analyzed on the basis of diffraction reflection curves and triple-crystal spectra. The relative variation of the lattice period and its distribution as a function of depth are calculated, and the type of defects appearing and the behavior of the implanted impurity in response to high-temperature annealing are determined. © *1998 American Institute of Physics.* [S1063-7842(98)01706-1]

INTRODUCTION

Ion implantation is an effective tool for altering the electrical properties of semiconductor materials. Boron is employed quite often as the implantant. Ion implantation causes damage to the subsurface structure of crystals subjected to irradiation. This structural damage is investigated by different methods. Among the nondestructive methods, double- and triple-x-ray diffractometry are very informative. These methods have been used in numerous studies (see, for example, Refs. 1–5) of the kinds of defects formed in silicon single crystals as a result of implantation. At the same time, the development of a method for treating experimental spectra, particularly triple-crystal spectra, would make it possible to obtain new data on defects. The present paper analyzes the results of systematic triple-crystal x-ray diffractometric investigations of silicon single crystals that were implanted with boron in various doses and subjected to isothermal annealing at various temperatures and for various annealing times.

EXPERIMENTAL METHOD

Nearly perfect single-crystal wafers of KDB-10 silicon with a thickness of 500 μm were investigated. The surfaces of the samples coincided with the (111) plane. The angle between them, which was measured by an x-ray diffractometric method, was about 10″. The samples were implanted with B^+ having an energy of 25 keV in doses from $D=6.2 \times 10^{14}$ to 6.25×10^{15} cm^{-2}. The implantation was carried out at room temperature under conditions which rule out channeling. The use of a fairly weak ion current with a density of 0.2 μA/cm^2 also ruled out the phenomenon of self-annealing during implantation. After implantation, some of the samples were annealed in a nitrogen atmosphere. The annealing temperature was varied from 300 to 1000 °C. The annealing times were 10, 60, and 120 min.

The structure of the subsurface layers of the silicon single crystals was diagnosed using an automatic double- and triple-crystal x-ray diffractometer, which was assembled on the basis of a DRON-UM1 x-ray diffractometer. Dispersion-free double-crystal $(n,-n)$ and triple-crystal $(n,-n,n)$ geometries[6,7] and Cu $K_{\alpha 1}$ radiation were employed. The double-crystal diffraction reflection curves and the triple-crystal x-ray diffraction spectra were measured in the ε scan mode (rotation of the analyzer). The sample azimuthal angle (ω) was varied from $-500''$ to $+500''$. High-perfection silicon single crystals with a symmetric (111) reflection served as the monochromator and analyzer. The half-widths of the diffraction reflection curves of the

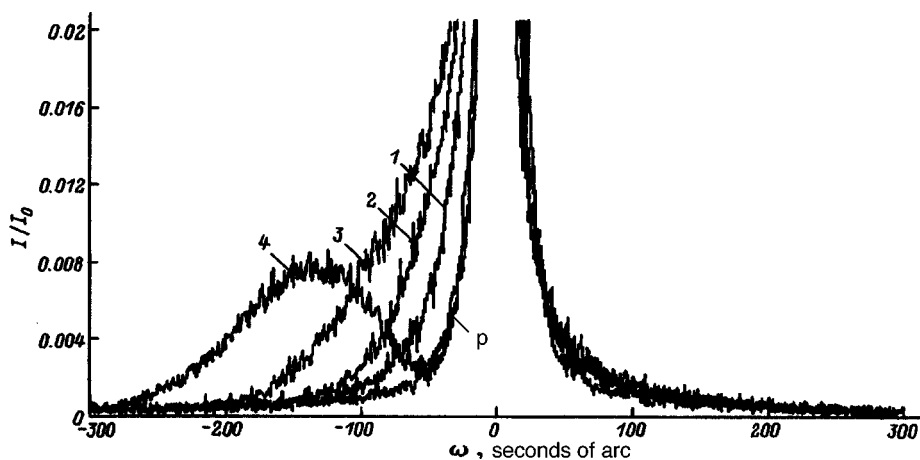FIG. 1. Diffraction reflection curves of crystals. Implantation dose, cm^{-2}: *1*—6.25 $\times 10^{14}$, *2*—1.875$\times 10^{15}$, *3*—3.125$\times 10^{15}$, *4*—6.25$\times 10^{15}$; *P*—curve for a perfect crystal, I/I_0—ratio of the reflected to the incident intensity.
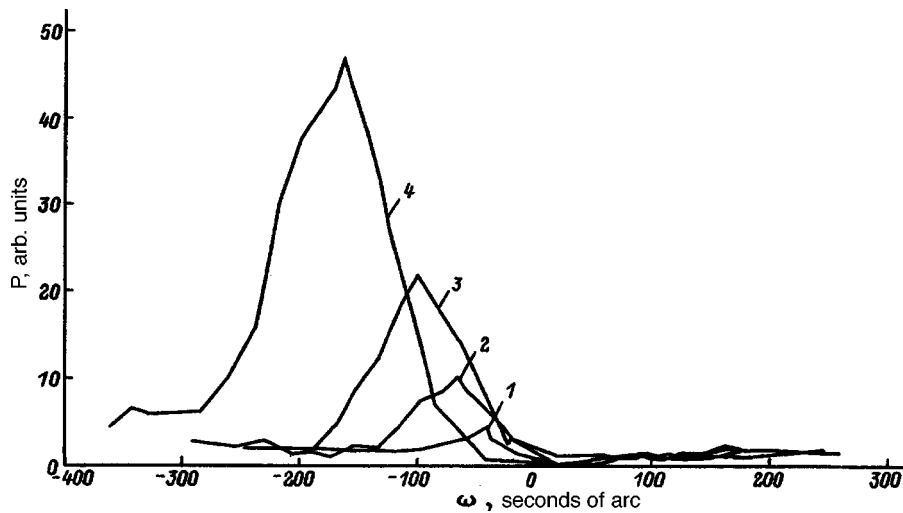
FIG. 2. Plots of the reduced intensity function $P(\omega)$. The notation corresponds to Fig. 1.

monochromator and the analyzer amounted to $10''$, which is close to the theoretical value.

EXPERIMENTAL RESULTS

A. Dependence of the structural damage on the dose. Samples with dose loads from 6.25×10^{14} to 6.25×10^{15} cm^{-2} were investigated to find the dose dependence of the structural damage in ion-implanted silicon. Figure 1 presents the ''tails'' of the corresponding diffraction reflection curves. The zero corresponds to the exact Bragg position of the reflection from the (111) plane of an undistorted crystal. There is an appreciable increase in the intensity of the diffraction reflection curve at small angles. Here the intensity increases with the dose, and a well resolved additional peak is observed for the largest dose. The triple-crystal x-ray diffraction spectra did not display diffuse peaks. Plots of the reduced intensity function $P(\omega) = I \cdot \omega^2 / P_{id}$ (I is the intensity of the principal peak, ω is the sample azimuthal angle, and $P_{id} = I^{id} \cdot \omega^2$ is the reduced intensity function for an ideal crystal and is approximately constant for all values of the sample azimuthal angle) constructed from the triple-crystal x-ray diffraction spectra are shown in Fig. 2. All the samples display a distinct maximum on the negative-angle side, whose intensity increases with the dose. In addition, the shift of the maxima toward negative angles increases. The position of the peaks on the plot coincides with the region of elevated intensity on the diffraction reflection curve at negative angles. The presence of a maximum on the plot of $P(\omega)$ attests to the occurrence of coherent scattering from a layer with an altered lattice constant. The mean relative strains were calculated from the positions of the maxima in Fig. 2 using the formula $\Delta d/d = \cot \Theta_b \cdot \Delta \Theta$ and are presented in Table I.

A plot of the dependence of the mean strain on the dose is presented in Fig. 3. Figure 4 presents the strain profile calculated using a program developed according to the method described in Ref. 8 for a sample irradiated with a dose equal to 6.25×10^{15} cm^{-2}. The program calculates the variation of the strain as a function of depth from the sample surface on the basis of an analysis of the additional peak on the diffraction reflection curve caused by coherent scattering

from a layer with an increased lattice constant. As is seen from the plot, the maximum value of $\Delta d/d$, which is approximately equal to 0.003, corresponds to a depth of 0.05 μm.

The effective thicknesses[9] of the damaged layers were also calculated for the irradiated samples: L_i from the half-widths of the maxima on the reduced intensity function, and L_p by the integral-characteristic method, without consideration of a region of $20''$ or $30''$ around 0. These data are presented in Table I.

B. Influence of annealing on the structure of ion-implanted silicon. To investigate the influence of annealing on the structure of boron-implanted silicon, we measured the diffraction reflection curves and triple-crystal x-ray diffraction spectra of samples which were irradiated with a dose of 1.875×10^{15} cm^{-2} and subjected to annealing in a nitrogen atmosphere at 300, 600, 800, 900, and 1000 °C with annealing times equal to 10, 60, and 120 min. Figure 5 presents the ''tails'' of the diffraction reflection curves for the samples subjected to annealing at 600 and 800 °C for 10 min. Annealing at the lower temperatures did not alter the form of the diffraction reflection curve shown in Fig. 1 for a dose equal to 1.875×10^{15} cm^{-2}. Increasing the temperature to 600 °C led to an increase in intensity on both the negative- and positive-angle sides. Annealing at 800 °C led to an appreciable drop on the negative-angle side.

The plots of the reduced intensity function $P(\omega)$ constructed from the triple-crystal x-ray diffraction spectra reveal the presence of a maximum on the negative-angle side (Fig. 6), which coincides with the region of elevated inten-

TABLE I. Values of the strain and thickness of the damaged layer as a function of the implantation dose.

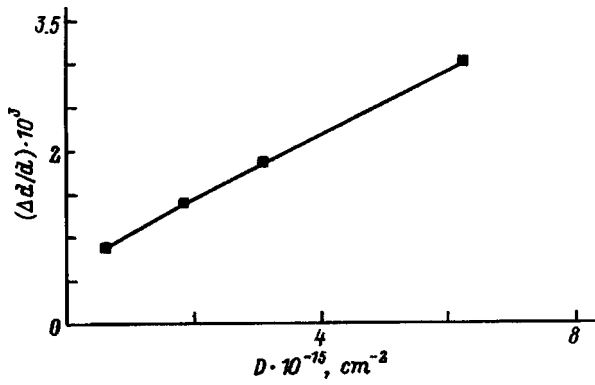| Dose (cm^{-2})	$\dfrac{\Delta d}{d} \times 10^3$	L_p [from $P(\omega)$], μm	$L_i(20)$, μm	$L_i(30)$, μm
6.25×10^{14}	0.8	0.16	0.043	0.031
1.875×10^{15}	1.2	0.14	0.067	0.054
3.125×10^{15}	1.9	0.14	0.11	0.09
6.25×10^{15}	2.9	0.13	0.12	0.11

FIG. 3. Dependence of the mean relative strain $\Delta d/d$ on the implantation dose D.
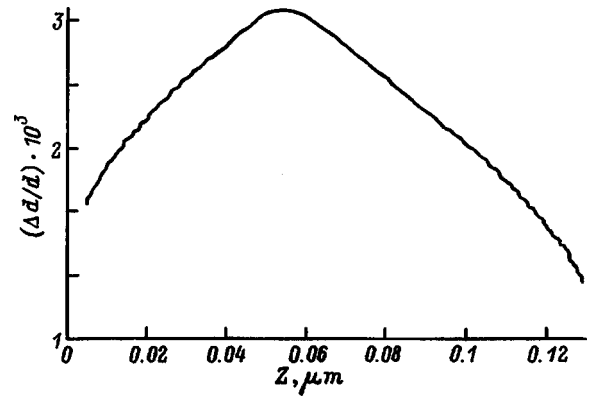

FIG. 4. Strain profile as a function of depth Z calculated from diffraction reflection curves.

sity on the corresponding diffraction reflection curve. At positive angles the plots of the reduced intensity function (for annealing temperatures equal to 300 and 600 °C) show a tendency to rise, which reflects the increase in intensity on the diffraction reflection curve (Fig. 5). However, annealing at 800 °C leads to the appearance of a maximum in this range of angles. The values of the mean strain $\Delta d/d$ calculated from the positions of the maxima on the plot of $P(\omega)$ are presented in Table II.

Thus, annealing at 800 °C led to the appearance of two regions with strains of different sign in the subsurface layer. It should also be noted that there was no diffuse peak on the triple-crystal x-ray diffraction spectra obtained in the analyzer scan mode for any of the samples subjected to heat treatment at temperatures from 300 to 900 °C.

Annealing at higher temperatures leads to further changes in the structure of the subsurface layer. At 1000 °C the triple-crystal x-ray diffraction spectra display a diffuse peak, which intensifies as the annealing time is increased. The plot of $P(\omega)$ for 1000 °C displays only one maximum, which attests to the presence of a layer with a negative mean strain (Table II), whose magnitude decreases.

Table II presents the values of L_i and L_p calculated for all temperatures. The values of L_i for the heavily annealed samples (1000 °C, 60 and 120 min) are not presented, since

the use of the integral method to estimate the thickness of the damaged layers is incorrect because of the presence of diffuse scattering, which makes a contribution to the diffraction reflection curve.

Plots of the dependence of the intensity of the diffuse peak on the sample azimuthal angle were constructed in log-log coordinates from the diffuse peak on the triple-crystal x-ray diffraction spectra for the samples annealed at 1000 °C for 120 min (Fig. 7). Each plot has two linear segments with slopes close to 1 and 3.

The most significant structural changes are observed at doses close to the doses ($\approx 10^{16}$ cm^{-2}) which cause amorphization of a subsurface layer under the implantation conditions employed. This can be observed for the samples implanted with boron in a dose equal to 6.25×10^{15} cm^{-2}. Figures 1 and 2 show the diffraction reflection curves and plots of $P(\omega)$ for this implantation dose. Annealing at 400 -700 °C for 10 min led to the appearance of one or two additional peaks on the diffraction reflection curve at angles smaller than the Bragg angles. Two maxima associated with coherent scattering are observed on the plot of $P(\omega)$ at negative sample azimuthal angles. After annealing at 800 °C for 10 min the diffraction reflection curve of the sample practically coincided with the analogous curve for an unimplanted
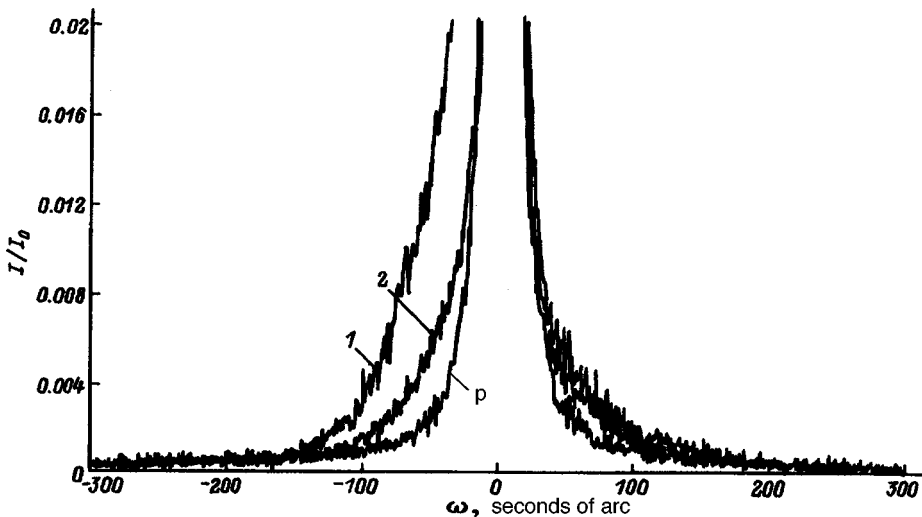

FIG. 5. Diffraction reflection curves of crystals implanted with a dose equal to 1.875×10^{15} cm^{-2}. Annealing temperature, °C: *1*—600, *2*—800 (the annealing time was 10 min); *P*—curve for a perfect crystal.
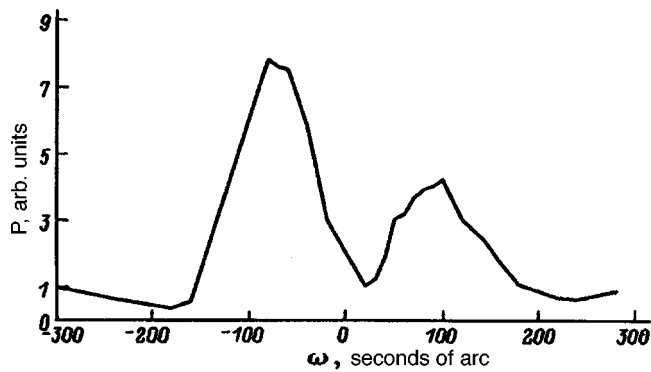
FIG. 6. Plot of the reduced intensity function $P(\omega)$. The implantation dose was 1.875×10^{15} cm^{-2}, the annealing temperature was 800 °C, and the annealing time was 10 min.

FIG. 7. Dependence of the logarithm of the intensity of the diffuse peak $\ln I_d$ on the logarithm of the sample azimuthal angle $\ln \omega$. The annealing temperature was 1000 °C, and the annealing time was 120 min; *1*—positive sample azimuthal angles, *2*—negative angles.

sample. In this case the plot of $P(\omega)$ did not display any peaks. A further increase in the temperature to 900 °C again led to an increase in the intensity on the diffraction reflection curve at large and small angles, and the plot of $P(\omega)$ shows four maxima (Fig. 8). There were no diffuse peaks on the triple-crystal x-ray diffraction spectra in the range of annealing temperatures 400–900 °C. Annealing at 1000 °C for 60 min led to the appearance of an intense diffuse peak.

DISCUSSION OF RESULTS

The presence of maxima on the plots of the reduced intensity function (Fig. 2) and the absence of a diffuse peak on the triple-crystal x-ray diffraction spectra are caused by coherent scattering in a subsurface layer with an increase in the lattice constant. This quite trivial result is associated with the generation of a large number of point defects upon the implantation of boron. Silicon ions displaced from regular positions into interstitial sites cause expansion of the lattice. As would be expected, the lattice strain increases with increasing dose (Fig. 3). The absence of a diffuse peak on the triple-crystal x-ray diffraction spectra indicates that no extended defects form in silicon following implantation by boron ions with doses up to 6.25×10^{15} cm^{-2}.

The thickness of the subsurface layer damaged by implantation, which was estimated from the half-width of the reduced intensity function, varies only slightly.

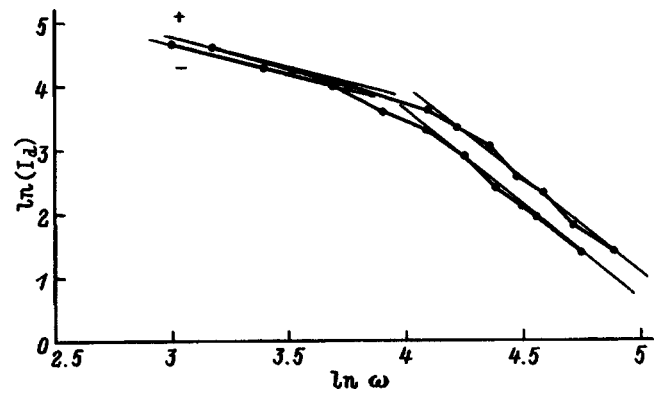The increase in the value of L_i calculated by the integral method can be attributed to the higher sensitivity of this method toward highly strained layers. The value of L_i is calculated from the difference between the areas under the diffraction reflection curves of the damaged and perfect crystals. Because the reflection coefficient of the latter crystal is higher than that of the ion-implanted crystal, the part of the diffraction reflection curve in the region of the Bragg angle must be excluded. The data in Table I demonstrate the critical nature of this exclusion. It should also be noted on the basis of the data in Table I that the total thickness of the damaged layer with ''large'' and ''small'' degrees of strain varies only slightly (0.20–0.25 μm) for all the samples except the sample irradiated with a dose equal to 6.25×10^{15} cm^{-2}. This can be attributed to amorphization of a certain part of the layer, which takes place in the region of the maximum on the defect distribution profile.

When the samples are annealed, restructuring of the defects in the subsurface structure should be expected. Annealing at 300 and 600 °C does not lead to significant changes in the x-ray diffraction pattern. However, annealing at 800 °C leads to the appearance of a layer with a negative strain in the subsurface region (Fig. 6) along with the layer with a positive strain formed as a result of the implantation of ions. Its formation can be attributed to a significant increase in the concentration of substituent boron ions in lattice sites. The layer with an increased lattice constant is probably located somewhat closer to the surface because of the stronger dis-

TABLE II. Variation of the strain and the thickness of the damaged layer as a result of annealing.

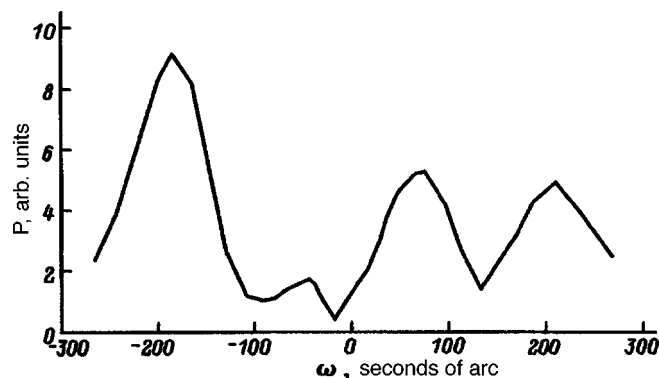| Temperature, °C | Annealing time, min | $\dfrac{\Delta d}{d} \times 10^3$ | | L_p, μm | | $L_i(20)$, μm | $L_i(30)$, μm |
		positive strain	negative strain	positive strain	negative strain		
300	10	1.2	⋯	0.13	⋯	0.079	0.059
600	10	1.4	⋯	0.14	⋯	0.070	0.050
800	10	1.4	−1.8	0.13	0.09	0.057	0.043
900	60	1.6	−0.97	0.12	0.14	0.045	0.009
1000	60	⋯	−0.38	⋯	0.18	⋯	⋯
1000	120	⋯	−0.29	⋯	0.26	⋯	⋯

FIG. 8. Plot of the reduced intensity function $P(\omega)$. The implantation dose was 6.25×10^{15} cm^{-2}, and the annealing temperature was 900 °C.

placement of the defect distribution toward it in comparison to the distribution of the implanted impurity.

The gradual decrease in the thickness of the layer with a positive strain revealed by the integral-characteristic method as the annealing temperature is increased from 300 to 900 °C is most likely attributable to the diffusion of interstitial silicon ions, which leads to spreading of the distribution of interstitial silicon, the passage of silicon ions into regular positions, and the occupation of lattice positions by boron ions, which actively displace silicon ions at high annealing temperatures.[10,11]

Heat treatment at 1000 °C leads to the disappearance of the layer with a positive strain due to the implantation of an even larger amount of boron in lattice sites. In addition, the diffusion of boron ions into the bulk leads to a decrease in the negative strain in the subsurface layer (Table II).

The presence of a diffuse peak on the triple-crystal x-ray diffraction spectra following high-temperature annealing attests to the association of point defects in extended defects.

The data presented in Fig. 7 show that, for the most part, defects of the dislocation-loop type make contributions to the diffuse peak for the crystal annealed at 1000 °C for 120 min. Dislocation loops are associated with stacking faults formed by interstitial silicon atoms.[12]

It should be stressed that defects of these types are not the only defects which make a contribution to the x-ray diffraction. Other crystal lattice imperfections of silicon in the form of rod-shaped defects, vacancy clusters, interstitial atoms, pores, cracks, tetrahedra of stacking faults, etc. are known.[11]

[1] A. Yu. Kazimirov, M. V. Koval'chuk, and V. G. Konn, Metallofizika, **9**(4), 54 (1987).

[2] V. Holy and J. Kubena, Czech. J. Phys. **32**, 750 (1982).

[3] M. Servidori and F. Cembaly, J. Appl. Crystallogr. **21**(5), 176 (1988).

[4] P. Zaumseil and U. Winter, Phys. Status Solidi A **120**, 67 (1990).

[5] V. A. Bushuev and A. P. Petryakov, Kristallografiya **40**, 1043 (1995) [Crystallogr. Rep. **40**, 968 (1995)].

[6] A. Iida and K. Kohra, Phys. Status Solidi A **51**, 533 (1979).

[7] A. M. Afanas'ev, P. A. Aleksandrov, and R. M. Imamov, *X-Ray Diffraction Diagnostics of Submicron Layers* [in Russian], Nauka, Moscow (1989), 152 pp.

[8] V. G. Kohn, M. V. Kovalchuk, R. M. Imamov, and E. F. Labonovich, Phys. Status Solidi A **64**, 435 (1981).

[9] V. A. Bushuev and A. P. Petrakov, Fiz. Tverd. Tela (St. Petersburg) **35**, 355 (1993) [Phys. Solid State **35**, 181 (1993)].

[10] G. A. Kachurin, I. E. Tyschenko, and M. Voelskow, Fiz. Tekh. Poluprovodn. **21**, 1193 (1987) [Sov. Phys. Semicond. **21**, 725 (1987)].

[11] J. W. Mayer, L. Eriksson, and J. A. Davies, *Ion Implantation in Semiconductors: Silicon and Germanium*, Academic Press, New York (1970); Mir, Moscow (1973), 296 pp.

[12] F. F. Komarov, A. P. Novikov, V. S. Solov'ev, and S. Yu. Shiryaev, *Structural Defects in Ion-Implanted Silicon* [in Russian], Universitetskoe Izd., Minsk (1990), 319 pp.

Translated by P. Shelnitz

Conversion of optical modes in an absorbing magnetogyrotropic waveguide

A. M. Shutyĭ and D. I. Sementsov

Ulyanovsk Branch of the M. V. Lomonosov Moscow State University, 432700 Ulyanovsk, Russia
(Submitted March 28, 1996)
Zh. Tekh. Fiz. **68**, 97–104 (June 1998)

The features of mode conversion in an absorbing magnetogyrotropic waveguide are investigated
by a coupled-wave approach. It is shown that absorption leads to an additional contribution
to the coupling of both identically and orthogonally polarized modes. A new waveguide regime for
mode conversion, in which there is no oscillatory energy exchange between modes, is
revealed. The possibility of controlling the damping of the total field in a waveguide by varying
the orientation of the magnetization in the waveguide layer is demonstrated. © *1998
American Institute of Physics.* [S1063-7842(98)01806-6]

INTRODUCTION

In most planar waveguides in use the losses associated with optical absorption in the film material are insignificant and are generally disregarded in theoretical analyses.[1–3] However, in waveguides based on epitaxial iron garnet films, the absorption in the near-IR range has values in the range $\alpha \sim 1–10 \text{ cm}^{-1}$, which cannot always be considered small.[2,4] The research reported in the literature deals mainly with the influence of absorption on waveguide eigenmodes, and the features of mode conversion in an absorbing waveguide has scarcely been analyzed. For example, mode damping coefficients were obtained on the basis of the ray approach in Ref. 5, and waveguide light propagation was investigated in Ref. 6 with consideration of the absorption in the metallic coating of the waveguide. The contributions from various absorption mechanisms to the damping coefficients of modes of different orders were considered in Refs. 2 and 7. Mode damping associated with the conversion of part of the energy of a waveguide mode into radiant energy was described in Ref. 8. Experimental results on mode conversion in a magnetogyrotropic waveguide with consideration of the complex character of the propagation constants of the modes were discussed in Ref. 9. Finally, an exact solution and numerical analysis of the problem of waveguide light propagation in an absorbing, transversely magnetized waveguide, for which the TE and TM modes are eigenmodes, were presented in Ref. 10. In the present work the features of mode conversion in an absorbing planar magnetogyrotropic waveguide are investigated on the basis of the coupled-mode theory for an arbitrary orientation of the magnetization in the waveguide layer.

ORTHOGONALITY RELATION FOR AN ABSORBING WAVEGUIDE

The orthogonality relations of modes play an important role in constructing waveguide solutions on the basis of the coupled-wave theory. Let us find the form that they take in the presence of absorption. We consider a planar waveguide structure consisting of a substrate, a waveguide layer, and a cladding. We direct the x axis perpendicularly to the inter-

faces between the layers. The $x=0$ and $x=-L$ planes separate the waveguide layers from the cladding and the substrate. The dielectric constant of the cladding (ε_1) and the substrate (ε_2) are assumed to be real, and the dielectric constant of the waveguide layer is assumed to be the complex quantity $\varepsilon = \varepsilon' - i\varepsilon''$. The magnetic permeabilities of all the layers for the optical range used are virtually equal to unity. The field of a mode propagating along the z axis with the complex propagation constant $\beta = \beta' - i\beta''$ is proportional to the factor $\exp(-i\beta z)$. It is easy to show (Ref. 3) that two different waveguide modes with the fields $\mathbf{E}_i$ and $\mathbf{H}_i$ obey the relation

$$\nabla \mathbf{b} \equiv \nabla (\mathbf{E}_1 \mathbf{H}_2^* + \mathbf{E}_2^* \mathbf{H}_1) = 2k_0 \varepsilon^* \mathbf{E}_1 \mathbf{E}_2^*, \qquad (1)$$

where $k_0 = \omega/c$, ω is the frequency of the radiation, and c is the speed of light in free space.

Separating the vector $\mathbf{b}$ and the operator ∇ into transverse and longitudinal components, we integrate (1) over a transverse cross section of the waveguide

$$\int \left(\frac{\partial b_z}{\partial z} + \nabla_t \mathbf{b}_t \right) ds = 2k_0 \varepsilon'' \int \mathbf{E}_1 \mathbf{E}_2^* ds, \qquad (2)$$

where $ds = dxdy$, $\nabla_t = (\partial/\partial x, \partial/\partial y, 0)$, and the integration is carried out over the cross section of the waveguide.

Replacing the integral over the area from the second term on the left-hand side of (2) by an integral over a contour enclosing the waveguide, and taking into account the exponential damping of the field of waveguide modes with increasing distance from the boundaries of the waveguide layer, we obtain

$$\int (\nabla_t \mathbf{b}_t) ds = \oint (\mathbf{b}_t \mathbf{e}_t) dl = 0, \qquad (3)$$

where $\mathbf{e}_t$ is a unit vector that is perpendicular to the integration contour.

With consideration of (3) we arrive at the following system of equations for the real and imaginary parts of the component b_z:

$$\int \, [(\beta_2'-\beta_1')b_z'-(\beta_2''+\beta_1'')b_z''-2k_0\varepsilon'' \, \text{Im}(\mathbf{E}_1\mathbf{E}_2^*)]ds=0,$$

$$\int \, [(\beta_2'-\beta_1')b_z''+(\beta_2''+\beta_1'')b_z'+2k_0\varepsilon'' \, \text{Re}(\mathbf{E}_1\mathbf{E}_2^*)]ds=0. \tag{4}$$

It follows from the relations obtained that only modes with orthogonal polarization $(\mathbf{E}_1\perp\mathbf{E}_2)$ are orthogonal, i.e., do not interact with one another in an absorbing scalar waveguide. The orthogonality relation $\int b_z ds=0$ holds for these modes. The orthogonality relation does not hold between different modes of the same polarization; therefore, a relation stipulated by the absorption of the waveguide should exist for them.

EQUATIONS OF COUPLED MODES

For a further analysis of the influence of absorption on mode coupling, a waveguide without absorption having a dielectric constant $\varepsilon=\varepsilon'$ should be taken as the unperturbed structure, and the imaginary part of the dielectric constant of the waveguide layer should be taken as the perturbation $\Delta\varepsilon(\alpha)=-i\varepsilon''$, where $\varepsilon''\approx\alpha\sqrt{\varepsilon'}/k_0$ and α is the absorption of the material. If the dichroism of TE and TM waves is taken into account, the perturbed part of the dielectric constant should be regarded as a diagonal tensor with nonidentical diagonal components, and the absorption will be different for modes with different polarization. In a magnetogyrotropic waveguide the perturbation of the dielectric constant is also determined by the orientation of the magnetization.[11] The total perturbation of the dielectric constant in this case can be represented in the following manner:

$$\Delta\hat{\varepsilon}(\alpha,\mathbf{M})=\Delta\hat{\varepsilon}(\alpha)+\Delta\hat{\varepsilon}(\mathbf{M}). \tag{5}$$

Assuming that there are no radiative modes, we expand the components of the electric field of the perturbed waveguide in the complete set of modes of the unperturbed waveguide:

$$E_j=\sum_\nu \, [A_\nu(z)\exp(-i\beta_\nu z)\pm B_\nu(z)\exp(i\beta_\nu z)]\mathscr{E}_j(x). \tag{6}$$

Here the upper signs in the square brackets refer to the transverse field components $(j=x,y)$; the lower signs refer to the longitudinal components $(j=z)$; $A_\nu(z)$ and $B_\nu(z)$ are the amplitudes of the forward and backward eigenmodes, which propagate in the $+z$ and $-z$ directions and vary along the waveguide as a result of mode coupling and perturbation of their propagation constants; and the $\mathscr{E}_j(x)$ are profile functions, which define the distribution of the field across the thickness of the waveguide. The coupling of counterpropagating modes is significant in waveguides with a periodically varying dielectric constant.[12] In the case considered here we shall confine ourselves to an analysis of the interaction of only the forward modes.

The representation of a field in a waveguide in the form (6) permits derivation of the following equation for the mode amplitudes:

$$A_\mu'=-ik_0 \exp(i\beta_\mu z)\int \, (\Delta\hat{\varepsilon}\mathbf{E})\mathscr{E}_\mu dx, \tag{7}$$

where the prime denotes the derivative with respect to the coordinate z and the integration is carried out over the a section of the waveguide layer of thickness L.

Of the complete set of modes, the strongest coupling is observed between the modes having the greatest phase synchronism. There are generally two such modes, and energy is transferred between them as they propagate in the waveguide. We write the coupling equations for these modes

$$A_\mu'=-i\Delta\beta_\mu A_\mu-i\gamma_{\mu\nu}A_\nu \, \exp[i(\beta_\mu-\beta_\nu)z],$$

$$A_\nu'=-i\Delta\beta_\nu A_\nu-i\gamma_{\mu\nu}^* A_\mu \, \exp[-i(\beta_\mu-\beta_\nu)z], \tag{8}$$

where the coupling coefficient $\gamma_{\mu\nu}=k_0\int\mathscr{E}_\mu^*\Delta\hat{\varepsilon}\mathscr{E}_\nu dx$.

The corrections to the propagation constants of the TE and TM modes are defined in the following manner

$$\Delta\beta_\nu^E=k_0\int \, \Delta\varepsilon_{yy}|\mathscr{E}_{\nu y}|dx,$$

$$\Delta\beta_\nu^M=k_0\int \, [\Delta\varepsilon_{xx}|\mathscr{E}_{\nu x}|+\Delta\varepsilon_{zz}|\mathscr{E}_{\nu z}|$$
$$+\mathscr{E}_{\nu x}^*\Delta\varepsilon_{xz}\mathscr{E}_{\nu z}+\mathscr{E}_{\nu z}^*\Delta\varepsilon_{zx}\mathscr{E}_{\nu x}]dx. \tag{9}$$

The solution of Eqs. (8) with the boundary conditions $A_\mu=A_\mu(0)$ and $A_\nu=0$, which were taken at $z=0$, have the following form:

$$A_\mu(z)=A_\mu(0)\left(\cos \chi_{\mu\nu}z-\frac{i\Delta_{\mu\nu}}{\chi_{\mu\nu}} \sin \chi_{\mu\nu}z\right)$$
$$\times\exp[i(\Delta_{\mu\nu}-\Delta\beta_\mu)z],$$

$$A_\nu(z)=-iA_\mu(0)\frac{\gamma_{\mu\nu}^*}{\chi_{\mu\nu}} \sin \chi_{\mu\nu}z \, \exp[-i(\Delta_{\mu\nu}+\Delta\beta_\nu)z],$$

$$2\Delta_{\mu\nu}=\beta_\mu+\Delta\beta_\mu-\beta_\nu-\Delta\beta_\nu, \quad \chi_{\mu\nu}^2=|\gamma_{\mu\nu}|^2+\Delta_{\mu\nu}^2. \tag{10}$$

With consideration of (10) the mode conversion efficiency $\eta_{\mu\nu}=|A_\nu(z)/A_\mu(0)|^2$ takes the form

$$\eta_{\mu\nu}=\left|\frac{\gamma_{\mu\nu}}{\chi_{\mu\nu}} \sin \chi_{\mu\nu}z\right|^2 \exp[-(\beta_\mu''+\beta_\nu'')z]. \tag{11}$$

DAMPING CONSTANT AND COUPLING OF IDENTICALLY POLARIZED MODES

The damping constant specifying the damping of a mode in a waveguide is found as the imaginary part of the perturbation of the propagation constant, i.e., as $\beta_\nu''=-\text{Im}\,\Delta\beta_\nu$. Figure 1 presents the dependence of the damping constant β_ν'' for the TE (solid curve) and TM (dashed curve) modes of the first three orders $\nu=0,1,2$ on the waveguide thickness L. The following parameters are used here and below in the calculations: $\varepsilon'=4.5371$, $\varepsilon_2=3.8$, $\varepsilon_1=1$; wavelength of the radiation in free space, $\lambda=1.15 \, \mu$m; damping parameter, $\alpha=1 \, \text{cm}^{-1}$. The symbols for the zeroth- and second-order modes in the figure show the solutions obtained numerically from the dispersion expressions found by matching the fields on the interfaces of the media. Both methods give coinciding
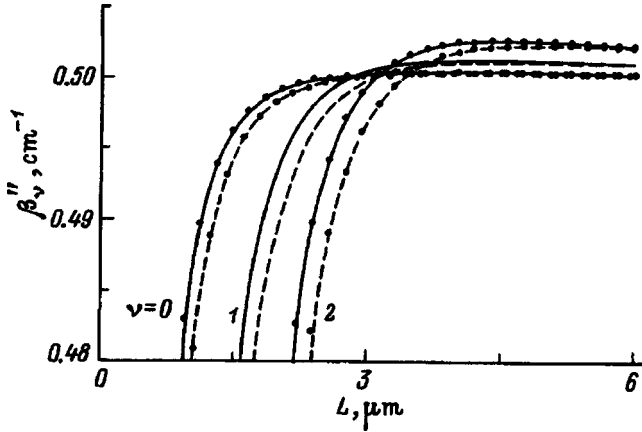
FIG. 1. Dependence of the damping constant β_ν'' on the waveguide thickness L.

values for $\beta_\nu''(L)$, attesting to the good approximation achieved using the coupled-mode formalism. It is seen from the figure that in the case of equality between the imaginary parts of the diagonal components of the dielectric tensor, the damping of the TM modes differs significantly from the damping of the TE modes only near the cutoff thickness, where the following relation holds:

$$\frac{(\beta_\nu^E)''}{(\beta_\nu^M)''} \approx \left(\frac{C_\nu^E \varepsilon' k_0 q_\nu^E}{C_\mu^M \beta_\nu^M h_\nu^E} \right)^2 \frac{v_\nu^E}{v_\nu^M}, \tag{12}$$

where the $C_\nu^{E,M}$ are normalized mode coefficients, $q_\nu^E = (\beta_\nu^2 - k_0^2 \varepsilon_1)^{1/2}$, $q_\nu^M = q_\nu^E \varepsilon'/\varepsilon_1$, $h_\nu^{E,M} = (d_0^2 \varepsilon' - \beta_\nu^2)^{1/2}$, and $v_\nu \equiv L + 2 \sin^2(h_\nu L)/q_\nu - \sin(2h_\nu L)/2h_\nu$.

Figure 2 presents the dependence of the damping constant β_ν'' of the zeroth ($\nu = 0$, curves *1* and *2*) and first ($\nu = 1$, curves *3* and *4*) TE modes on the absorption α of the waveguide-layer material for two waveguide thicknesses: $L = 1.5$ μm (curves *1* and *3*) and $L = 1$ μm (curves *2* and *4*). An analysis of these curves reveals that the linear dependence of $\beta_\nu''(\alpha)$ has its smallest slope at the mode cutoff
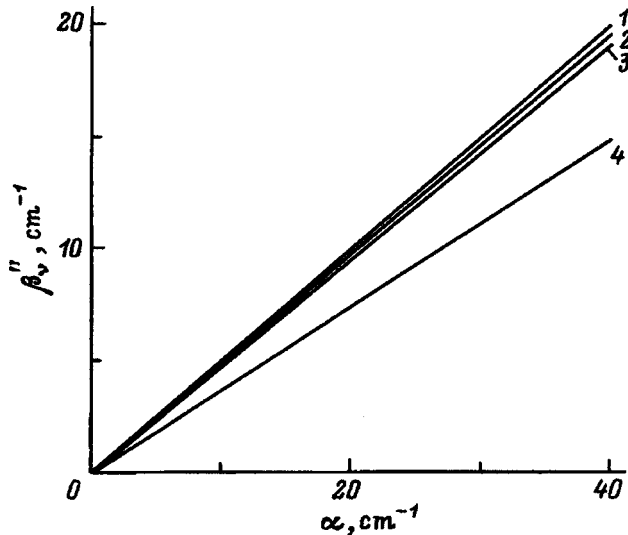

FIG. 2. Dependence of the damping constant β_ν'' on the absorption α for two different waveguide thicknesses.

thickness $L_\nu = L_{\nu 0}$ and that this dependence asymptotically approaches the straight line $\beta_\nu''(\alpha) = \alpha/2$ as the deviation from that thickness increases.

The coupling coefficients of the identically polarized modes are defined by the expressions

$$\gamma_{\mu\nu}^{EE} = k_0 \int \Delta\varepsilon_{yy} \mathscr{E}_{\mu y}^* \mathscr{E}_{\nu y} dx,$$

$$\gamma_{\mu\nu}^{MM} = k_0 \int [\mathscr{E}_{\mu x}^* (\Delta\varepsilon_{xx} \mathscr{E}_{\nu x} + \Delta\varepsilon_{xz} \mathscr{E}_{\nu z})$$
$$+ \mathscr{E}_{\mu z}^* (\Delta\varepsilon_{zz} \mathscr{E}_{\nu z} + \Delta\varepsilon_{zx} \mathscr{E}_{\nu x})] dx. \tag{13}$$

It follows from (5) that the perturbation of the dielectric constant contains a component which depends on the absorption α and leads to the coupling of identically polarized modes even in a scalar waveguide, i.e., in the absence of magnetization. This is confirmed by the orthogonality relations obtained above (4). When $\alpha \sim 1 - 10$ cm^{-1}, $\Delta\varepsilon(\alpha)$ is of the same order as the perturbation which is quadratic with respect to the magnetization in the diagonal terms of the dielectric constant in the iron garnet film. Therefore, the contribution of the absorption introduced to the coupling of identically polarized modes is comparable to the coupling caused by the magnetization,[11] but the large phase mismatch of modes of different orders greatly weakens the mode interaction. The mode coupling coefficient appearing as a result of absorption is of the order of $|\gamma_{\mu\nu}| \sim 10^{-2}$ cm^{-1} for $\alpha = 1$ cm^{-1}, while the phase mismatch $\Delta'_{\mu\nu} \sim 10^2$ cm^{-1}. For a mismatch $\Delta'_{\mu\nu} \gg |\gamma_{\mu\nu}|$ the mode conversion efficiency $\eta_{\mu\nu} \lesssim |\gamma_{\mu\nu}|/\Delta'_{\mu\nu} \ll 1$. Therefore, in an absorbing waveguide, as in a transparent waveguide, identically polarized modes with different propagation constants ($\beta_\mu \neq \beta_\nu$) can be considered noninteracting modes, i.e., eigenmodes, of the waveguide in the absence of other (for example, periodic) perturbations of the dielectric constant.

COUPLING OF DIFFERENTLY POLARIZED MODES

In an absorbing waveguide, as in a transparent waveguide, orthogonally polarized modes are coupled only when the dielectric tensor has off-diagonal components. In a magnetogyrotropic waveguide the TE_μ and TM_ν modes are coupled with the coupling coefficient

$$\gamma_{\mu\nu} = k_0 \int \mathscr{E}_{\mu y}^* (\Delta\varepsilon_{yz} \mathscr{E}_{\nu z} + \Delta\varepsilon_{yx} \mathscr{E}_{\nu x}) dx. \tag{14}$$

To find the profile functions appearing in (14), we should take into account the perturbation of the dielectric constant of the waveguide layer and the propagation constants of the modes, which lead to a dependence of the coupling coefficient $\gamma_{\mu\nu}$ on the absorption. Figure 3 presents the dependence of the coupling coefficient which provides for $TE_\mu \rightarrow TM_\nu$ conversion in the first three orders ($\mu = \nu = 0, 1, 2$) on the thickness of the waveguide layer for orientation of the magnetization along the x axis and $\alpha = 20$ cm^{-1} (the dashed curves correspond to a transparent waveguide, $\alpha = 0$). It is seen that the difference between the values of $|\gamma_{\mu\nu}|$ for absorbing and transparent films increases as the deviation from the cutoff thickness increases, i.e., as
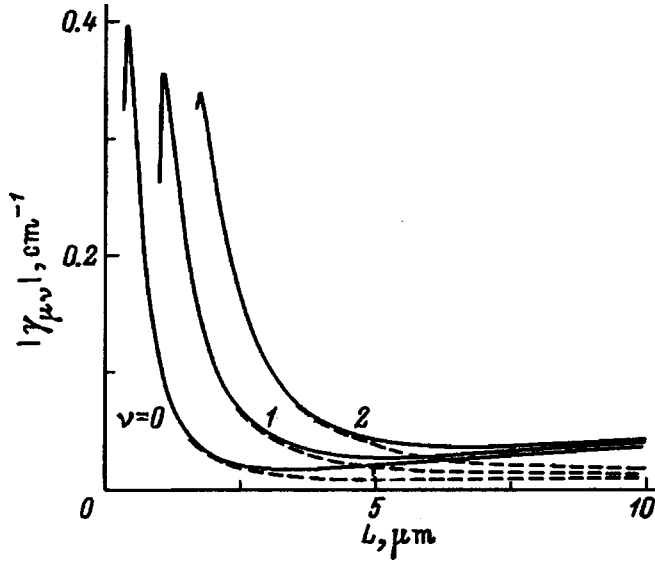
FIG. 3. Dependence of the coupling coefficient for $TE_\mu \to TM_\nu$ conversion on the thickness of the waveguide layer.

the film thickness increases. This is due to the fact that, as the deviation from the cutoff thickness increases, the localization of each mode in the film increases with resultant enhancement of the influence of absorption on the coupling coefficient, which is determined by the overlap of the profile functions in the layer with the perturbed dielectric constant. Deflection of the magnetic moment away from a normal toward the Faraday orientation leads to a significant (by more than two orders of magnitude) increase in $|\gamma_{\mu\nu}|$, while the difference $|\gamma_{\mu\nu}(\alpha)| - |\gamma_{\mu\nu}(0)|$ varies only slightly; therefore, it can be stated that the influence of absorption on the coupling of modes, i.e., on $(|\gamma_{\mu\nu}(\alpha)| - |\gamma_{\mu\nu}(0)|)/|\gamma_{\mu\nu}(0)|$, decreases. Here and in the following we shall use a magnetogyrotropic waveguide with the [111] crystallographic axis normal to the surface of the film for the calculations; its linear and quadratic magnetooptical parameters have the following values: $f = 3.07 \times 10^{-4}$, $\Delta g = -0.73 \times 10^{-4}$, $g_{11} = 5.07 \times 10^{-4}$, and $g_{44} = 2.4 \times 10^{-4}$.

We bring the expression for the mode conversion efficiency (11) into the form

$$\eta_{\mu\nu} = \left| \frac{\gamma_{\mu\nu}}{\chi_{\mu\nu}} \right|^2 [\sin^2(\tau z \cos \psi) + \sinh^2(\tau z \sin \psi)]$$

$$\times \exp[-(\beta_\mu'' + \beta_\nu'')z],$$

$$\tau = (\chi_{\mu\nu} \chi_{\mu\nu}^*)^{1/2}, \quad 2\Delta_{\mu\nu}' = \beta_\mu - \beta_\nu + \mathrm{Re}(\Delta\beta_\mu - \Delta\beta_\nu),$$

$$2\Delta_{\mu\nu}'' = \beta_\nu'' - \beta_\mu'',$$

$$\psi = \frac{1}{2} \arctan[2\Delta_{\mu\nu}' \Delta_{\mu\nu}'' (\Delta_{\mu\nu}'^2 + |\gamma_{\mu\nu}|^2 - \Delta_{\mu\nu}''^2)^{-1}]. \quad (15)$$

While the conversion efficiency for a transparent waveguide is a periodic function, the conversion efficiency for a waveguide with absorption is a decreasing oscillating function. When the condition $|\Delta_{\mu\nu}''| \geq |\gamma_{\mu\nu}|$ holds for $\Delta_{\mu\nu}' \approx 0$, the conversion efficiency ceases to be an oscillating function. This effect is similar to the effect considered in a system of

coupled waveguides with strongly different absorption properties.[13] An analysis of (15) shows that the conversion efficiency reaches its maximal values at the waveguide lengths $z \approx \pi(1/2 + n)/(\tau \cos \psi)$ $(n = 0,1,2,\dots)$ and its minimal values at the lengths $z \approx \pi n/(\tau \cos \psi)$, where, in contrast to the situation in a transparent waveguide, for unsynchronized modes the intensity of the νth mode $I_\nu = |A_\nu(z)|^2$ is nonzero:

$$I_\nu^{(\min)} = \left| A_\mu(0) \frac{\gamma_{\mu\nu}}{\chi_{\mu\nu}} \right|^2 \sinh^2(\pi n \tan \psi)$$

$$\times \exp[-(\beta_\mu'' + \beta_\nu'')\pi n/\tau \cos \psi]. \quad (16)$$

Hence it follows that in an absorbing waveguide the polarization of the total field of two coupled modes for $\psi \neq 0$ and any waveguide length differs from the polarization of the field of the input mode. In the case of complete phase synchronism ($\Delta_{\mu\nu}' = 0$) the expressions for the intensities of the input mode $I_\mu = |A_\mu(z)|^2$ and the intensity of the excited mode I_ν take the form

$$I_\mu = |A_\mu(0)|^2 \left(\cos \chi_{\mu\nu} z + \frac{\Delta_{\mu\nu}''}{\chi_{\mu\nu}} \sin \chi_{\mu\nu} z \right)^2$$

$$\times \exp[-(\beta_\mu'' + \beta_\nu'')z],$$

$$I_\nu = |A_\mu(0)|^2 \left(\frac{|\gamma_{\mu\nu}|}{\chi_{\mu\nu}} \sin \chi_{\mu\nu} z \right)^2 \exp[-(\beta_\mu'' + \beta_\nu'')z], \quad (17)$$

where $\chi_{\mu\nu} = (|\gamma_{\mu\nu}|^2 - \Delta_{\mu\nu}''^2)^{1/2}$.

The energy of the νth mode in a waveguide of length $z = \pi n/\chi_{\mu\nu}$ is equal to zero, while the energy of the μth mode remains nonzero over the entire course of the waveguide. If the condition $|\Delta_{\mu\nu}''| \ll |\gamma_{\mu\nu}|$ holds at the minima at the lengths $z = \pi(1/2 + n)/\chi_{\mu\nu}$ $(n = 0,1,2,\dots)$, the intensity of the input mode takes the values

$$I_\mu^{(\min)} = |A_\mu(0)|^2 \left(\frac{\Delta_{\mu\nu}''}{\chi_{\mu\nu}} \right) \exp\left[-\pi\left(\frac{1}{2} + n \right)(\beta_\mu'' + \beta_\nu'')/\chi_{\mu\nu} \right]. \quad (18)$$

To analyze mode coupling it is useful to introduce the quantity $r_{\mu\nu}(z) = |A_\nu(z)/A_\mu(z)|^2$, which characterizes the contribution of each mode to the intensity of the total field in the waveguide. The form of the function $r_{\mu\nu}(z)$ is largely determined by the difference $\Delta_{\mu\nu}''$ between the damping parameters of the modes. In real waveguide structures the absorption of modes of different polarization can differ by several fold even far from the cutoff thickness.[9] The parameter $|\Delta_{\mu\nu}''|$ increases significantly when waveguides with a cladding of a conducting material or waveguides obtained by proton exchange are used.[3,14]

Figure 4 presents plots of $r_{\mu\nu}(z)$ for $TE_0 \to TM_0$ conversion and various values of $\Delta_{\mu\nu}''$ and the waveguide thicknesses $L = 3.6$ (a), 3.7 (b), and 2.0 μm (c). The difference between the mode damping parameters takes the values $\Delta_{\mu\nu}'' = -0.25$, -0.125, 0, 0.125, and 0.25 cm^{-1} (Fig. 4a, curves *1–5*), $\Delta_{\mu\nu}'' = -0.25$ cm^{-1} (Fig. 4b), and $\Delta_{\mu\nu}'' = -0.25$, -0.5, -0.75, and -1 cm^{-1} (Fig. 4c, curves *1–4*). The growth anisotropy of the iron garnet film was taken into
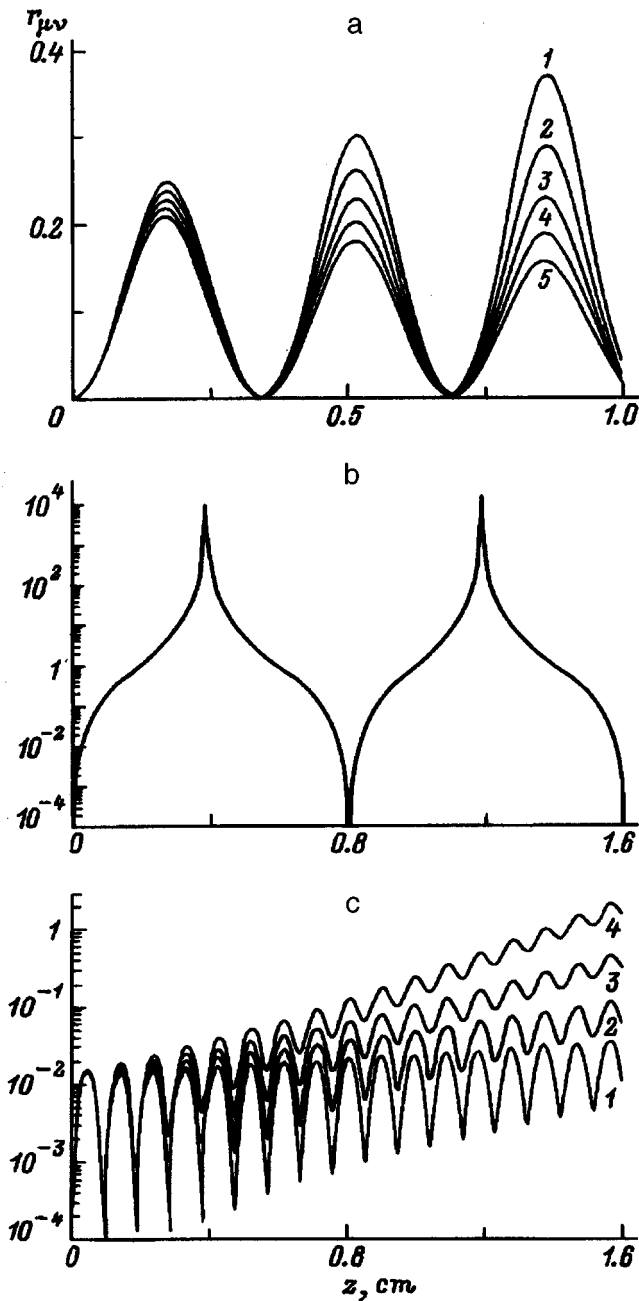
FIG. 4. Plot of $r_{\mu\nu}(z)$ for TE$_0 \to$ TM$_0$ conversion.

$r_{\mu\nu}(z)$ curve with increasing maxima was obtained as a result of an investigation of TE$_0 \to$ TM$_0$ conversion for $\Delta''_{\mu\nu} \approx 0.29$ cm^{-1}. However, our analysis (Fig. 4a) shows that the maxima of $r_{\mu\nu}(z)$ should obey a dependence which decreases with the waveguide length in this case. When there is complete phase synchronism between the modes (Fig. 4b) and the condition $|\Delta''_{\mu\nu}| < |\gamma_{\mu\nu}|$ holds, the intensity of the μth input mode reaches a minimum, and the function $r_{\mu\nu}(z)$ takes very large values at certain waveguide lengths. The polarization of the total field approximates the polarization of the νth mode in this case. The strong mode mismatch at the waveguide thickness $L = 2$ μm (Fig. 4c) makes the maxima more frequent and dramatically lowers their height. Also, if $\beta''_\mu > \beta''_\nu$, as the difference between the damping parameters of the input and excited modes $|\Delta''_{\mu\nu}|$ increases, the contribution of the excited mode to the total intensity increases, rising with the waveguide length z.

NONOSCILLATING INTERMODE INTERACTION REGIME

In the case of phase synchronism ($\Delta'_{\mu\nu} = 0$) and a fairly large difference between the damping parameters of the coupled modes ($|\beta''_\mu - \beta''_\nu| \geqslant 2|\gamma_{\mu\nu}|$), the trigonometric functions in (16) transform into hyperbolic functions. As a result, the periodic energy exchange between modes propagating in a waveguide that takes place in a transparent waveguide and in the case of a small difference between the mode damping parameters is not observed. In this case the intensity of the νth excited mode $I_\nu = |A_\nu(z)|^2$ has one maximum at the waveguide length

$$z_1 = \frac{1}{2\sigma_{\mu\nu}} \ln\left(\frac{\beta''_\mu + \beta''_\nu + 2\sigma_{\mu\nu}}{\beta''_\mu + \beta''_\nu - 2\sigma_{\mu\nu}} \right), \tag{19}$$

and the intensity of the μth input mode I_μ has one minimum and one maximum at the waveguide lengths

$$z_2 = \frac{1}{2\sigma_{\mu\nu}} \ln\left(\frac{\delta^-_{\mu\nu}}{\delta^+_{\mu\nu}} \right),$$

$$z_3 = \frac{1}{2\sigma_{\mu\nu}} \ln\left(\frac{(\beta''_\mu + \beta''_\nu + 2\sigma_{\mu\nu})\delta^-_{\mu\nu}}{(\beta''_\mu + \beta''_\nu - 2\sigma_{\mu\nu})\delta^+_{\mu\nu}} \right), \tag{20}$$

respectively, where

$$\sigma_{\mu\nu} = (\Delta''^2_{\mu\nu} - |\gamma_{\mu\nu}|^2)^{1/2} \quad \text{and} \quad \delta^\pm_{\mu\nu} = \Delta''_{\mu\nu} \pm \sigma_{\mu\nu}.$$

It is significant that the interaction of modes results in changes in their damping, which is determined for the portion of the waveguide at $z > z_3$ by the quantity

$$\bar{\alpha}_{\mu,\nu} = 2(\beta''_\mu + \beta''_\nu \pm \sigma_{\mu\nu}). \tag{21}$$

This permits regulation of the mode damping $\bar{\alpha}_{\mu,\nu}$ in a certain range by varying the coupling coefficient. In addition, the total intensity $I_\mu + I_\nu$ at the waveguide exit can be regulated by varying the coupling coefficient, although this could not be done when identically absorbed modes are coupled. The results presented are also valid for a small phase mis-

account in constructing the curves in Fig. 4b by adding 1.2 $\times 10^{-3}$ to the diagonal component $\Delta\varepsilon_{xx}$.[11] As a result of this anisotropy, the TE$_0$ and TM$_0$ modes are completely synchronized, i.e., $\Delta'_{00} = 0$, when $L \approx 3.7$ μm. It follows from the plots presented that if the absorption of the μth input mode is greater than the absorption of the νth excited mode, the maxima and minima of $r_{\mu\nu}$ increase with the waveguide length. This means that the contribution of the νth mode to the field propagating in the waveguide increases with the distance z. When there is no dichroism, the height of the maxima of $r_{\mu\nu}$ does not vary, but the minima take a zero value, at which the field takes on the original polarization. If the absorption of the input mode is weaker than the absorption of the excited mode, the height of the maxima decreases with only a slight increase in the minima of $r_{\mu\nu}$. In Ref. 9 an

match $2\Delta'_{\mu\nu} < |\gamma_{\mu\nu}|$. Under the conditions $\Delta'_{\mu\nu} = 0$ and $\Delta''_{\mu\nu} = |\gamma_{\mu\nu}|$ the expressions for the intensities of the modes take the form

$$I_\mu = I_0 (z\Delta''_{\mu\nu} + 1)^2 \, \exp[-(\beta''_\mu + \beta''_\nu)z],$$

$$I_\nu = I_0 (z\Delta''_{\mu\nu})^2 \, \exp[-(\beta''_\mu + \beta''_\nu)z], \tag{22}$$

where $I_0 = |A_\mu(0)|^2$, and the intensity maxima of the νth and μth modes and the minimum of the μth mode, respectively, are realized at the lengths

$$z_1 = 2/(\beta''_\mu + \beta''_\nu), \quad z_3 = 2\beta''_\mu/(\beta''^2_\mu - \beta''^2_\nu),$$

$$z_2 = 1/(\beta''_\mu - \beta''_\nu).$$

If the μth input mode damps more slowly than the excited mode, its intensity does not have extremum values and decays monotonically.

As follows from an analysis of the relations presented, in the absence of intermode coupling the possibilities of regulating the energy damping in a waveguide are increasingly better the more strongly the absorption coefficients of the modes differ. To obtain a quantitative estimate of the effect described, let us consider conversion of the $TM_1 \leftrightarrow TE_1$ type in a magnetogyrotropic waveguide consisting of an iron garnet film on a gadolinium–gallium garnet substrate with a cladding of a conducting material. The conducting layer mainly influences the absorption of the modes by strongly enhancing the dichroism ($\beta''_M = -4.5$ cm^{-1} and $\beta''_E = -0.5$ cm^{-1}) and has practically no influence on other parameters of the modes.[14] The film thickness $L = 6.8$ μm, at which the TE_1 and TM_1 modes are synchronized ($\Delta'_{\mu\nu} \approx 0$) in the waveguide considered, was used for the calculations. Figure 5 presents the dependence of the relative (normalized to I_0, i.e., the intensity of the input mode at $z = 0$) intensity of the input mode $J_\mu(z) = I_\mu(z)/I_0$, (curves 1), the relative intensity of the excited mode $J_\nu(z)$ (curves 2), the total relative intensity $J_\mu(z) + J_\nu(z)$ (curves 3), and the relative intensity of the input mode in the absence of coupling $J_\mu(z) = \exp(-2\beta''_\mu z)$ ($\gamma_{\mu\nu} = 0$, curves 4) on the waveguide length z for $TM_1 \rightarrow TE_1$ (a) and $TE_1 \rightarrow TM_1$ (b) conversion. The magnetization lies in the plane of the film at the angle $\varphi = 59°$ to the z axis, at which the mode-synchronism condition ($\Delta'_{\mu\nu} \approx 0$) and the equality $\Delta''_{\mu\nu} = |\gamma_{\mu\nu}|$ hold. It follows from the dependences presented in Fig. 5a that the intensity of a strongly absorbed input mode interacting with a less strongly absorbed mode passes through the minimum value $J_\mu(z_2)$ and then damps more slowly than the intensity of the μth eigenmode. If the input mode is absorbed more slowly than the excited mode (Fig. 5b), mode coupling accelerates its damping. The total intensity is greater in the former case and less in the latter case than the intensity of the noninteracting mode.

When the magnetization deviates from the angle $\varphi = 59°$, the phase synchronism is violated to a slight extent ($\Delta'_{\mu\nu} = -2.13$ cm^{-1} for $\varphi = 0$ and $\Delta'_{\mu\nu} = 0.8$ cm^{-1} for $\varphi = 90°$), and the character of the mode interaction in this case is determined mainly by the relation of the coupling coefficient to the difference between the damping parameters $\Delta''_{\mu\nu}$. The total energy propagated in a waveguide can be effec-
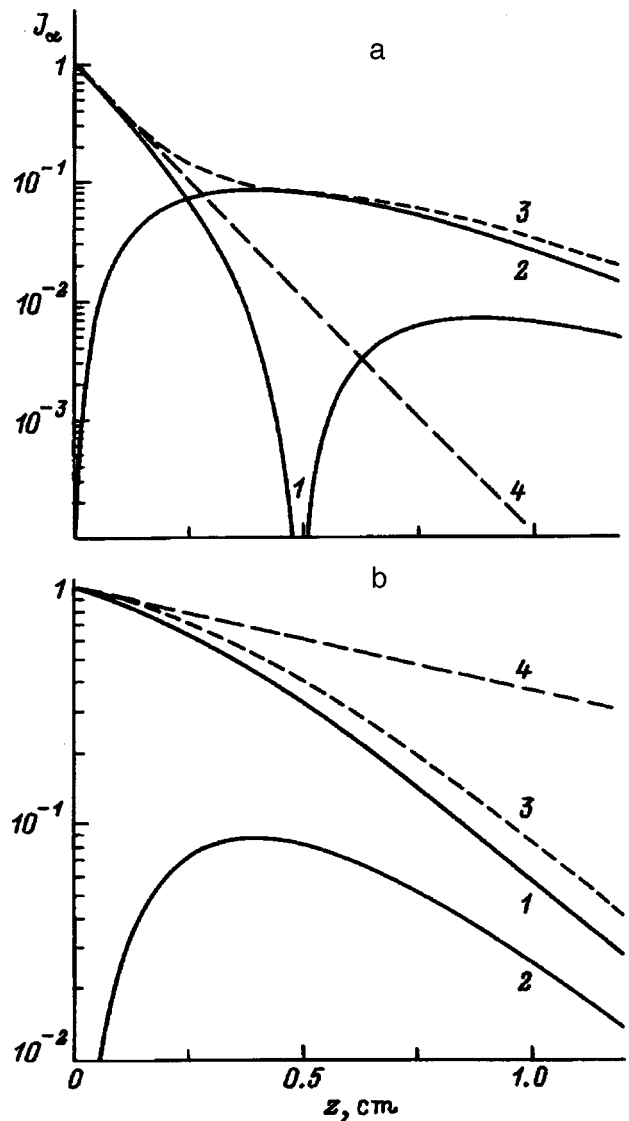

FIG. 5. Dependence on z of the relative intensities J_μ (1), J_ν (2), $J_\mu + J_\nu$ (3), and $J_\mu = \exp(2\beta''_\mu z)$ for $\gamma_{\mu\nu} = 0$ (4).

tively varied by altering the coupling coefficient as a result of rotation of the magnetic moment of the film. Figure 6 shows the dependence of the total relative intensity of the modes on the waveguide length for $TM_1 \rightarrow TE_1$ conversion and various orientations of the magnetic moment in the plane of the film: $\varphi = 0$, 30, 50, 59, 80, 85, and 90° for curves $1-7$, respectively. Under conditions close to phase synchronism of the coupled modes, the most effective control of the energy transferred in the waveguide is possible, if the coupling coefficient $|\gamma_{\mu\nu}|$ is smaller than $|\Delta''_{\mu\nu}|$ (curves $4-7$) or exceeds it only slightly (curve 3). When the coupling coefficient is increased further (curves 1 and 2), the possibilities of controlling the waveguide energy by varying $|\gamma_{\mu\nu}|$ are greatly reduced, and in the limiting case of $|\gamma_{\mu\nu}| \gg |\Delta''_{\mu\nu}|$ they vanish completely, since the energy of two synchronized modes damps according to an $\exp[-(\beta''_\mu + \beta''_\nu)z]$ law, being determined only by the imaginary parts of the propagation parameters of the modes. For large coupling coefficients ($|\gamma_{\mu\nu}| \gg |\Delta''_{\mu\nu}|$) variation of the damping of the total energy is possible only by significantly increasing the phase mismatch

strongly different damping parameters, the oscillatory energy exchange between the modes vanishes and the damping of the modes themselves becomes dependent on the coupling coefficient under conditions close to phase synchronism. Effective control of the damping of the field in the waveguide becomes possible under such conditions. The effect described can be utilized to create radiation amplitude modulators based on absorbing planar waveguides.
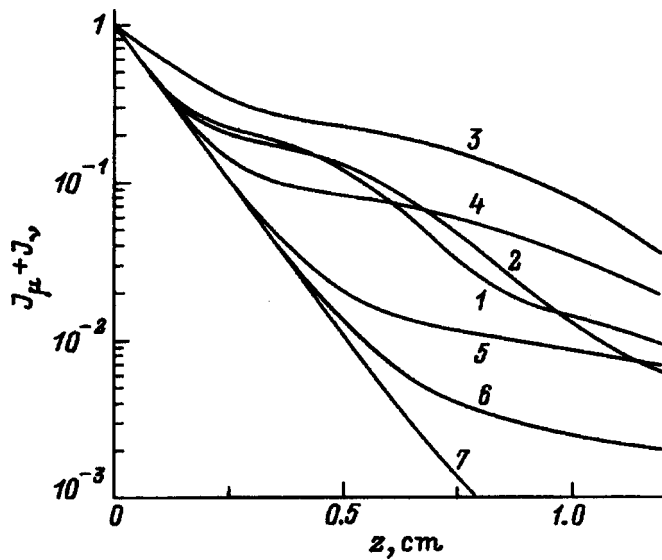

FIG. 6. Dependence of the total relative intensity of the coupled modes for TM_1-TE_1 conversion on the waveguide length.

$\Delta'_{\mu\nu}$, the mode coupling is then destroyed, and the energy transferred by the waveguide, consisting practically completely of the energy of the μth input mode, damps according to an $\exp[-2\beta''_\mu z]$ law.

The analysis performed here allows us to conclude that absorption in a magnetogyrotropic waveguide influences the interaction of both identically and orthogonally polarized modes, making an additional contribution to the intermode coupling and altering its character when certain conditions are satisfied. For example, for interacting modes with

[1] *Integrated Optics*, T. Tamir (Ed.) [Springer-Verlag, Berlin (1975); Mir, Moscow (1978), 575 pp.].
[2] A. M. Prokhorov, G. A. Smolenskiĭ, and A. N. Ageev, Usp. Fiz. Nauk **143**, 33 (1984) [Sov. Phys. Usp. **27**, 339 (1984)].
[3] D. I. Sementsov, Zh. Tekh. Fiz. **56**, 2157 (1986) [Sov. Phys. Tech. Phys. **31**, 1294 (1986)].
[4] A. K. Zvezdin and V. A. Kotov, *Magnetooptics of Thin Films* [in Russian], Nauka, Moscow (1988), 192 pp.
[5] P. V. Adamson, Opt. Spektrosk. **66**, 1172 (1989) [Opt. Spectrosc. (USSR) **66**, 684 (1989)].
[6] A. Reisinger, Appl. Opt. **12**, 1015 (1973).
[7] É. P. Stinser, A. N. Ageev, and S. A. Mironov, Pis'ma Zh. Tekh. Fiz. **3**, 913 (1977) [Sov. Tech. Phys. Lett. **3**, 373 (1977)].
[8] H.-G. Unger, *Planar Optical Waveguides and Fibres*, Clarendon Press, Oxford–New York (1978); Mir, Moscow (1980), 656 pp.
[9] G. Hepner, J. P. Castera, and B. Desormiere, Physica D (Amsterdam) **89**, 264 (1977).
[10] D. I. Sementsov and A. M. Shutyĭ, Opt. Spektrosk. **84**(2), 280 (1998) [Opt. Spectrosc. **84**, 238 (1998)].
[11] D. I. Sementsov, A. M. Shutyĭ, and O. V. Ivanov, Radiotekh. Elektron. **41**(3), 1 (1996).
[12] D. I. Semenotsov, Izv. Vyssh. Uchebn. Zaved. Fiz. No. **2**, 94 (1993).
[13] P. V. Adamson, Opt. Spektrosk. **69**, 453 (1990) [Opt. Spectrosc. (USSR) **69**, 269 (1990)].
[14] M. Adams, *An Introduction to Optical Waveguides* [Wiley, New York (1981); Mir, Moscow (1984), 512 pp.].

Translated by P. Shelnitz

Temperature stability and radiation resistance of holographic gratings on photopolymer materials

T. N. Smirnova, O. V. Sakhno, I. A. Strelets, and E. A. Tikhonov

Institute of Physics, Academy of Sciences of Ukraine, 252650 Kiev, Ukraine
(Submitted January 8, 1997)
Zh. Tekh. Fiz. **68**, 105–110 (June 1998)

Data are presented from studies of the temperature dependence of the diffraction efficiency and radiation resistance of volume phase hologram/transmission gratings. The reversible diffraction efficiencies are described by the phase equilibrium diagram for the polymer–diffusate system. The radiation resistance of these hologram/gratings is determined by the thresholds for photothermolytic decay of the diffusate, bromonaphthalene, that was used. Composites containing diffusates with high thresholds for photolysis and thermolysis are studied. As a result, modified versions of the photopolymer recording medium with radiation resistances exceeding $200 \, \mathrm{MW/cm^2}$ are proposed. © *1998 American Institute of Physics.*
[S1063-7842(98)01906-0]

INTRODUCTION

Photosensitive polymer composites (PPCs) for holographic recording are constantly being developed and improved.[1–6] The advances of recent years have led to new possibilities for photopolymers, such as the recording of negative holograms in real time[4,5] and a thermal technique for developing latent holographic images.[6] The known advantages of PPCs compared to bichromated gelatin include the ability to obtain holograms without having to go through the stages of chemical development of a latent image and extensive possibilities for optimizing the composites and achieving better reproducibility of the basic parameters. The phase character of the recorded information over a wide range of thickness and area of the recording medium makes these materials attractive for the fabrication and development of various kinds of holographic optical elements.

We have for sometime been using the basic photopolymer holographic recording composite FPK-488 and its modifications for fabricating various types of holographic optical elements. These include transmission and autocollimating diffraction gratings,[7] diffraction optical filters,[8] and selectively reflecting diffraction gratings.[9]

Applications of optical devices based on polymer holographic optical components in instruments or systems inevitably rely on the stability of the main operating parameters. The primary factor which affects photopolymer holographic optical elements is the temperature. Humidity or mechanical effects are not as important, since this composite does not swell in water and the optical element is a rigid structure consisting of a pair of substrates with the holograms between them. The temperature, on the other hand, is an unavoidable parameter, since local temperature rises may occur because of dissipative losses during channeling of high power laser beams.

In the first stage, the effect of general and local heating of holographic optical components on their spectral and angular characteristics is reversible. With increasing temperature, they may cause irreversible changes and destroy a hologram. Researchers encounter a similar situation when gratings interact with high-power laser beams. During interactions with high-power laser beams, a third stage may develop between these two, owing to the nonlinear optical response of the grating material (a cubic nonlinearity with respect to the field in an isotropic medium). This stage will be the subject of later studies; the results of a study of the first and second stages are presented in this paper.

One of the important practical achievements of this work has been the creation of a modification of the basic composite which has a higher radiation resistance than bichromated gelatin.

STRUCTURE AND PROPERTIES OF PHOTOSENSITIVE POLYMER COMPOSITE HOLOGRAPHIC MEDIA

Holographic media based on FPK-488 are made in the form of a triplex: substrate–PPC–substrate, with a gap ranging from a few microns to 1 mm, depending on the angular selectivity specifications for the fabricated holographic optical element. The gap is typically formed by spacers which do not inhibit shrinkage of the polymer layer. This kind of medium is prepared before recording and, for sizes up to 10 cm, presents no technical difficulty. PPCs of this class incorporate three interacting subsystems: a monomer–oligomer mixture which is capable of polymerizing; a mixture of two or more components which form an effective initiator for a radical polymerization process; and, a chemical component which plays a fundamental role in the diffusion-induced irreversible spatial modulation of the refractive index—a chemically neutral diffusate (CND).

Some characteristics of the photosensitive polymer composites are listed in Table I. FPK-488M differs from the basic FPK-488 composite[5] in the monomer–oligomer component, and FPK-488N differs in the diffusate.

The basis of the mechanism for holographic recording in these composites is the radical photopolymerization of the

TABLE I.

Type of PPC	Sensitivity,* mJ/cm^2	Resolution, mm^{-1}	Amplitude of the modulaion of the refractive index, δn**
FPK-488	300	>6000	0.012–0.015
FPK-488M	60	>6000	0.025
FPK-488N	150	>6000	0.012–0.014

*The photosensitivity of the recording medium is defined as the exposure required to attain a maximum diffraction efficiency for write beam intensities of 0.2 mJ/cm^2 at a writing wavelength of $\lambda = 0.488$ μm.

**The amplitude of the modulation of the refractive index was calculated for grating readout by a $\lambda = 0.6328$ μm beam with the electric field vector E parallel to the ''rulings'' (s polarization).

FIG. 1. Temperature-induced variations in the amplitude of the modulation of the refractive index of a grating based on FPK-488 (*1*), and in the diffraction efficiency of gratings based on FPK-488M (*2*) and FPK-488 (*3*).

monomer–oligomer mixture and interdiffusion of the monomer and the CND between the phase planes of the incipient hologram which correspond to the maximum and minimum light intensities in the interference light field.[10,11] The resulting spatial modulation in the concentration of the CND enhances, by many times, and stabilizes the modulation of the refractive index of the polymer layer.

TEMPERATURE STABILITY OF THE PARAMETERS OF FPK-488 HOLOGRAPHIC GRATINGS

We have studied the temperature dependences of the diffraction efficiency η and transmission T_0 of holographic gratings recorded on layers of FPK-488 and FPK-488M. For this purpose we measured the variation in the diffraction efficiency over temperatures of -50 to $+250$ °C. A transmission diffraction grating was placed in a temperature controlled cabinet with a temperature T that was controlled over the range $+20–+250$ °C. The diffraction efficiency was measured over the interval $T = -100$ to $+20$ °C using a liquid nitrogen cryostat with a variable temperature. The variation in the diffraction efficiency of the grating was tested using a He–Ne laser beam ($\lambda = 0.6328$ μm) incident on the sample grating at the Bragg angle. The powers of the incident (P_{in}), transmitted (P_{out}), and diffracted beams (P_{dif}) were monitored using FD-26 photodiodes whose outputs were fed to an automatic data recording and processing system.

The diffraction efficiency was defined as the ratio $P_{dif}/(P_{dif} + P_{out})$. Losses due to light scattering ($\rho = 1 - T_0$) were taken into account by measuring the grating transmission $T_0 = (P_{in} - P_{fr} - P_{out})/P_{in}$, where P_{fr} represents the Fresnel losses measured when the probe beam was incident on the grating at the angle for which $P_{dif} = 0$.

These studies yielded the following data. Heating the gratings to 100 °C causes a negligible change in the diffraction efficiency. On going to $T > 100$ °C, a reduction in the diffraction efficiency is observed, with an approach to a steady value which remains invariant for fixed T throughout the observation time (up to 8 h). Cooling the grating to the initial (room) temperature is accompanied by a return of the diffraction efficiency to its original value.

Figure 1 shows the diffraction efficiency $\eta = f(T)$ of transmission gratings for the FPK-488 and FPK-488M composites. A temperature-reversible recovery of the diffraction efficiency occurs up to temperatures T such that the grating
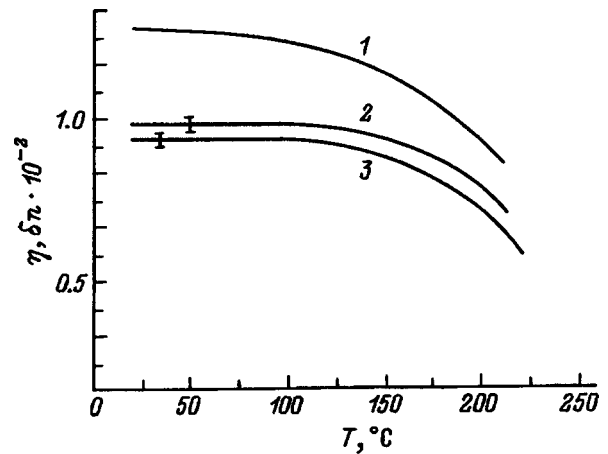
is destroyed. For gratings based on FPK-488, destruction sets in at $T = 218$ °C; for the gratings recorded on FPK-488M, at 244 °C.

A polymer layer without the neutral component remained undamaged up to the highest observation temperature, $T = 250$ °C. The damage for all the gratings (with $\nu = 2000$ mm^{-1}) had a specific and fixed character: cracks formed parallel to the direction of the phase planes of the grating, with a division ratio of roughly 20:1. The cracks occur periodically and thus convert the thick transmission grating into a thin one (according to the Klein–Cook criterion) and lead to the appearance of a multiwave diffraction pattern of the probe beam. Polymer layers of the same photopolymer recording material with no hologram recorded on them crack in a disordered fashion at the same temperatures.

As a grating is cooled to around -5 °C, its diffraction efficiency remains practically constant. At low temperatures ($T < 0$ °C), a reversible reduction in the diffraction efficiency is observed (Fig. 2). For gratings with FPK-488 with an initial $\eta = 98\%$, the minimum diffraction efficiency stabilizes at ~ 0.81. When a grating is heated at an average rate of ~ 0.5 deg/min, such that a quasiequilibrium temperature distribution can develop over the layer, the diffraction efficiency increases. The initial value is recovered at $T = +8$ °C. Hysteresis appears in the diffraction efficiency for gratings of FPK-488M on cooling and heating. Thus, when a grating (with an initial $\eta = 0.95$) is initially cooled to -60 °C and then heated, its efficiency at first increases but then, without reaching its original value, stabilizes at a level of ~ 0.8. However, the initial diffraction efficiency is recovered when the grating is kept $T = 20$ °C for a long time (over an hour) or heated to $+30–40$ °C. In addition, for gratings based on the two modified PPCs, the light scattering increases as the temperature is reduced.

The diffraction efficiency of a transmission volume phase grating at the Bragg angle β is given by the Kogelnik formula[12] for a given temperature

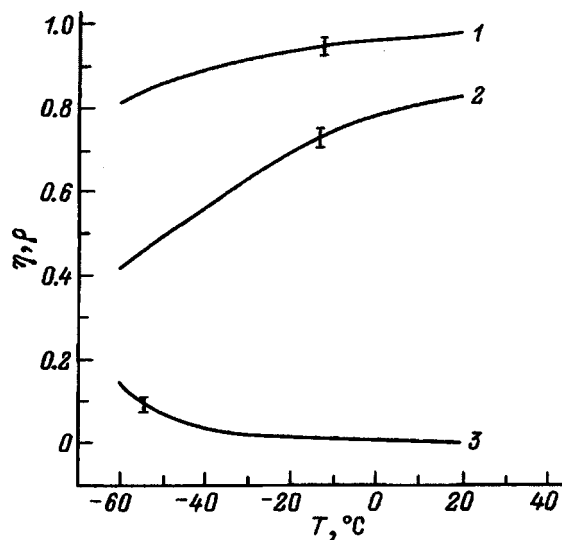$$\eta(T) = \sin[\pi \delta n(T) d(T)/\lambda \cos \beta(T)], \qquad (1)$$

FIG. 2. Temperature-induced variations in the diffraction efficiency (*1,2*) and light scattering losses (*3*) for gratings based on FPK-488 (*1,3*) and FPK-488M (*2*).
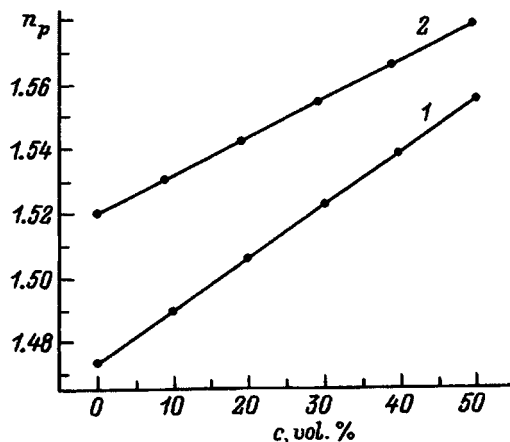

FIG. 3. The refractive indices of FPK-488 (*1*) and of a polymer based on it (*2*) as functions of the concentration of the chemically neutral diffusate (CND).

where δn is the amplitude of modulation of the refractive index and d is the grating thickness.

It might be expected that $\eta(T)$ is determined by temperature variations in the grating parameters which enter in Eq. (1).

As our measurements showed, the diffraction efficiency of the grating is essentially unchanged over temperatures of 1–100 °C. This appears to be related to the specific structural features of the holographic optical element and its thermal mechanics: because of the adhesion of the polymer to the substrate, the change in the size of the polymer layer along the grating vector is determined principally by the thermal expansion (contraction) of the glass substrate. For $T = 200$ °C and a coefficient of linear thermal expansion of the glass of $\alpha_c = 8 \times 10^{-6}$ deg^{-1}, the change in the grating period is roughly 0.07% and the change in the diffraction efficiency is 0.02%. For $T < 0$, α_c is smaller and the effect of this factor on the grating period is even more negligible.

The changes in the grating thickness is determined by the coefficient of linear thermal expansion of the polymer, α_p, and not of the substrate. The PPCs used here are new systems of polymer (more often, copolymer) and solvent (CND). There are no published data on their temperature coefficients of expansion. The estimates given here are based on values of α_p for polymers with similar properties. According to data[13] for polymers similar to those studied here, the maximum value of $\alpha_p = 5 \times 10^{-4}$ deg^{-1}. For this value of α_p, the grating thickness d increases by ~10% when T is raised to 200 °C.

The observed decrease of the diffraction efficiency (rather than an increase with grating thickness) is evidence that its temperature behavior is determined primarily by the dependence $\delta n(T)$, which decreases by more than the thickness d increases.

The temperature variation $\delta n(T)$ calculated from Eq. (1) on the basis of the measured $\eta(T)$, with allowance for the change in d for $\beta =$ const, is shown in Fig. 1 (curve *1*).

Let us see which properties of the recording material determine the magnitude and temperature behavior of $\delta n(T)$. By definition,

$$\delta n = |(n_{\max} - n_{\min})|/2, \tag{2}$$

where $n_{\max,\min}$ are the refractive indices of the polymer in the phase planes formed at the antinodes and nodes of the interference field, where the concentration of the CND is different.

Our measurements (Fig. 3) show that the dependence of the refractive index of the polymer on the amount of CND within this range of concentrations c is quite well described by a linear dependence of the form[14]

$$n = n_0 + c(dn/dc), \tag{3}$$

where dn/dc is the rate of increase of the refractive index and n_0 is the refractive index of the polymer without the neutral component.

Substituting Eq. (3) in Eq. (2) yields

$$\delta n(T) = |dn(T)/dc| c_{1N}(T), \tag{4}$$

where c_{1N} is the amplitude of the modulation of the concentration of the neutral component.

The absolute value of the rate of increase of the refractive index is known[14] to increase in proportion to the difference between the refractive indices of the polymer and solvent. The magnitude of Δc depends on the thermodynamic affinity of the neutral component of the polymer. (We shall discuss this in more detail below.) Thus, $\delta n(T)$ is determined by the thermodynamic properties of the polymer–CND system and by the difference in their refractive indices. The rate of increase of the refractive index is constant over a wide range of temperatures for various polymer–solvent systems.[13] The slight change in δn over 20–100 °C indicates that the change in $|dn(T)/dc|$ is small for the PPCs studied here. Based on this, we may assume that the character of the temperature dependence of $\delta n(T)$ is determined mainly by that of $\Delta c(T)$.

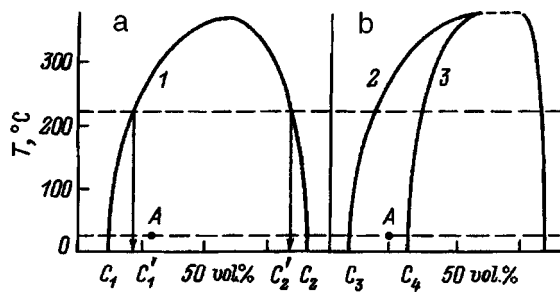The temperature dependence of $\Delta c = f(T)$ is explained by a change in the phase equilibrium of the polymer–CND

FIG. 4. Phase diagrams of the states of a polymer-CND system for two models described in the text (*c* is the concentration.)

system. The phase equilibrium of such a system is described by a state diagram. In the following, we shall consider two models which are capable of describing $\Delta c = f(T)$ qualitatively (Fig. 4a and 4b). We consider phase equilibrium curves with an upper critical temperature of displacement T^* as the most common type of diagram for systems similar to those studied here.[13,15]

Our measurements imply that for FPK-488 and FPK-488M, T^* lies above the temperature at which the holograms are destroyed. Prior to recording, these composites are multicomponent single phase solutions. Polymerization in the gradient interference field destroys the equilibrium state of the system and shifts it to point A (Fig. 4a). Here system *1* breaks up into two phases. When equilibrium is reached, the compositions of the phases correspond to the points C_1 and C_2. The phase planes of the gratings in this case differ in having different amounts of these phases.

Diagrams *2* and *3* of Fig. 4b allow for the difference in solubility of the polymer formed at the antinodes and nodes of the interference field. Such a difference may be caused by a dependence of the degree of polymerization (density factor of the polymer network) and, therefore, of the solubility of the polymer, on the spatial distribution of the intensity of the light field.[16]

During the recording process, at the antinodes of the field the system becomes two-phased, and microsyneresis (squeezing out) of the excess neutral component from the polymer mass takes place. For the polymer formed at the nodes of the field, the concentration of the neutral component remains in equilibrium. As a result, the phase planes of the grating consist of polymer phases with different equilibrium concentrations of the CND (C_3 and C_4 in Fig. 4b).

For the system described by curve *1*, an elevated temperature causes the thermodynamic affinity of the CND to the polymer to increase and, accordingly, the compositions of the phases move closer together (points C_1' and C_2'). Then $\Delta c(T)$ and $\delta n(T)$ decrease, which leads to a reduction in the diffraction efficiency. In the second case (Fig. 4b), the equilibrium concentration of the neutral component in both polymer phases increases, with an accompanying drop in the concentration gradient of the CND and, therefore, a reduction in δn. In both cases, the change in $\Delta c(T)$ is reversible, so that the diffraction efficiency recovers when the grating is cooled.

The choice of a specific model for the holographic recording process in these photosensitive polymer composite requires additional study of their thermodynamic properties.

The situation is different when the gratings are cooled below 20 °C. The monotonic drop in the diffraction efficiency can be explained by a decrease in the thickness of the grating with temperature, since c_{1N} changes little according to the phase diagram. In addition, the drop in the diffraction efficiency may be related to increased light scattering in the grating (Fig. 2, curve *3*). This last effect may be caused by crystallization of the CND, whose freezing point is $-6\,°C$. The further drop in the diffraction efficiency as the temperature is brought below the freezing point of the CND and the above-described hysteresis may be related to a change in the diffusion times of the system owing to the high viscosity of the polymer phase and the peculiarities of the crystallization kinetics of the neutral component in the polymer matrix.

RADIATION RESISTANCE OF POLYMER HOLOGRAPHIC GRATINGS

The radiation resistance of the photosensitive polymer composite gratings was studied by irradiating them with pulsed Nd^{3+}:YAG ($\lambda = 1.064,\ 0.532\ \mu m$) and dye ($\lambda = 0.630\ \mu m$) lasers. The transmission and diffraction efficiency of the gratings and their thresholds for *N*-pulse and many-pulse damage were measured. The threshold for *N*-pulse damage is defined as the intensity (I_N) at which *N* pulses on a given point of the sample will produce microscopic damage on its surface and in its volume. The multipulse damage threshold is defined as the intensity I^* which will cause damage for $N > 1000$ pulses.

The way the diffraction efficiency and transmission T_0 varied was similar for laser irradiation at different wavelengths, intensities, and pulse repetition rates. As a rule, the diffraction efficiency and transmission of a grating are observed to decrease in an irradiated region (the latter is measured at orientations off the Bragg angle) all the way up until destruction of the grating. Visually, a blackening of the grating is observed where later a microscopic hole is formed owing to evaporation of the polymer. Here the transmission of a polymer layer without a neutral component does not change until it is destroyed at $I^* \sim 1\ GW/cm^2$.

Our results indicate that the damage threshold for holographic optical components made of these photosensitive polymer compositions is determined primarily by the photophysical and thermal properties of the CND component of the polymer. Here the processes leading to destruction of the grating have a cumulative character. A reduction in the intensity is observed as the frequency of the laser light approaches the edge of the fundamental absorption band of the diffusate.

Without pretending to an exhaustive description of the damage mechanism, we may assume that under intense irradiation, photochemical processes develop in the polymer owing to multiphoton absorption. The consequence of these processes is the formation of photodissociation products of the component with the lowest photodissociation threshold. The products of photodissociation that exhibit fundamental absorption at the wavelength of the high-power laser are extremely important, since in this case there is positive feedback in the chain of events formed by absorption–photoly-

TABLE II.

No	CND	n_d	$n_p - n_d$	C_{opt}, vol.%	l^*, MW/cm^2	Diffraction efficiency, $\lambda = 0.6328\ \mu m$
1	Bromonaphthalene	1.66	−0.14	30	16.0	0.99
2	Quinolene	1.627	−0.107	35	12	0.85
3	Petachlor Diphenyl	1.636	−0.116	30	15	0.9
4	Trifluorethanol	1.29	0.23	35	>200	0.94
5	Acetonitryl	1.344	0.176	40	>200	0.04
6	Methanol	1.328	0.192	40	>200	0.03
7	Heptane	1.387	0.133	12	>200	0.98
8	Hexane	1.375	0.145	20	>200	0.98

Note: The multi-pulse damage threshold I^* was measured using light from a Nd laser with the following parameters: $\lambda = 1.06\ \mu m$, pulse length 10–15 ns, pulse repetition rate 12.5 Hz; ''>'' means that no damage to the sample was observed during continuous operation for 6 h.

sis–absorption–heating–thermolysis, etc. An increase in the concentration of these products leads to local heating of the polymer and its evaporation in the region of the interaction with the light beam. Apparently in the case of bromonaphthalene as the CND, bromine is photodissociated from the main naphthalene molecule and this then opens up the possibility of direct absorption of visible light by molecular bromine. The process proceeds with steady autoacceleration and thus is cumulative. Soot formation and the evaporation of the polymer, as a final stage of damage, take place owing to thermolysis during the rapid local heating of the polymer in the irradiated regions.

In order to enhance the radiation resistance of holographic optical components based on these photopolymers, we have studied replacing the CND contained in the basic composite FPK-488. A list of the diffusates that were studied and some characteristics of the new FPK-488N are shown in Table II.

The preliminary choice of the compounds for use as the CND was based on the following considerations. Aromatic 1–3 and aliphatic 4–8 compounds were investigated. Since the aromatic compounds undergo thermolysis more easily than the aliphatic,[17,18] using them as a diffusate in a photopolymer recording medium lowers the radiation resistance of holographic optical components. In order to ensure highly efficient recording, compounds were chosen for which the absolute value of the difference between the refractive indices of the polymer-forming part (n_p) and the CND (n_d) was at least 0.1.

In addition, materials with a different solubility relative to the polymer were examined, and this made it possible to establish a relation between the holographic parameters and the thermodynamic properties of the photosensitive polymer composite. The Huggens parameters, which characterize the thermodynamic affinity of a CND to a polymer, were not determined. However, the change in the affinity in a series of additives can be estimated qualitatively from the value of the optimum concentration c_{opt} of the CND which ensures maximum recording efficiency. A reduction in this efficiency is evidence of a drop in the equilibrium concentration of the neutral component in the polymer and, accordingly, of a reduction in the thermodynamic affinity in the system. As Table II implies, the groups of compounds Nos. 1–4, 5, 6,

and 7, 8 have similar affinities to the polymer, and the affinity decreases in the series (5,6)–(1,4)–(7,8).

Diffusates *2* and *3* ensure highly efficient recording at roughly the same optimum concentrations as for the bromonaphthalene in the basic FPK-488. The multi-pulse damage thresholds for them lie between 12 and 16 MW/cm^2 and are roughly the same as for the basic FPK-488. In both cases, damage to the grating was accompanied by the formation of absorbing products, in agreement with earlier observations.[17] On the other hand, the CNDs 4–8 greatly increased the radiation resistance of the gratings, although the photosensitive polymer composites based on them differed in their recording efficiencies. Thus, the maximum diffraction efficiency for additives 5 and 6 was less than 30–40%. CND 4 yielded a diffraction efficiency of up to 96%, but noticeably reduced the photosensitivity of the PPC. This was a result of a chemical interaction between this diffusate and the initiator.

The best results for the magnitude of the diffraction efficiency and the radiation resistance, simultaneously, were obtained using the normal paraffins 7 and 8 as CNDs. Besides increasing the radiation resistance, going to the normal paraffins made it possible to increase the photosensitivity of the material by a factor of 1.5–2 as a result of increasing the rate of polymerization of the composite while the concentration of the CND was lowered. This last result suggests that for materials such as FPK-488, reducing the thermodynamic affinity of the neutral component leads to an improvement in such parameters of the materials as its photosensitivity.

CONCLUSION

Based on the results obtained here, we reach the following conclusions:

1. The temperature variation in the diffraction efficiency of gratings based on photopolymer materials such as FPK-488 is characterized by the existence of two temperature ranges. Within the range $+100 > T > -10\ °C$, the grating efficiency is essentially constant. The decrease in the diffraction efficiency for $T < -10\ °C$ and for $T > 100\ °C$ up until the grating is destroyed is reversible. The change in the diffraction efficiency with temperature in the latter case is de-

termined by temperature-induced changes in the phase equilibrium conditions for the polymer–diffusate system.

2. The radiation resistance of gratings using FPK-488 polymers is determined by the properties of the chemically neutral diffusate. By choosing an optimum diffusate from the class of aliphatic compounds, we were able to raise the laser damage threshold of the gratings to $\gtrsim 200$ MW/cm^2, which makes it possible to use gratings based on FPK-488N for controlling the output from high-power lasers.

3. By choosing constituents of the photosensitive polymer composite with an optimum combination of physical and thermodynamic parameters, it is possible to minimize the concentration of the diffusate in the medium without lowering the recording efficiency and, thereby, to raise the photosensitivity of these materials.

4. The recording of holographic gratings can serve as a method for determining the relative change in the affinity of the solvents to the polymer, based on the change in the optimum concentration of solvent which ensures a maximum recording efficiency for a given pair.

[1] H. J. Caulfield, *Handbook of Optical Holography*, New York (1979), 375 pp.

[2] W. J. Tomlinson and E. A. Chandross, Adv. Photochem. **12**, 202 (1980).

[3] V. A. Barachevskiĭ, Zh. Nauch. Prikl. Fotogr. Kinematogr. **38**, 75 (1993).

[4] T. N. Smirnova, E. A. Tikhonov, and E. S. Gulnazarov, Proc. Soc. Photo-Opt. Instrum. Eng. (SPIE) **1017**, 190 (1988).

[5] E. A. Tikhonov, T. N. Smirnova, and É. S. Gyul'nazarov, Kvant. Élektron. (Moscow) **40**, 1 (1991) [*sic*].

[6] A. M. Weber, W. K. Smouthers, T. J. Truot, and D. J. Mickish, Proc. Soc. Photo-Opt. Instrum. Eng. (SPIE) **1212**, 30–39 (1990).

[7] T. N. Smirnova, E. A. Tikhonov, and É. S. Gyul'nazarov, *Abstacts of the VI Conference on Holography* [in Russian], Vitebsk (1990), p. 20.

[8] O. V. Sakhno, T. N. Smirnova, and E. A. Tikhonova, Zh. Tekh. Fiz. **63**(12), 70 (1993) [Tech. Phys. **38**, 1071 (1993)].

[9] T. N. Smirnova, T. A. Sarbaev, and E. A. Tikhonov, Kvant. Elektron. (Moscow) **21**, 373 (1994).

[10] É. S. Gyul'nazarov, T. N. Smirnova, and E. A. Tikhonova, Opt. Spektrosk. **67**, 175 (1989) [Opt. Spectrosc. (USSR) **67**, 99 (1989)].

[11] É. S. Gyul'nazarov, V. V. Obukhovskiĭ, and T. N. Smirnova, Opt. Spektrosk. **69**, 178 (1990) [*sic*].

[12] R. Kohler, K. Berhart, and L. Lin, *Optical Holography*, Moscow (1973), 686 pp.

[13] A. E. Nesterov, *Handbook of the Physical Chemistry of Polymers. Properties of Solutions and Mixtures of Polymers* [in Russian], Kiev (1984), 374 pp.

[14] B. V. Ioffe, *Refractometric Techniques in Chemistry* [in Russian], Leningrad (1974), 400 pp.

[15] A. E. Nesterov and Yu. S. Lipatov, *Phase States of Solutions and Mixtures of Polymers* [in Russian], Kiev (1987), 168 pp.

[16] J. A. Jenney, J. Opt. Soc. Am. **6**, 1155 (1970).

[17] A. V. Butenin and B. Ya. Kogan, Kvant. Élektron. (Moscow) **3**, 1136 (1976) [Sov. J. Quantum Electron. **6**, 611 (1976)].

[18] R. Z. Magaril, *Mechanisms and Kinetics of Homogeneous Thermal Conversion of Hydrocarbons* [in Russian], Moscow (1970), 175 pp.

Translated by D. H. McNeill

Experimental study of the focusing of submicrosecond pressure pulses in liquids

Yu. V. Sud'enkov

St. Petersburg State University, 198904 St. Petersburg, Russia

É. V. Ivanov

*St. Petersburg State Institute of Precision Mechanics and Optics (Technical University,
197101 St. Petersburg, Russia*
(Submitted March 28, 1997)
Zh. Tekh. Fiz. **68**, 111–117 (June 1998)

Experimental data are presented from a study of the focusing of single, submicrosecond pressure
pulses in water. The effects of the initial amplitude distribution, the initial pressure level,
and the geometric parameters of the opto-acoustic concentrator are studied. It is found that the
focusing efficiency can be substantially enhanced by going from a bell-shaped distribution
of the initial amplitude to an annular distribution. © *1998 American Institute of Physics.*
[S1063-7842(98)02006-6]

INTRODUCTION

The focusing acoustic pulses in liquids currently has a wide range of applications in different areas of science, technology, and medicine.[1–3] Thus the problem of concentrating pulsed pressures in the smallest possible spatial region is extremely timely.

The Kirchhoff–Helmholtz integral equation[4] is used to describe the focusing of monochromatic waves in the linear approximation. If the local distance F is much greater than the wavelength λ, while the geometric angle of convergence, α, is not too large ($\alpha \lesssim 1$ rad), then this approach yields the following result:[5]

$$(p_f/p_0) \sim F\beta^2/\lambda, \quad L_f \sim \lambda/\beta^2, \quad D_f \sim \lambda/\beta, \qquad (1)$$

where p_f and p_0 are, respectively, the maximum pressure amplitudes in the focal plane and at the surface of the focusing source, L_f and D_f are the length and diameter of the focal region at a level of $0.5p_f$, and β is the effective convergence angle.

The parameter β is defined by the formulas[5]

$$\beta^2 = \pi^{-1} \int_0^{2\pi} d\varphi \int_0^{\alpha} f_s(\theta,\varphi) \sin\theta \, d\theta,$$

$$f_s(\theta,\varphi) = p_s(\theta,\varphi)/p_0, \qquad (2)$$

where $p_s(\theta,\varphi)$ and $f_s(\theta,\varphi)$ are, respectively, the pressure amplitude at the surface of the focusing source and its distribution function, while θ and φ are the latitude and longitude angles in a spherical coordinate system. (The origin of the coordinate system lies at the geometric focus, with the angle θ reckoned from the acoustic axis and φ in the focal plane.)

Thus in order to enhance the focusing efficiency for linear monochromatic waves, it is necessary to reduce the wavelength λ and increase the effective convergence angle β.

The criterion for applicability of the linear approach, which takes only diffraction into account, is the following:[4]

$$(4L_D/L_N) \ll 1, \quad L_D = \pi(F\beta)^2/\lambda,$$

$$L_N = \rho_0 c_0^2 \lambda/(2\pi\varepsilon p_0), \qquad (3)$$

where L_D and L_N are, respectively, the diffraction length and the distance for formation of a discontinuity in a plane wave, $\varepsilon = (1+\gamma)/2$ is the acoustic nonlinearity parameter, and ρ_0 and c_0 are the density of the liquid and the speed of sound in it.

Therefore, for sufficiently short waves and large effective angles of convergence, nonlinear effects, whose combined effect on the focusing efficiency for monochromatic waves is complicated, must be taken into account.[6–8]

It has been found that the processes involved in the propagation of pulsed and monochromatic waves in both the linear and nonlinear regimes have many features in common.[4,9,10] In particular, to examine the focusing of an initially monopolar pressure pulse of duration τ, one can use Eqs. (1)–(3) with the wavelength λ replaced by the "pulse length" $c_0\tau$.[11,12] Thus, according to the linear theory, the efficiency with which a single pulse is focused will increase as its duration is reduced. Then, however, the influence of nonlinear effects should become greater, and this may be qualitatively different, depending on the ratio L_D/L_N.

Hence, studies of the focusing of short, single pressure pulses ($\tau \lesssim 10^{-7}$ s), which are most conveniently generated using an opto-acoustic approach,[13] are of great interest from the standpoint of both fundamental and applied problems.

In both theoretical and experimental studies of the focusing of pressure pulses it is customary to consider bell-shaped profiles of their initial amplitude. An analysis of the available experimental data[14–18] indicates that for enhancing the focusing efficiency for submicrosecond pulses it is necessary to increase the geometric angle of convergence. This conclusion is consistent with Eqs. (1) and (2): as α is increased, so does β, and therefore the focusing conditions are improved.

There is, however, another approach to the problem of increasing the effective convergence angle. According to
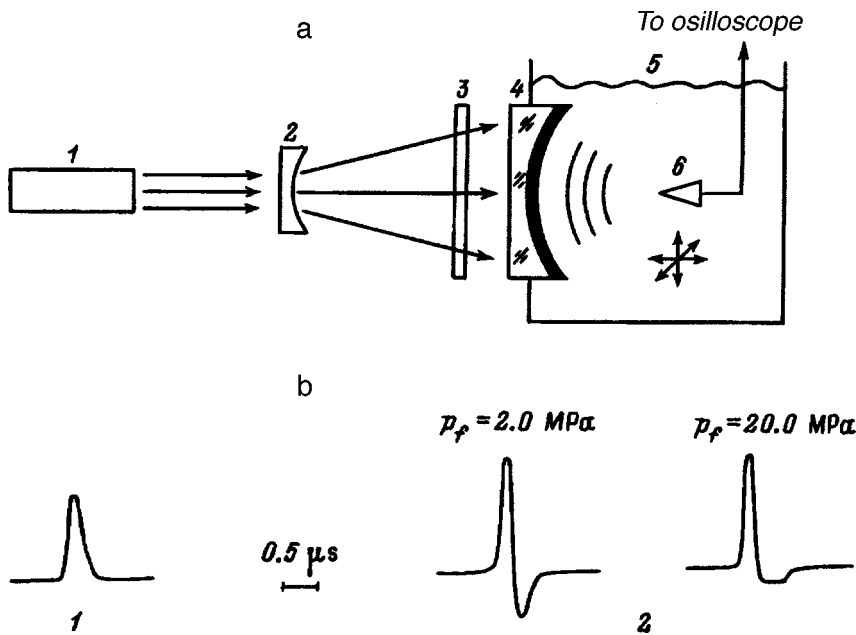
FIG. 1. The experimental apparatus (a) and temporal profiles of the pressure pulses (b): *1*—near the surface of the absorbing layer, *2*—at the focus of the opto-acoustic concentrators.

Eq. (2), the largest contribution to β is from the periphery of the beam, rather than from its axial region. Thus, the effective angle β can be increased without changing α, by going from a bell-shaped distribution of the initial amplitude to an annular distribution.

Under nonlinear conditions a bell-shaped initial distribution is clearly even more inferior from the standpoint of focusing because of the negative effect of nonlinear refraction. In the case of a bell-shaped distribution of the initial amplitude, nonlinear refraction straightens the wave front and thereby makes focusing more difficult. At the same time, the nonlinear distortion of the pulse profile, which enriches the initial spectrum with higher-frequency harmonics, promotes focusing until a shock front develops. Then the deleterious effect of nonlinear refraction is augmented by nonlinear absorption. Thus, different situations can occur, depending on the intensity of these nonlinear processes.

It appears that using beams with an annular distribution of the initial pressure can greatly attenuate the deleterious influence of nonlinear effects. In this case, in the initial stage of propagation of the wave, nonlinear refraction will not inhibit the focusing process but, on the contrary, will facilitate it. In addition, for a fixed acoustic energy, p_0 and, therefore, the influence of nonlinear effects will be smaller for larger values of β.

In order to verify these assumptions, we have made an experimental study of the dependence of the focusing of submicrosecond pressure pulses in water on the level of p_0 and on the geometric parameters of the opto-acoustic concentrator for the cases of bell-shaped and annular initial distributions.

EXPERIMENTAL TECHNIQUE

A multimode Q-switched neodymium glass laser (wavelength 1.06 μm) was used to excite the pressure pulses. The duration of the laser pulse at half maximum was 20 ns and its energy was varied over the interval 0.1–0.3 J.

Figure 1a shows a sketch of the experimental apparatus. The laser beam *1* was expanded by a negative lens *2*. In order to smooth out its transverse multimode structure, a diffuser *3* was placed at the inlet of the opto-acoustic concentrator *4*.

The laser pulse was converted into a single pressure pulse of duration 0.2 μs at half maximum in a solid absorbing layer of thickness $\cong 0.4$ mm deposited on the concave spherical surface of the glass substrate. The layer material was well matched with water in terms of its acoustic impedance ($\rho \approx 10^3$ kd/m^3, $c \approx 1.8 \times 10^3$ m/s) and this essentially eliminated energy losses owing to reflection as the pressure pulse passed from the opto-acoustic layer into the water tank *5*.

The thermoelastic mechanism for opto-acoustic generation ensured that the absorbing layer could be used many times and provided good reproducibility of the parameters of the exciting pressure pulses with a high light-to-sound conversion efficiency. The parameter $\eta = p_0/J_0$ (where J_0 is the maximum energy density of the laser pulse at the aperture of the concentrator) was ~ 20 MPa/(J/cm^2) for $0 < J_0 < 0.5$ J/cm^2.

The change in the distribution of the initial amplitude of the pressure pulse was measured by transforming the intensity distribution of the laser light at the inlet of the opto-acoustic concentrator.

The pressure in the water was recorded using a piezoceramic probe *6* with a sensitive area of diameter 0.5 mm. The signal from the probe was fed to the input of an S8-14 recording oscilloscope (bandwidth ~ 50 MHz). A differential interferometer with stabilized sensitivity and photoelectron pulse counting, operating in a linear mode, was used to calibrate the probe.

Studies of focusing were done on two opto-acoustic concentrators: 1) $F = 68$ mm and $D = 54$ mm, and 2) $F = 91$ mm and $D = 110$ mm, where D is the diameter of the concentrator.

Figure 1b shows typical profiles of the pressure pulses recorded near the surface of the absorbing layer and at the focus of the opto-acoustic concentrators for different levels of the initial amplitude. At low pressures, one can observe a characteristic diffractional transformation of the temporal profile, i.e., the pulse is differentiated. As the initial amplitude is increased, the influence of nonlinearity shows up distinctly:[4] the leading edge of the pulse becomes steeper, the duration of the compression phase is shorter, and its relative amplitude increases, while the rarefaction phase is extended and smoothed out.

It should be noted that in our experiments, as opposed those of Musatov and Sapozhnikov,[16] no shock front formation was observed, so only two nonlinear effects operated: nonlinear distortion of the temporal profile, and nonlinear refraction.

RESULTS AND DISCUSSION

The axially symmetric functions $f_s(\theta,\varphi)=f_s(\theta)$ constructed on the basis of the experimental data for the case of a bell-shaped distribution of the initial amplitude are shown in Fig. 2a. The changes in the amplitude of the compression phase of the pressure pulse along the acoustic axis and in the focal plane are shown in Figs. 2b and 2c for two laser pulse energies E. The absolute magnitudes of the initial and focal pressures, the gain coefficient $G=p_1/p_0$, the length and diameter of the focal region, and the effective convergence angle are listed in Table I.

These data show that as the effective convergence angle and the focal length are increased (i.e., on going from the first concentrator to the second), the focal pressure and the gain increase substantially. At the same time, there is a noticeable reduction in the size of the focal region, both the dimension along the acoustic axis (2) and the diameter.

For low initial pressures ($p_0 \sim 0.2-0.3$ MPa), these data are in fairly good agreement with the predictions of the linear theory. According to Eq. (1), going from the first to the second concentrator should yield an increase in G by a factor of 2.4 as L_f and D_f are reduced by factors of 1.8 and 1.3, respectively. In the experiment the gain increased by a factor of 2.4 and the size of the focal region was reduced to roughly half.

At higher initial pressures ($p_0 \sim 2-3$ MPa), the agreement between the linear theory and the experimental data is poorer owing to the greater role of nonlinear phenomena. Here the changes indicating an enhancement in the focusing efficiency on going from the first concentrator to the second are less marked than at low p_0: the gain increases by a factor of 1.7, while the size of the focal region is reduced by a factor of 1.3.

The influence of nonlinear effects on the focusing process is clearly demonstrated by the plots of $p_f(p_0)$ and $G(p_0)$ for the second concentrator shown in Fig. 3. For $p_0 \leqslant 1$ MPa, these curves are still linear, while for higher initial pressures the focusing efficiency falls off. That the range of linearity was much wider than predicted by Eq. (3) is apparently explained by the fact that nonlinear refraction and nonlinear distortion of the temporal profile counterbalance each
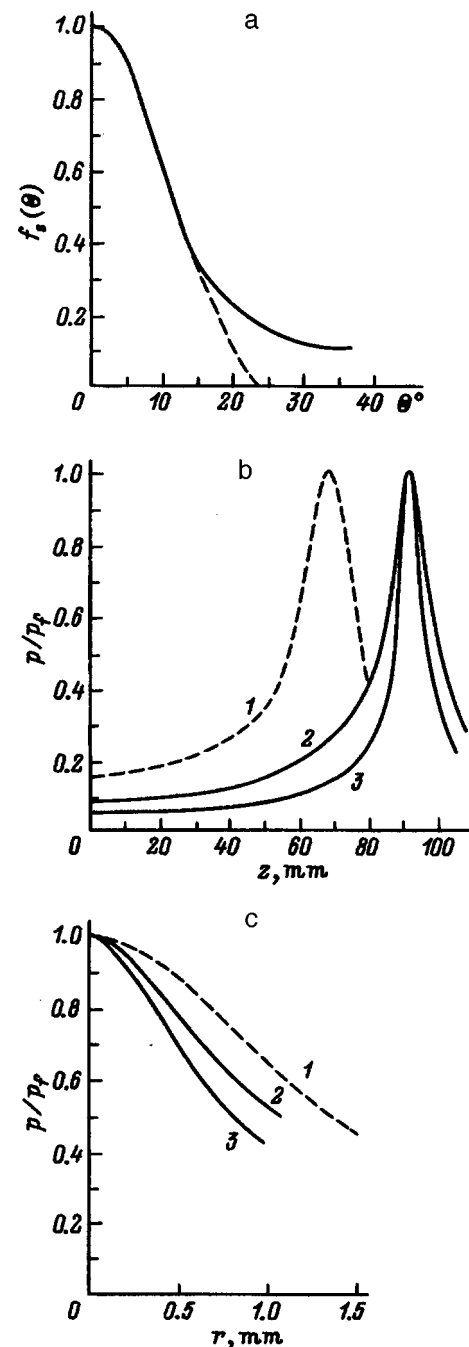

FIG. 2. Bell-shaped distribution of the initial pressure in the opto-acoustic concentrator. The dashed curve is the first concentrator and the smooth curve, the second. E (J): 0.3 and 3.0 (*1*), 1.85 (*2*), 0.15 (*3*).

other at low initial pressures. As p_0 is increased, nonlinear refraction begins to predominate and the focusing deteriorates.

Let us examine the experimental data for the case of an annular distribution of the initial amplitude. The functions $f_s(\theta,\varphi)=f_s(\theta)$ for this sort of distribution are shown in Fig. 4a. Figure 4b shows the variations in the amplitude of the compression phase of the pressure pulse along the acoustic axis for the first and second concentrators for two laser pulse energies. The corresponding radial distributions in the focal planes of the opto-acoustic concentrators are given in Fig. 4c. The normalization parameters, gain coefficient, length

TABLE I.

Opto-acoustic concentrator	E, J	p_0, MPa	p_f, MPa	G	L_f, mm	D_f, mm	β^0
1	0.30	0.30	2.0	6.7	22	2.8	13.5
1	3.00	3.00	20.0	6.7	22	2.8	13.5
2	0.15	0.17	2.7	15.9	10	1.6	18.0
2	1.85	2.00	22.9	11.5	18	2.2	18.0

and diameter of the focal region, and the effective convergence angle are listed in Table II.

The data for an annular distribution of the initial amplitude also show that as the effective divergence angle and focal length are increased (on going from the first to the second concentrator), there is a significant rise in the gain and focal pressure with a simultaneous decrease in the size of the focal region. As opposed to the case of a bell-shaped initial distribution, these changes are larger than predicted by the linear theory. In the experiment, even for $p_0 \sim 1-2$ MPa, an increase in the focusing efficiency by a factor of 2.7 was observed with respect to all three parameters (G, L_f, and D_f), while for $p_0 \sim 0.1-0.2$ MPa these indicators were still higher.

Here it should be noted that the measured value of the diameter of the focal region for the second concentrator was comparable to the size of the sensitive area of the pressure probe. Thus, the dimensions of the focal volume were overstated, while the pressures measured in the neighborhood of the focus were understated.

The effect of the finite probe size on the measurement results can be estimated by writing the amplitude distribution of the compression phase of the pressure pulse in the focal plane in the form

$$p(r) = p_f^* F(r/D_f^*), \qquad (4)$$

where $F(r/D_f^*)$ the dimensionless profile function, p_f^* is the true value of the pressure at the focus, and D_f^* is the actual diameter of the focal region at the $0.5p_f^*$ level.

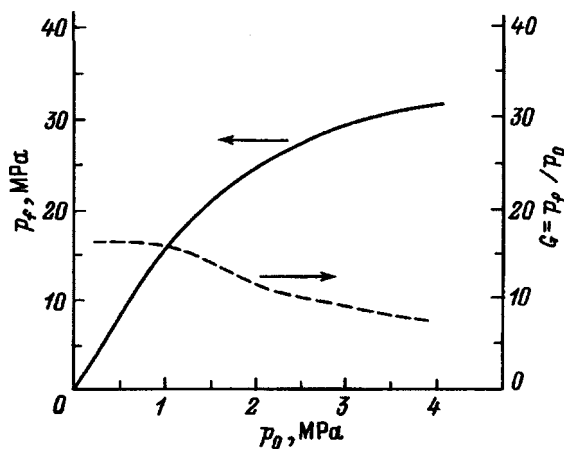FIG. 3. The focal pressure and gain coefficient as functions of the initial pressure for the second opto-acoustic concentrator in the case of a bell-shaped distribution of the initial amplitude.
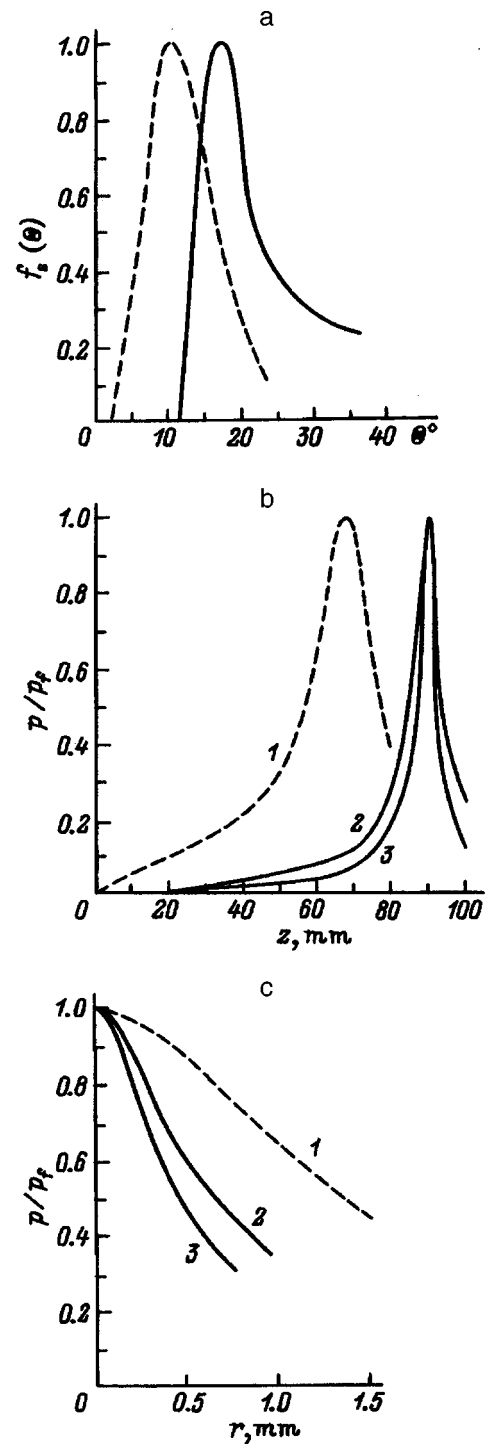
FIG. 4. Annular distribution of the initial pressure in the opto-acoustic concentrator. The dashed curve is the first concentrator and the smooth curve, the second. E (J): 0.3 and 3.0 (*1*), 1.85 (*2*), 0.15 (*3*).

TABLE II.

Opto-acoustic concentrator	E, J	p_0, MPa	p_f, MPa	G	L_f, mm	D_f, mm	β^0	M_{LG}
1	0.30	0.18	2.2	12.2	22	2.8	17.0	
1	3.00	1.80	22.0	12.2	22	2.8	17.0	
2	0.15	0.08	2.7	33.8	5	1.0	22.5	1.42
2	1.85	1.00	29.2	29.2	9	1.3	22.5	1.16

Introducing the dimensionless averaging coefficient $M = p_f^*/p_f$, we find

$$M = d^2 \left[8 \int_0^{d/2} F(r/D_f^*) r\, dr \right]^{-1}, \qquad (5)$$

where d is the probe diameter.

Since $D_f > d$, for the purpose of estimates it is sufficient to approximate $F(r/D_f^*)$ by some bell-shaped function, assuming that $D_f^* \approx D_f - d$. Let us compare the values of M calculated for Lorentzian (L) and Gaussian (G) focal distributions

$$M_L = x^2 \ln^{-1}(1 + x^2),$$

$$M_G = x^2 \ln 2 [1 - \exp(-x^2 \ln 2)]^{-1}, \qquad (6)$$

where $x = d/D_f^*$.

M_L and M_G begin to differ noticeably only for $x > 1.5$ and, therefore, the parameter $M_{LG} = (M_L + M_G)/2$ introduced in Table II can be used for our estimates. For $D_1 \geq 1.6$ mm, the calculation yields values of M_{LG} which differ little from unity ($M_{LG} \leq 1.09$), so they have not been indicated in Table II and were not taken into account in analyzing the experimental data.

Thus, in the case of the second type of opto-acoustic concentrator, the growth in the focusing efficiency was underestimated by roughly factors of 1.5 and 1.2, respectively, for low and high initial pressures. In order to determine the focal parameters more accurately, the spatial resolution of the measurements must be increased severalfold; that is an independent and by no means simple problem.

Let us analyze the enhancement in the focusing efficiency on going from a bell-shaped distribution of the initial amplitude to an annular one. For the first concentrator, the gain increases by a factor of 1.8, while the changes in the dimensions of the focal region are negligible. At the same time, for the second concentrator, the gain increases by a factor of 3.0, while the length and diameter of the focal region are reduced to half.

Going from a bell-shaped distribution of the initial amplitude to an annular distribution, therefore, makes it possible to greatly enhance the focusing efficiency, and these changes are most noticeable for the second concentrator. In the meantime, according to Eq. (1), the increase in the focusing efficiency should be the same for the first and second concentrators, since on replacing a bell-shaped initial distribution by an annular one, β increases by the same factor for both of the concentrators (Tables I and II).

The observed additional improvement in the focusing conditions for the second opto-acoustic concentrator appears to be related to a lessening of the negative contribution of nonlinear phenomena in the case of an annular distribution of the initial pressures. An analysis of the geometry of the annular distributions of the initial amplitude (Fig. 4a) shows that for the second concentrator, the ring of the initial distribution stands substantially farther away from the center than for the first concentrator, a circumstance which also showed up in the character of the nonlinear refraction. This assumption is confirmed by the plots of $p/p_f(z)$ in Fig. 4b. For the first concentrator the pressure equals zero only in the neighborhood of the point $z = r = 0$, while for the second concentrator the region in which there is no acoustic perturbation on the beam axis is much wider ($z \leq 20$ mm).

In order to determine the influence of the geometry of an annular initial distribution on the focusing efficiency, we performed the following experiment. At $p_0 \sim 0.1$ MPa, an annular distribution of the initial amplitude for the second concentrator was shifted from the center to the edge in a way such that the effective width of the annulus remained constant. The diameter of the focal region was found to decrease to 0.5 mm, i.e., to the size of the sensitive area of the pressure probe, with the measured focal pressure p_f essentially constant owing to the strong spatial averaging. These results confirm the important influence of the geometric parameters of an annular distribution on the focusing efficiency and the possibility of further increasing the efficiency by optimizing $f_s(\theta, \varphi)$.

The suppression of the deleterious contribution from nonlinear refraction compared to the positive influence of the nonlinear distortion of the temporal profile shows up most clearly in the behavior of $p_f(p_0)$ and $G(p_0)$ for the case of an annular distribution of the initial amplitude (Fig. 5). At initial pressures below 1 MPa, $G(p_0)$ has a characteristic maximum. This confirms the earlier conclusion[6,8,10,11,14] that at low initial pressures, when a shock front is unable to develop, nonlinear effects can improve the focusing conditions compared to the linear case, if the nonlinear distortion of the temporal profile predominates over nonlinear refraction. At higher p_0, as in the case of a bell-shaped initial distribution, the focusing efficiency falls off owing to the dominant contribution of nonlinear refraction.

CONCLUSIONS

These studies have shown that the distribution of the initial amplitude of submicrosecond pressure pulses in a liquid has a significant effect on the process by which they are focused. In particular, we have demonstrated the possibility of greatly increasing the focusing efficiency by going from a
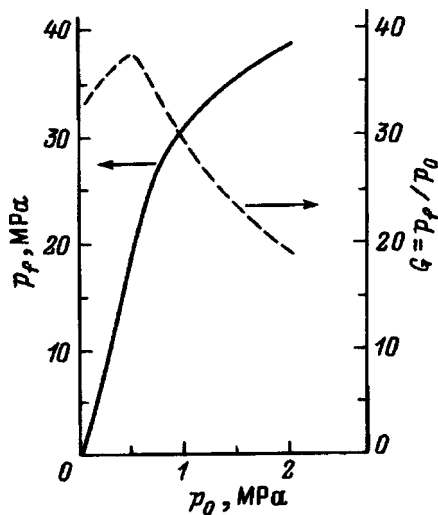
FIG. 5. The focal pressure and gain coefficient as functions of the initial pressure for the second opto-acoustic concentrator in the case of an annular distribution of the initial amplitude.

bell-shaped initial distribution to an annular one. This phenomenon is observed over a wide range of initial pressures.

At low levels of p_0, the positive effect originates primarily in an enhancement in the effective convergence angle β when the distribution function is changed appropriately. At higher initial pressures there is an additional factor: the suppression of the negative influence of nonlinear refraction compared to the nonlinear distortion of the temporal profile of the pulse. Then the degree of suppression depends significantly on the geometric parameters of the annular distribution of the initial amplitude.

[1] L. D. Rozenberg (Ed.), *High-Power Ultrasonic Waves* [in Russian], Nauka, Moscow (1968).

[2] H. Reichenberger, Proc. IEEE **76**, 1236 (1988).

[3] C. R. Hill (Ed.), *Physical Principles of Medical Ultrasonics* [Halstead Press, New York, 1986; Mir, Moscow, 1989].

[4] M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *The Theory of Waves* [in Russian], Nauka, Moscow (1990), 432 pp.

[5] I. N. Kanevskiĭ, *The Focusing of Acoustic and Ultrasonic Waves* [in Russian], Nauka, Moscow (1977), 336 pp.

[6] L. A. Ostrovskiĭ and A. M. Sutin, Dokl. Akad. Nauk SSSR **221**, 1300 (1975).

[7] N. S. Bakhvalov, Ya. M. Zhileĭkin, and E. A. Zabolotskaya, *Nonlinear Theory of Sound Beams* [in Russian], Nauka, Moscow (1982).

[8] S. N. Makarov, ''Finite amplitude near-field modelling of ultrasonic fields using a transfer matrix formulation,'' PTB-Bericht MA-42, Physikalische-Technische Bundesanstalt, Braunschweig; Berlin (1995), 57 pp.

[9] O. V. Vasil'eva, A. A. Karabutov *et al.*, *Interaction of One-Dimensional Waves in Dispersionless Media* [in Russian], Izd. MGU, Moscow (1983), 152 pp.

[10] Al. M. Kolomenskiĭ, A. A. Maznev, and V. G. Mikhalevich, Izv. Akad. Nauk SSSR, Ser. Fiz. **54**, 2451 (1990).

[11] O. A. Sapozhnikov, Akust. Zh. **37**, 760 (1991).

[12] A. G. Musatov, O. V. Rudenko, and O. A. Sapozhnikov, Akust. Zh. **38**, 502 (1992).

[13] V. É. Gusev and A. A. Karabutov, *Laser Optoacoustics* [in Russian], Nauka, Moscow (1991), 304 pp.

[14] A. I. Bozhkov, F. V. Bunkin *et al.*, Izv. Akad. Nauk SSSR, Ser. Fiz. **46**, 1624 (1982).

[15] I. I. Komissarova, G. V. Ostrovskaya *et al.*, Zh. Tekh. Fiz. **62**(2), 34 (1992) [Tech. Phys. **37**, 130 (1992)].

[16] A. G. Musatov and O. A. Sapozhnikov, Akust. Zh. **39**, 315 (1993).

[17] A. G. Musatov and O. A. Sapozhnikov, Akust. Zh. **39**, 510 (1993).

[18] A. G. Musatov, Akust. Zh. **41**, 117 (1995).

Translated by D. H. McNeill

Magnetostatic surface waves produced by an inhomogeneity of the anisotropy with a turning point of the spectral function on a ferromagnet surface

I. A. Kaĭbichev and V. G. Shavrov

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 103907 Moscow, Russia

Magnetostatic surface waves with fixed frequency and wave vector are predicted to exist in a ferromagnet with an inhomogeneity of the magnetic anisotropy such that the spectral function has a turning point on the surface. This result is most important for the case when an external magnetic field magnetizes the ferromagnet perpendicular to its surface. The frequency of the surface wave is determined by the frequency of the magnetostatic volume wave at the surface of the ferromagnet, and the wave vector is determined by the surface values of the local magnetic anisotropy field and its derivative. © *1998 American Institute of Physics.* [S1063-7842(98)02106-0]

INTRODUCTION

Ferrite epitaxial films, mainly of yttrium iron garnet, are widely used in microwave electronics. They are grown on nonmagnetic substrates, such as gallium gadolinium garnet. The existing film growth technologies mainly yield films that are inhomogeneous over the sample thickness.[1,2] Inhomogeneities can also be created artificially by ion implantation.[3–8] In this regard, a need has arisen for studying the properties of magnetostatic waves in films that are inhomogeneous over their thickness. The distribution of the magnetization in the ground state of a ferromagnet is determined to a great extent by the profile of the magnetic anisotropy constant; here the inhomogeneous exchange constant and the saturation magnetization can be regarded as constant.[6] We will therefore consider a ferromagnet with only an inhomogeneity in the magnetic anisotropy. We limit ourselves to inhomogeneities with a single turning point of the spectral function on the surface of the ferromagnet. A magnetostatic volume wave is thereby allowed to propagate on the surface of the ferromagnet. We shall choose the profile of the inhomogeneity in the magnetic anisotropy constant to be such that magnetostatic volume waves cannot propagate inside the ferromagnet. In that way, we obtain exponential damping of the magnetic potential of the wave in the interior of the ferromagnet, and this leads to localization of the wave near the surface. A wave of this sort can be classed as a magnetostatic surface wave. This result is of greatest interest for a ferromagnet magnetized by an external magnetic field perpendicular to its surface, since then only magnetostatic volume waves will exist in the homogeneous case,[9–11] while surface excitations of the spin system show up only when exchange is included.[12–16] Thus, in a normally magnetized ferromagnet, magnetostatic surface waves can exist as a result of the inhomogeneity in the magnetic anisotropy.

EXPERIMENTAL AND THEORETICAL STUDIES OF MAGNETOSTATIC SURFACE WAVES IN INHOMOGENEOUS FILMS

Most studies have been made of the case where the external magnetic field and the magnetization are in the same direction and lie in the plane of the film, while the wave, itself, propagates perpendicular to them.[17–24] The inhomogeneity in the distribution of the magnetization is modeled by a multilayer film consisting of two or three homogeneous layers with different magnetic parameters.[19–23] Numerical calculations have been done for a specified inhomogeneity profile by several authors.[17,18,24] It is important to obtain analytical results for a rather general form of the inhomogeneities.

The case in which an external magnetic field magnetizes an inhomogeneous ferromagnet in a direction perpendicular to its surface is of greatest interest. An inhomogeneity in a film of yttrium iron garnet with this geometry has only been considered once,[25] in an examination of the direct and inverse spectral problems for magnetostatic forward volume waves, while the experimental spectra were interpreted in terms of a spatial inhomogeneity in the uniaxial anisotropy.

STATEMENT OF THE PROBLEM AND BASIC EQUATIONS

Let us consider a uniaxial semi-infinite ferromagnet which occupies the region $z>0$ (see Fig. 1) and has an arbitrary profile of the inhomogeneity of the local magnetic anisotropy field $H_A(z)=2K(z)/M_0$, where $K(z)$ is the magnetic anisotropy constant and M_0 is the saturation magnetization. We assume that the magnetic anisotropy field $H_A(z)$ is less than the absolute value of the demagnetizing field $|H_M|=4\pi M_0$. Then in weak magnetic fields $H_0<4\pi M_0-H_A(z)$ the ground state is inhomogeneous:

$$\psi_0(z)=\pm\arccos(H_0/[4\pi M_0-H_a(z)]), \tag{1a}$$

while in strong fields $H_0>4\pi M_0-H_A(z)$ a homogeneous ground state is realized,
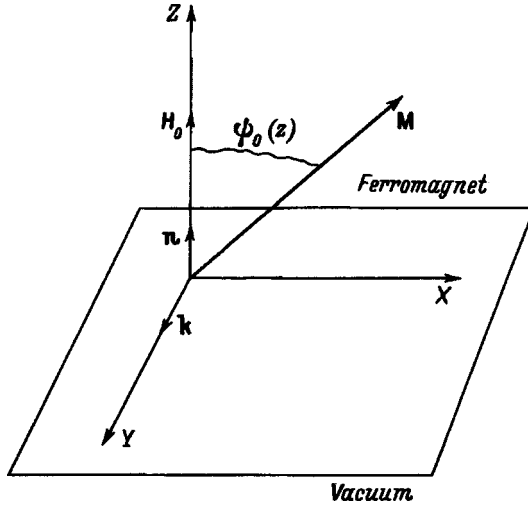
FIG. 1. The geometry of the problem. $\mathbf{H}_0$ is the external magnetic field; $\mathbf{n}$ is a unit vector characterizing the direction of the easy axis z of the ferromagnet, which is perpendicular to its surface; $\mathbf{M}$ is the magnetization (in general it deviates from the easy axis by an angle $\psi_0(z)$); and, $\mathbf{k}$ is the wave vector of the magnetostatic surface wave.

$$\psi_0(z)=0. \tag{1b}$$

This latter condition also occurs in the case of $H_A(z) > 4\pi M_0$ and $H_0 > 0$. In a situation with $H_A(z) > 4\pi M_0$ and $4\pi M_0 - H_A(z) < H_0 < 0$, the homogeneous phase (1b) is metastable. Thus, the results for the magnetostatic surface wave spectrum in the latter case will be valid only when the energy of the magnetic excitations of the ferromagnet is small compared to the potential barrier preventing a transition of the ferromagnet into the homogeneous stable state $\psi_0(z)=\pi$. The exchange interaction was not taken into account in determining the ground state. This is valid if the dimensions of the ferromagnet film and the characteristic scale length of the inhomogeneity of the magnetic anisotropy field (the length over which $H_A(z)$ changes from its surface to its bulk value), L, exceed the exchange length $\delta=\sqrt{2A/(H_{A0}M_0)}$. ($A$ is the inhomogeneous exchange constant and H_{A0} is the anisotropy field in the interior of the film.)

Surface magnetostatic waves propagate along the Y axis, so we assume that all the variables in the problem are proportional to $\exp(i\omega t - iky)$, where ω is the frequency and k is the wave vector. We consider frequencies up to several GHz, since they are usually employed in practice.[19-23] At these frequencies, the wave vector $k < 10^5$ cm^{-1}. In this region, the contribution of the exchange interaction is small compared to the other terms in the magnetic energy: dipole–dipole and Zeeman. In an examination of magnetostatic waves, it can be neglected. We shall begin with the system of equations of magnetostatics:

$$\text{curl } H_M = 0, \quad \text{div}(H_M + 4\pi M) = 0, \tag{2}$$

where H_M is the demagnetization field.

The magnetization M satisfies the Landau–Lifshitz equation

$$\partial M/\partial t = -\gamma [M \cdot H_{\text{eff}}(z)], \tag{3}$$

where γ is the gyromagnetic ratio, $H_{\text{eff}}(z)=H_0+H_M+H_A(z) \times (Mn)n/M_0$ is the effective magnetic field, and n is the unit vector characterizing the direction of the anisotropy axis of the ferromagnetic crystal (directed along the easy axis z of the ferromagnet).

The magnetization vector in the ground state has components $(M_0 \sin \psi_0(z), 0, M_0 \cos \psi_0(z))$, and the demagnetization field is $(0,0,-4\pi M_0 \cos \psi_0(z))$. We assume that the deviations of the magnetization vector $\mathbf{m}$ and the demagnetizing field $\mathbf{h}$ from these equilibrium values are small. We linearize the Landau–Lifshitz equation (3) to obtain the coupling between the components of the vectors $\mathbf{m}$ and $\mathbf{h}$, which we write in the form

$$m_i = \chi_{ij} h_j, \quad i,j = x,y,z, \tag{4}$$

where χ_{ij} is the high-frequency magnetic susceptibility tensor of the ferromagnet.

Its components are given by

$$\chi_{XX} = \Gamma(z)\Omega_1(z)\cos^2 \psi_0(z),$$

$$\chi_{XY} = -\chi_{YX} = \Gamma(z)i\omega \cos \psi_0(z),$$

$$\chi_{YY} = \Gamma(z)\Omega_2(z), \quad \chi_{YZ} = -\chi_{ZY} = \Gamma(z)i\omega \sin \psi_0(z),$$

$$\chi_{ZX} = \chi_{XZ} = -\Gamma(z)\Omega_1(z)\cos \psi_0(z)\sin \psi_0(z),$$

$$\chi_{ZZ} = \Gamma(z)\Omega_1(z)\sin^2 \psi_0(z).$$

Here we have used the notation

$$\Gamma(z) = \gamma M_0 /[\Omega_1(z)\Omega_2(z) - \omega^2],$$

$$\Omega_1(z) = \gamma[H_0^{(i)} + H_A(z)\cos \psi_0(z)]\cos \psi_0(z),$$

$$\Omega_2(z) = \gamma[H_0^{(i)} \cos \psi_0(z) + H_A(z)\cos 2\psi_0(z)],$$

while $H_0^i = H_0 - 4\pi M_0 \cos \psi_0(z)$ is the internal magnetic field. Note that on going to a system of coordinates with its Z axis coincident with the magnetization of the ground state of the ferromagnet, we essentially obtain a form of the high-frequency magnetic susceptibility tensor which was known previously.[9] The differences are related to the dependence of the magnetic anisotropy field on the vertical coordinate z. The components of the magnetic susceptibility tensor also become functions of z. Substituting Eq. (4) into the equation of magnetostatics (2), and then introducing the magnetic scalar potential $\Phi(\mathbf{h}=-\nabla\Phi)$, we obtain a second-order differential equation with variable coefficients:

$$D^2\Phi(z) - k^2 Q(\omega,z)\Phi(z) = 0, \tag{5}$$

where $Q(\omega,z) = 1 + 4\pi\gamma M_0 \Omega_2(z)/[\Omega_1(z)\Omega_2(z) - \omega^2] + \omega D\{4\pi\gamma M_0 \sin \psi_0(z)/[\Omega_1(z)\Omega_2(z) - \omega^2]\}/k$ is a function which determines the character of the solutions, which we shall refer to as the spectral function, and $D^2 = \partial^2/\partial z^2$. The form of the spectral function $Q(\omega,z)$ depends on the choice of ground state:

$$Q(\omega,z) = -[\omega_{V1}(z) - \omega][\omega - \omega_{V2}(z)]/\omega^2 \tag{6a}$$

for the inhomogeneous ground state (1a) and

$$Q(\omega,z) = [\omega_{V2}^2(z) - \omega^2]/[\omega_{V1}^2(z) - \omega^2] \tag{6b}$$

for the homogeneous phase (1b). The quantities $\omega_{V1}(z)$ and $\omega_{V2}(z)$ have the same physical significance in both cases: they are the lower and upper local boundaries of the magnetostatic volume wave spectrum in a homogeneous slab with parameters equal to their values at the point z. Their forms, however, depend on the choice of ground state. For the inhomogeneous state (1a),

$$\omega_{Vj} = A(z) + (-1)^j \sqrt{A^2(z) - B(z)},$$

$$A(z) = 2\pi\gamma M_0 \cos \psi_0(z) D\psi_0(z)/k,$$

$$B(z) = 4\pi\gamma^2 M_0 H_A(z) \sin^2 \psi_0(z), \quad D = \partial/\partial z,$$

while for the homogeneous phase (1b),

$$\omega_{V1}(z) = \Omega_0(z), \quad \omega_{V2}(z) = \sqrt{\Omega_0(z)[\Omega_0(z) + 4\pi\gamma M_0]},$$

where $\Omega_0(z) = \gamma[H_0 - 4\pi M_0 + H_A(z)]$ is the local ferromagnetic resonance frequency.

In the beginning we examine a uniform material, where there is no dependence on z and $Q(\omega,z) = Q(\omega)$. In an infinite material there is a single magnetostatic volume wave, whose frequency is determined from the condition $Q(\omega) = 0$. For a slab with $Q(\omega) < 0$, the solutions of Eq. (5) are expressed in terms of a linear combination of the sine and cosine and describe an infinite set of magnetostatic volume wave spectral modes.[10,11] It should be noted that these waves, by analogy with elastic waves,[26,27] occupy an intermediate position between surface and volume waves, but the term ''magnetostatic volume waves in a slab'' has become established in the literature. In a half space with $Q(\omega) > 0$, the solution of Eq. (5) is expressed in terms of a linear combination of exponents with different signs and describes a magnetostatic surface wave (if it satisfies the boundary conditions).

Everything is a bit more complicated in an inhomogeneous medium. For example, if there is point z_0, referred to as the turning point, such that $Q(\omega,z_0) = 0$, then only one magnetostatic volume wave can propagate at the point z_0. If $Q(\omega,z) < 0$ in some layer, then an infinite set of magnetostatic volume wave modes exist in it, while for $Q(\omega,z) > 0$ a magnetostatic volume wave cannot exist, but a magnetostatic surface wave can propagate. Let us choose a situation such that $Q(\omega,z=0) = 0$, while for any other $z > 0$ the spectral function $Q(\omega,z) > 0$. The conditions for propagation of a magnetostatic volume wave are thereby created on the surface of the ferromagnet, while inside it the magnetic potential will fall off exponentially. Ultimately, this should result in localization of the wave near the surface, so we shall refer to this as a surface wave.

We now formulate the boundary conditions for the problem. They involve the continuity at the ferromagnet surface ($z = 0$) of the normal component of the magnetic induction and the tangential component of the magnetic field strength, which reduces to the following conditions for the magnetic potential:

$$\Phi(0) = \Phi_B(0), \quad -D\Phi(0) + 4\pi m_z(0) = -D\Phi_B(0), \tag{7}$$

where $\Phi_B(z)$ is the magnetic potential in the vacuum region ($z \geq 0$).

Thus, the second-order differential equation (5) with the spectral function (6) and the boundary conditions (7) can be used to describe the propagation of a surface magnetostatic wave in an inhomogeneous ferromagnet.

MAGNETOSTATIC SURFACE WAVE SPECTRA

Of all the possible profiles of the local magnetic anisotropy field $H_A(z)$, we shall consider only those which have a single turning point in the spectral function $Q(\omega,z)$, at the surface of the ferromagnet, i.e., $Q(\omega,z=0) = 0$. This condition uniquely determines the frequency of the surface magnetostatic wave:

$$\omega_S = \omega_{V2}(0). \tag{8}$$

The frequency ω cannot take other values. The upper local limit on the magnetostatic volume wave spectrum in a slab, $\omega_{V2}(0)$, is given by different expressions (6a) or (6b), depending on whether an inhomogeneous (1a) or homogeneous (1b) state is realized. In the interior of the ferromagnet for $z > 0$ we assume that the spectral function $Q(\omega_S,z)$ is positive. Then the upper local limit $\omega_{V2}(z)$ for any $z \geq 0$ must satisfy the inequality

$$\omega_{V2}(z) \leq \omega_{V2}(0). \tag{9}$$

The equal sign holds only for $z = 0$. Thus, the situation we are examining is realized only for profiles of the local magnetic anisotropy field which ensure that the upper local limit ω_{V2} of the magnetostatic volume wave spectrum in the slab satisfies inequality (9).

In the short-wavelength approximation, $|k|L \gg 1$ (wavelength much shorter than the inhomogeneity scale length L), Eq. (5) has the solution[28,29]

$$\Phi(z) = \Phi_0 \, Ai[|k|^{2/3} \xi(z)],$$

$$\xi(z) = \left[\frac{3}{2} \int_0^z dt \sqrt{Q(\omega_S,t)} \right]^{2/3}, \quad z \geq 0, \tag{10}$$

where $Ai(\alpha)$ is the Airy function of the first kind.

Near the surface, for $z \to 0$, the function $\xi(z) = \sqrt[3]{DQ(\omega_S,0)}z$, while the Airy function is written in the form

$$Ai(\alpha) = Ai(0) + [\partial Ai(\alpha)/\partial\alpha]_{\alpha=0}\alpha = -\frac{1}{3^{2/3}\Gamma(2/3)}$$

$$-\frac{z|k|^{2/3}\sqrt[3]{DQ(\omega_S,0)}}{3^{4/3}\Gamma(4/3)},$$

where $\Gamma(\varepsilon)$ is the gamma function.

Then the distribution of the magnetic potential of the surface magnetostatic wave with frequency (8) has the asymptotic form

$$\Phi(z) = -\Phi_0 \left\{ \frac{1}{3^{2/3}\Gamma(2/3)} + \frac{z|k|^{2/3}\sqrt[3]{DQ(\omega_S,0)}}{3^{4/3}\Gamma(4/3)} \right\} \tag{11a}$$

near the surface ($z \to 0$) and

$$\Phi(z) = \frac{\Phi_0}{2\pi} \sqrt[4]{\frac{Q(\omega_S)}{Q(\omega_S, z)}} \exp\left\{ -|k| \int_0^z dt \sqrt{Q(\omega_S, t)} \right\}$$
$$(11b)$$

in the interior of the ferromagnet ($z \gg 0$). Outside the ferromagnet, for the vacuum region ($z \leqslant 0$), we have

$$\Phi(z) = \Psi_0 \exp\{|k|z\}. \tag{11c}$$

Substituting the expressions for the magnetic potential (11) into the boundary conditions (7) yields an expression for the wave vector of the magnetostatic surface wave,

$$|k| = \beta D Q(\omega_S, 0), \tag{12}$$

where $\beta = \{\Gamma(2/3)/\Gamma(4/3)\}^3/9 \approx 0.3874$.

In calculating β we have used tabulated values of the gamma function.[30] A result for the wave vector (12) was obtained in the short-wavelength range, where the condition $|k| \gg L^{-1}$ must be satisfied. This leads to a requirement for the surface value of the derivative of the spectral function,

$$DQ(\omega_S, 0) \gg [\beta L]^{-1}. \tag{13}$$

MAGNETOSTATIC SURFACE WAVE SPECTRUM IN A FERROMAGNET WITH AN INHOMOGENEOUS GROUND STATE

We now make the results for the frequency of the surface magnetostatic wave (8) and the absolute value of the wave vector (12), as well as the conditions (9) and (13) for the inhomogeneous state (1a), more specific. A magnetostatic volume wave spectrum with lower $\omega_{V1}(z)$ and upper $\omega_{V2}(z)$ local boundaries (7) exists in an inhomogeneous medium if $A^2(z) > B(z)$ or

$$H_A(z) < \pi M_0 \left\{ \frac{DH_A(z)}{[4\pi M_0 - H_A(z)]k} \cot^2 \psi_0(z) \right\}^2. \tag{14}$$

Restricting ourselves to low fields $H_0 \ll 4\pi M_0 - H_A(z)$, we have

$$H_A(z) < H_{kp} = \frac{\pi M_0 H_0^4 [DH_A(z)]^2}{[4\pi M_0 - H_A(z)]^6 k^2}$$
$$\approx \frac{\pi M_0 H_0^4}{[4\pi M_0 - H_A(z)]^4 [kL]^2}. \tag{15}$$

This condition is clearly satisfied in an "easy plane" ferromagnet with $H_A(z) < 0$. For short wavelengths, with $kL \gg 1$, the term $A(z)$ in the expression for the upper limit $\omega_{V2}(z)$ can be regarded as small. Then, for the magnetostatic surface wave frequency (8), we obtain

$$\omega_S = \omega_V(0)[1 - \alpha], \tag{16}$$

where $\omega_V(0) = 2\gamma |\sin \psi_0(0)| \sqrt{-\pi M_0 H_A(0)}$ is the frequency of a magnetostatic volume wave in an infinite homogeneous ferromagnet with parameters equal to their surface values in the inhomogeneous medium, while

$$\alpha = H_0^2 DH_A(0) \frac{\text{sign } \psi_0(0)}{[4\pi M_0 - H_A(0)]^3 k} \sqrt{-\frac{\pi M_0}{H_A(0)}} \tag{17}$$

is a small correction leading to a dispersion $\sim k^{-1}$.

In the case where there is no external magnetic field ($H_0 = 0$) one has $\omega_S = \omega_V(0)$. To determine the absolute value of the wave vector according to Eq. (12), we calculate the derivative of the spectral function on the surface; the dispersion of the surface magnetostatic wave at short wavelengths can be neglected. As a result, we obtain

$$|k| = -\beta \frac{DH_A(0)}{H_A(0)} \left\{ 1 - \frac{2H_0^2 H_A(0)}{[4\pi M_0 - H_A(0)]^3} \right\}. \tag{18}$$

All these results have been obtained for short wavelengths, such that $|k| \gg L^{-1}$. This leads to a condition on the surface value of the logarithmic derivative of the local magnetic anisotropy field,

$$D \ln[-H_A(0)] \gg \frac{1}{\beta L} \left\{ 1 + \frac{2H_0^2 H_A(0)}{[4\pi M_0 - H_A(0)]^3} \right\}, \tag{19}$$

i.e., the local magnetic anisotropy field must increase near the surface. For other values of the logarithmic derivative of the local magnetic anisotropy field at the surface, the short-wavelength approximation does not hold and our equations are not applicable. If condition (19) is satisfied, then the final expression for the magnetostatic surface wave frequency is obtained from Eq. (16) by substituting the wave vector, i.e.,

$$\omega_S = \omega_V(0) \left\{ 1 - \sigma \text{ sign } \psi_0(0) \frac{H_0^2 \sqrt{-\pi M_0 H_A(0)}}{[4\pi M_0 - H_A(0)]^3} \right\},$$
$$\sigma = \text{sign } k, \tag{20}$$

where

$$\omega_V(0) = 2\gamma \sqrt{-\pi M_0 H_A(0)} \left\{ 1 - \frac{H_0^2}{2[4\pi M_0 - H_A(0)]^2} \right\}$$

is the magnetostatic volume wave frequency in a homogeneous ferromagnet with parameters equal to their values on the surface of the inhomogeneous medium.

If $\psi_0(0) > 0$, then for propagation in the positive Y direction ($\sigma = +1$), the magnetostatic surface wave frequency ω_S (20) will be somewhat lower than the magnetostatic volume wave frequency $\omega_V(0)$, and if the wave propagates in the negative Y direction ($\sigma = -1$), then ω_S is higher than $\omega_V(0)$. In the case $\psi_0(0) < 0$, for the positive Y direction ($\sigma = +1$) the magnetostatic surface wave frequency ω_S (20) will be somewhat higher than the magnetostatic volume wave frequency $\omega_V(0)$, and for the negative direction ($\sigma = -1$), lower. Thus, a nonreciprocity effect shows up. It involves a difference in the frequencies for identical magnitudes of the wave vector but different propagation directions of the wave. Equation (20) implies that it is observed only in an external magnetic field H_0. If there is no field ($H_0 \neq 0$), then $\omega_S = \omega_V(0)$, and the nonreciprocity effect does not occur.

Condition (11), which ensures that the spectral function is positive in the absence of a magnetic field ($H_0 = 0$), is satisfied for any profile of the local magnetic anisotropy field with

$$H_A(z) \geqslant H_K = H_A(0). \tag{21}$$

At low fields $H_0 \ll 4\pi M_0 - H_A(z)$, the expression for the critical field is somewhat different,

$$H_k = H_A(0) - \frac{H_0^2 H_A(0)\sqrt{-\pi M_0 H_A(0)}}{[4\pi M_0 - H_A(0)]^3}\left\{1 - \frac{DH_A(z)}{DH_A(0)}\right\}. \tag{22}$$

If $DH_A(z)/DH_A(0)$ is not a constant equal to 1, then H_K should depend weakly on the propagation direction of the wave. Thus we have determined the fixed values of the frequency (20) and wave vector (18) of the magnetostatic surface wave in the case where the ferromagnet has an inhomogeneous ground state. We have shown that these values are different in an external magnetic field for different directions of propagation of the wave (i.e., a nonreciprocity effect occurs). The local magnetic anisotropy field inside the ferromagnet must be greater than the critical field and increase near the surface. The critical field is determined mainly by the local magnetic anisotropy field at the surface.

MAGNETOSTATIC SURFACE WAVES IN A FERROMAGNET WITH A HOMOGENEOUS GROUND STATE

We now obtain these results for the homogeneous state (1b). The surface magnetostatic wave frequency (8) in this case is given by

$$\omega_S = \omega_V(0), \tag{23}$$

where $\omega_V(0) = \sqrt{\Omega_0(0)[\Omega_0 + 4\pi\gamma M_0]}$ is the magnetostatic volume wave frequency for a homogeneous ferromagnet with parameters equal to their values on the surface of a inhomogeneous medium. For the wave vector k, Eq. (12) implies that

$$|k| = -\beta \frac{DH_A(0)}{18\pi M_0}\frac{\Omega_0(0) + 2\pi\gamma M_0}{\Omega_0(0)}. \tag{24}$$

Thus, this sort of magnetostatic surface wave can exist only for inhomogeneities in the local magnetic anisotropy field which fall off near the surface, i.e., $DH_A(0) < 0$. Since our results have been obtained for the short-wavelength case, Eq. (13) can be satisfied and is equivalent to the requirement that

$$DH_A(0) \ll -\frac{18\pi M_0}{\beta L}\frac{\Omega_0(0)}{\Omega_0(0) + 2\pi\gamma M_0}. \tag{25}$$

In the interior of the ferromagnet for $z > 0$, the spectral function $Q(\omega_S, z)$ was assumed positive. Then the upper local limit $\Omega_{V2}(z)$ for arbitrary $z \geqslant 0$ should satisfy the inequality (9), which leads to an inequality for the local magnetic anisotropy field,

$$H_A(z) \leqslant H_A(0). \tag{26}$$

Thus, a magnetostatic surface wave with the frequency (23) and wave vector (24) exists for a local magnetic anisotropy field which is smaller than that on the surface and falls off near the surface.

CONCLUSIONS

In summary, we have demonstrated the possible existence of a single magnetostatic surface wave with a fixed frequency and wave vector in an inhomogeneous ferromagnet for which the spectral function has a turning point on the surface. This kind of magnetostatic surface wave is obtained from a magnetostatic volume wave propagating at the surface of the ferromagnet by choosing a profile for the local magnetic anisotropy field which precludes the existence of a magnetostatic volume wave in the interior of the ferromagnet and ensures exponential damping of the magnetic potential. This significantly new result is of greatest interest for a ferromagnet that has been magnetized perpendicular to its surface, where only a magnetostatic volume wave can exist in the homogeneous case. The fact that the frequency and wave vector are fixed is nontraditional. They are uniquely determined by the values of the local magnetic anisotropy field and its derivative at the surface. The conditions for the existence of the predicted magnetostatic surface wave depend on the choice of the ground state of the ferromagnet.

[1] C. Vittoria and J. H. Schelleng, Phys. Rev. B **18**, 4020 (1977).
[2] C. Borghese, P. De Gasperies, and R. Tappa, Solid State Commun. **25**, 21 (1978).
[3] A. H. Eschenfel'der, *Magnetic Bubble Technology*, Springer-Verlag, Berlin (1980); Mir, Moscow (1983), 496 pp.
[4] V. S. Speriosu and C. H. Wilts, J. Appl. Phys. **54**, 3325 (1983).
[5] C. H. Wilts, H. Awano, and V. S. Speriosu, J. Appl. Phys. **57**, 2161 (1985).
[6] C. H. Wilts and S. Prased, IEEE Trans. Magn. **MAG-17**, 2045 (1981).
[7] V. É. Osukhovskiĭ, D. E. Linkova, Z. Z. Ditina *et al.*, Fiz. Tverd. Tela (Leningrad) **26**, 1533 (1984) [Sov. Phys. Solid State **26**, 933 (1984)].
[8] G. A. Shmatov, V. N. Filippov, V. B. Sadkov, and I. I. Kryukov, Pis'ma Zh. Tekh. Fiz. **15**(17), 86 (1989) [Sov. Tech. Phys. Lett. **15**, 699 (1989)].
[9] A. I. Akhiezer, V. G. Bar'yakhtar, and S. V. Peletminskiĭ, *Spin Waves*, North-Holland, Amsterdam (1968) [Russian original, Nauka, Moscow (1967), 368 pp.].
[10] V. G. Bar'yakhtar and M. I. Kaganov, ''Inhomogeneous resonance and spin waves,'' in *Ferromagnetic Resonance* [in Russian], Fizmatgiz, Moscow (1961), pp. 266–284.
[11] G. A. Vugal'ter and I. A. Gilinskiĭ, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **32**, 1187 (1989).
[12] B. N. Filippov, Fiz. Met. Metalloved. **32**, 911 (1971).
[13] B. N. Filippov and I. G. Tityakov, Fiz. Met. Metalloved. **35**, 28 (1973).
[14] R. E. De Wames and T. Wolfram, J. Appl. Phys. **41**, 987 (1970).
[15] B. R. Tittman, R. E. De Wames, R. D. Henry, and P. J. Besser, Trudy MKM-73, Vol. 2, pp. 19–27 (1974).
[16] J. T. Yu, R. A. Turk, and P. E. Wigen, Phys. Rev. B **5**, 420 (1972).
[17] F. R. Morgenthaler, IEEE Trans. Magn. **MAG-13**, 1252 (1977).
[18] F. R. Morgenthaler, J. Appl. Phys. **52**, 2267 (1981).
[19] P. Hartemann and D. Fontaine, IEEE Trans. Magn. **MAG-18**, 1595 (1982).
[20] N. I. Lyashenko and V. M. Talalaevskiĭ, Ukr. Fiz. Zh. **31**, 1716 (1986).
[21] Yu. V. Gulyaev, I. A. Ignat'ev, A. F. Popkov, and V. M. Shabunin, in *Abstracts of the 10th All-Union School–Seminar on New Magnetic Materials for Microelectronics* [in Russian], Riga (1986), Part 1, pp. 176–177.
[22] 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)].
[23] Yu. M. Yakovlev, E. G. Rzhakhina, T. A. Kyrlova *et al.*, Fiz. Tverd. Tela (Leningrad) **30**, 622 (1988) [Sov. Phys. Solid State **30**, 360 (1988)].
[24] I. V. Zavislyan and A. F. Kalaĭda, Vestn. Kievsk. Univ. Ser. Fiz., No. 23, pp. 75–79 (1982).
[25] I. Yu. Gaĭovich, G. P. Golovach, I. V. Zavislyan, and V. F. Romanyuk, Fiz. Tverd. Tela (St. Petersburg) **34**, 1680 (1992) [Sov. Phys. Solid State **34**, 893 (1992)].
[26] I. A. Viktorov, *Acoustic Surface Waves in Solids* [in Russian], Nauka, Moscow (1981), 288 pp.

[27] E. Dieulesant and D. Royer, *Elastic Waves in Solids*, Wiley, New York (1980); Nauka, Moscow (1982), 434 pp.

[28] A. Nayfeh, *Introduction to Perturbation Techniques*, Wiley, New York (1981); Mir, Moscow (1984), 534 pp.

[29] M. V. Fedoryuk, *Asymptotic Methods for Linear Ordinary Differential Equations* [in Russian], Nauka, Moscow (1983), 352 pp.

[30] E. Jahnke, F. Emde, and F. Loesch, *Tables of Higher Functions*, 6th ed., McGraw-Hill, New York (1960); Nauka, Moscow (1977), 344 pp.

Translated by D. H. McNeill

Influence of the edge field on the focusing properties of a coaxial cylindrical lens

L. P. Ovsyannikova and T. Ya. Fishkova

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

The effect of the distance between the grounded flat entrance diaphragm and the outer cylindrical electrode (which determines the edge field of the lens) is investigated over a wide range of variation of the geometry of a coaxial cylindrical lens. It is found that focusing of a charged particle beam on the lens axis is achieved over a wide range only in the case of small clearances between the diaphragm and the outer electrode. It is shown that for an alternative power feed arrangement, in which the outer cylindrical electrode is grounded and the voltage is applied to the inner electrode, beam focusing is in general degraded, in particular on the lens axis. © *1998 American Institute of Physics.* [S1063-7842(98)02206-5]

In Refs. 1 and 2 we identified the working regimes of a coaxial cylindrical lens focusing an annular charged-particle beam on its axis and investigated it over a wide range of variation of its geometrical and electrical parameters. We also obtained simple empirical formulas for the cardinal elements of this lens. The clearance between the grounded end-face diaphragm at the lens entrance and the external cylindrical electrode was significantly less than the transverse and longitudinal dimensions of the lens. Since this clearance determines the configuration of the edge field, it should have a substantial influence on the focusing properties of the lens. The influence of the position of the entrance diaphragm is in fact the subject of the present paper.

Figure 1a presents a diagram of a coaxial cylindrical lens consisting of two cylindrical electrodes, a flat end-face diaphragm at the entrance with an open back face. The inner electrode and the diaphragm were constructed as a single unit with a grounded housing, onto which the outer electrode was mounted through an insulator. Application of a voltage *V* to the outer electrode while the diaphragm and inner electrode remain grounded leads to the appearance of a field on the lens. Therefore, in what follows we will call the outer cylindrical electrode the field-assigning electrode.

In such a lens, the equipotentials of small absolute values of the potential pass near the inner electrode, then run nearly parallel to the end-face diaphragm, and finally close in the space between the housing and the inner electrode. Here, as a consequence of the open back face they penetrate into the space beyond the electrodes. The equipotentials of large absolute values of the potential have a simpler configuration, smaller extent, and surround the field-assigning electrode from the inside and out. It can be seen from Fig. 1b that in the region of the diaphragm located near the field-assigning electrode (the edge-field region at the lens entrance), the values of the potential at fixed radius grow with increasing value of the longitudinal coordinate whereas away from the diaphragm (the edge-field region at the lens exit), near the inner electrode, the potential at first grows and then falls. In the rest of the exit edge-field region the dependence of the potential on the longitudinal coordinate is analogous to this

dependence for the entrance edge field. It should be stated that near the diaphragm but far from the field-assigning electrode, the field distribution is similar to the field distribution away from the diaphragm. The values of the voltages determining the form of the equipotentials and the rate of falloff of the edge field away from the field-assigning electrode depend on the geometry of the lens.

The charged particle beam trajectories were calculated numerically using the TEO computer code for two-dimensional electrostatic fields. We considered beams entering the coaxial cylindrical lens parallel to its longitudinal
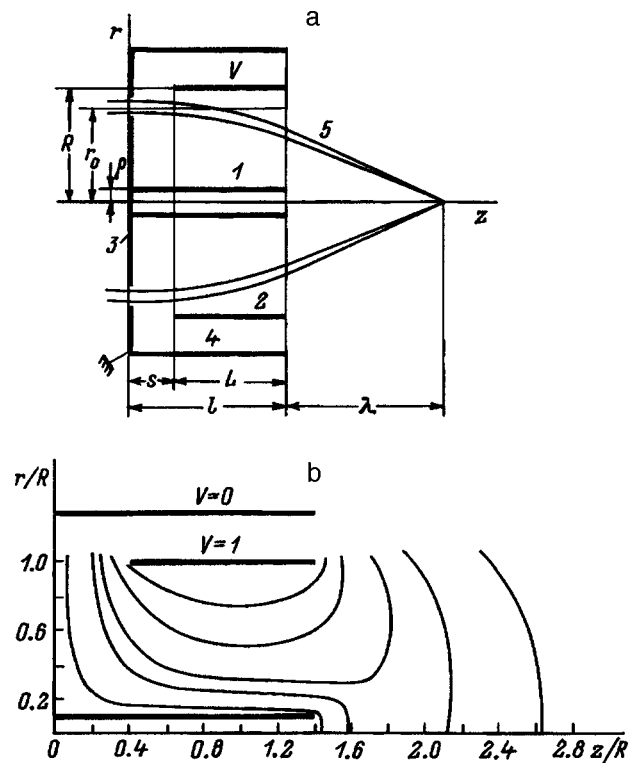

FIG. 1. a:—Coaxial cylindrical lens: *1,2*—cylindrical electrodes, *3*—flat diaphragm, *4*—housing, *5*—charged particle trajectories; b—equipotential plot.

axis and determined the working regimes focusing the beam on the lens axis. For the characteristic geometry of the lens with ratio of the radii of the large and small cylindrical electrodes $R/\rho = 10$ the distance from the diaphragm to the field-assigning electrode was varied over wide limits $0 < s/R \leqslant 1.5$. The length of the latter was varied within the limits $0.5 \leqslant L/R \leqslant 1.9$ so that the length of the lens from the diaphragm to the back face did not change ($l = L + s = 2R$ $=$ const). The minimum clearances s/R were governed by the magnitude of the breakdown voltage, the maximum clearances lead to vigorous growth of the lens excitation due to decrease of the length of the field-assigning electrode.

Figure 2 displays the calculated results. It plots the focal length, measured from the back of the lens, and the entrance radius of the paraxial beam trajectory—the trajectory around which the other beam trajectories are focused, as functions of the lens excitation F ($F = eV/\varepsilon$, where e is the charge of the particle and ε is its energy). The small brackets mark off the minimum and maximum excitations bounding the existence region of the focus. For small clearances ($s \leqslant 0.3R$) focusing is realized over a wide range of variation of the distance from lens exit to focus (focal length) ($100 \leqslant \lambda/R \leqslant 0.2$); here the value of the minimum achievable focal length decreases as the clearance is decreased. At large clearances $s \cong R$ the focal region substantially narrows, and for $s = 0.5R$ focusing is generally absent. The entrance radius of the paraxial beam trajectory in the focusing regime falls as the excitation is
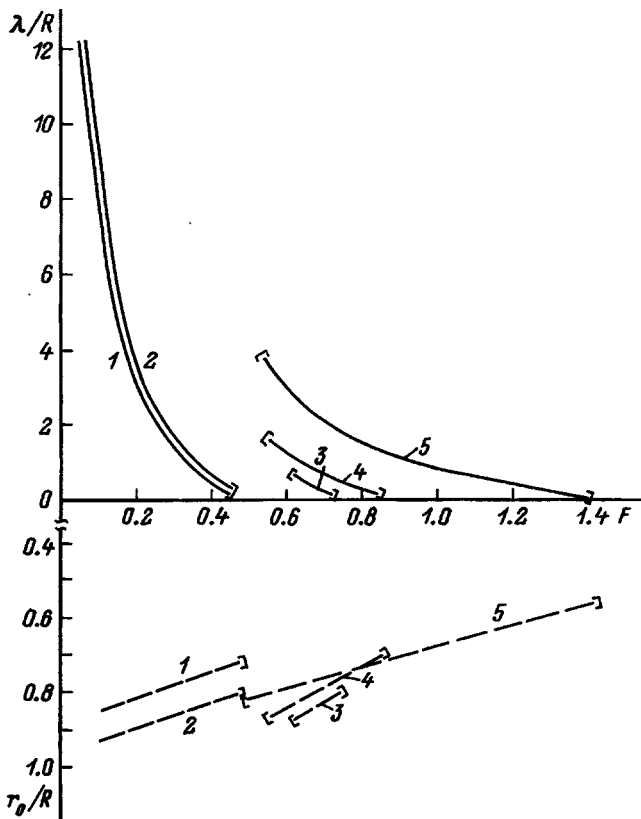

FIG. 2. Dependence of the focal length (solid curves) and radius of the paraxial beam trajectory (dashed curves) on the excitation for a lens with $R/\rho = 10$ and $l/R = 2 =$ const for various clearances between the diaphragm and the field-assigning electrode. s/R: 1—0.1, 2—0.25, 3—0.75, 4—1.0, 5—1.5.

increased according to a linear law, whereas with increase of the clearance it at first grows (curves 1–3) and then falls (curves 4 and 5).

From the aforesaid it can be concluded that for focusing of a parallel beam onto the lens axis, coaxial cylindrical lenses should be designed with a small clearance between the entrance diaphragm and the field-assigning electrode.

It is of interest to investigate the influence of the edge field on focusing of a coaxial cylindrical lens for which the length of the field-assigning electrode was not varied. The latter was chosen from the condition that the potential difference between the electrodes not exceed the accelerating potential, i.e., a lens excitation $F \leqslant 1$. We calculated a lens with $R/\rho = 10$ and $L = R =$ const for which we varied the clearance within the limits $0 < s/R \leqslant 2$. In this case the overall length of the lens from the diaphragm to the back face varied within the range $1 < lR \leqslant 3$. Results of the numerical calculation are plotted in Fig. 3. As can be seen, focusing takes place over the entire range of variation of the clearance which determines the edge field. However, starting at $s = 0.4R$ (curves 3–6) the focal region is found near the lens and does not extend beyond $\lambda \cong 2R$, and as the clearance is increased this region decreases in extent. Note that the focal length and the initial radius of the paraxial beam trajectory for the clearances $s = R$ and $s = 2R$ practically coincide. Consequently, from the point of view of focusing it is without sense to make the clearance greater than the radius of the outer cylindrical electrode.

For small clearances between the entrance diaphragm and the field-assigning electrode ($s \leqslant 0.3R$) focusing exists within wide limits, and while the focal length varies only slightly with the clearance (curves 1 and 2 in Fig. 3a), the radius of the paraxial trajectory at the entrance to the coaxial cylindrical lens varies considerably (curves 1 and 2 in Fig. 3b). Note that with increase of the ratio of the electrode radii the region in s in which focusing is realized increases somewhat and for $R/\rho = 100$ is $0 < s \leqslant 0.4R$. As this ratio is decreased, the indicated region narrows substantially, and for $R/\rho = 2$ the maximum clearance is $s = 0.1R$.

In the choice of the optimal geometry of a coaxial cylindrical lens, the quality of focusing is important, where the latter is defined by the radius of the spot formed by the lens when focusing a ring beam on the lens axis. Figure 4 plots the radius of the spot in the plane passing through the intersection point of the longitudinal axis of the lens and the paraxial beam trajectory (the trajectory around which the other trajectories are focused) as a function of the clearance between the diaphragm and the field-assigning electrode. The solid curves correspond to ring thickness at lens entrance $\Delta r_0 = 0.05R$, and the dashed ones, to $\Delta r_0 = 0.1R$. In the small clearance region the radius of the focal spot is two orders of magnitude smaller than the thickness of the entrance ring while for $s \geqslant R$ this radius grows substantially. In addition, for $s \geqslant 0.4R$, as can be seen from Fig. 3, the focal region is small. Therefore the geometry of a coaxial cylindrical lens with a large clearance between the entrance diaphragm and the field-assigning electrode is not of interest from the point of view of beam focusing on the lens axis.

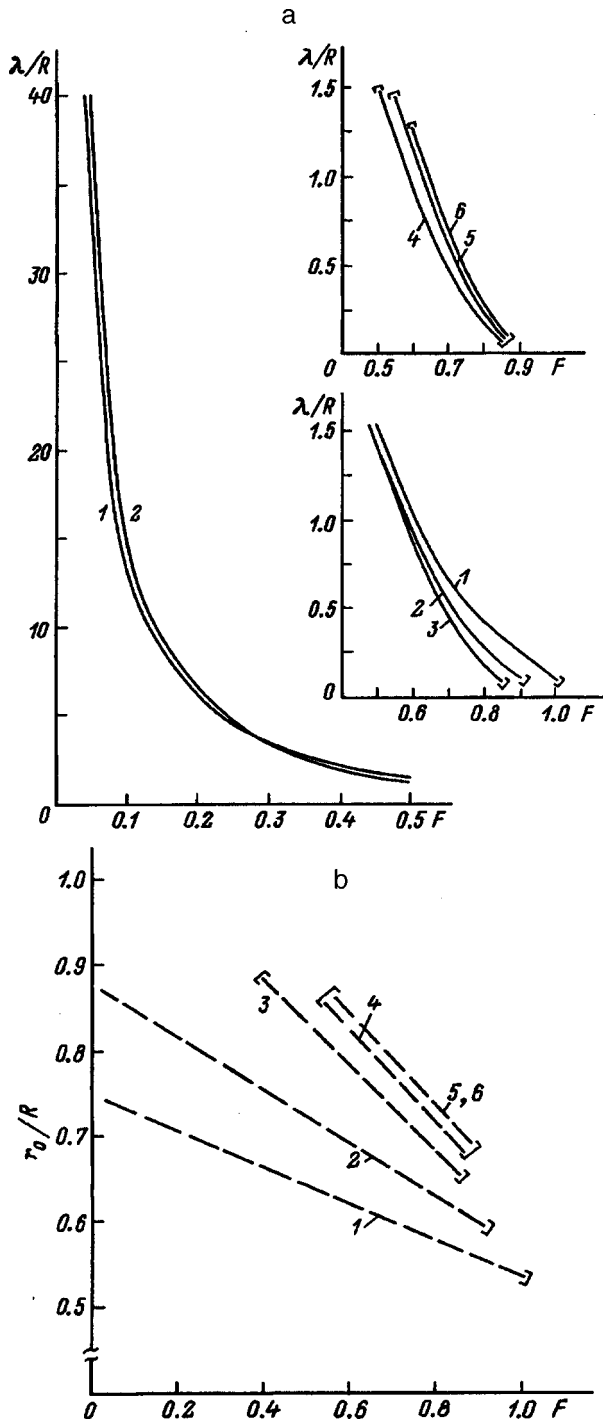Figure 5 plots the radius of the axially focused beam

a

b

FIG. 3. Same as in Fig. 2 for a lens with $R/\rho = 10$ and $L/R = 1$ (a,b). s/R: *1*—0.1, *2*—0.25, *3*—0.4, *4*—0.5, *5*—1.0, *6*—2.0.
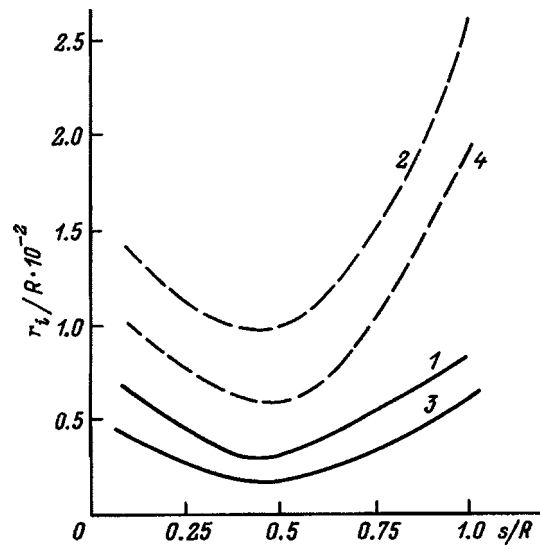

FIG. 4. Radius of the focused spot as a function of the clearance between the diaphragm and the field-assigning electrode for a coaxial cylindrical lens with $R/\rho = 10$ and $L/R = 1$. Solid curves—$\Delta r_0 = 0.05R$, dashed curves—$\Delta r_0 = 0.1R$; *1,2*—$\lambda = 0.5R$; *3,4*—$\lambda = R$.

Thus, for a coaxial cylindrical lens, over a wide range of variation of the ratio of electrode radii for various lengths of the field-assigning electrode, the optimum clearance between the entrance diaphragm and the field-assigning electrode from the point of view of focusing the beam on the axis is $s \leqslant 0.1R$. In this case, to determine the focal length and the radius of the paraxial beam trajectory (the trajectory about which focusing takes place) it is possible to use the corresponding formulas from Ref. 2.

In conclusion, note that we have also investigated a coaxial cylindrical lens with an alternative power feed arrangement, where the voltage is fed to the inner electrode while the outer electrode and diaphragm remain grounded. In this case the design of the lens simplifies since the need for a housing falls away and the inner electrode is mounted to the diaphragm through an insulator. However, such a power feed arrangement leads to a substantial change in the edge field at the entrance to the lens and, as a consequence, causes the beam to behave in a different way, leading to degradation of

over a wide range of variation of the focal length for small clearances between the diaphragm and the field-assigning electrode for a coaxial cylindrical lens with ratio of electrode radii $R/\rho = 10$. It can be seen from the figure that only in the region $2 \leqslant \lambda/R \leqslant 10$ is the spot radius for a lens with clearance $s = 0.25R$ smaller than for $s = 0.1R$. Outside this region of focal length values the clearance value $s = 0.1R$ provides the smallest spot size, whose magnitude varies only slightly (for $\Delta r_0 = 0.005R$ the spot radius $r_i \cong 0.003R$, while for $\Delta r_0 = 0.1R$ it is equal to $r_i \cong 0.009R$).
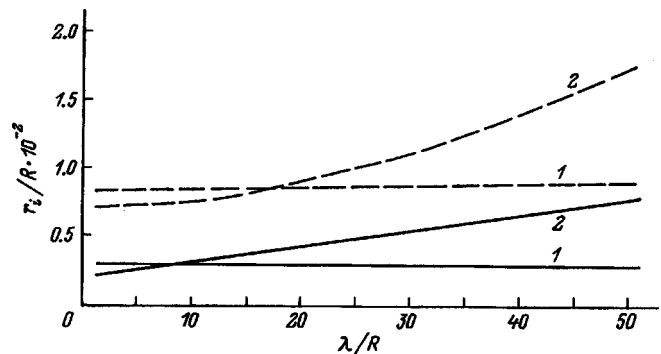

FIG. 5. Dependence of the radius of the focal spot on position measured from the back of the lens for small clearances between the diaphragm and the field-assigning electrode. s: *1*—0.1R, *2*—0.25R; solid curves—$\Delta r_0 = 0.05R$, dashed curves—$\Delta r_0 = 0.1R$.

focusing in general and on the axis of the lens in particular. This has to do with the fact that the particles in the edge field are accelerated, and this acceleration and also the radial force of the lens decrease with growth of the radius. For the main voltage feed arrangement (Fig. 1) the particles are slowed down, and this slowing down and the radial force of the lens

grow with the radius.

[1] L. P. Ovsyannikova and T. Ya. Fishkova, Pis'ma Zh. Tekh. Fiz. **22**(16), 39 (1996) [Tech. Phys. Lett. **22**, 660 (1996)].
[2] L. P. Ovsyannikova and T. Ya. Fishkova, Zh. Tekh. Fiz. **67**(8), 89 (1997) [Tech. Phys. **42**, 935 (1997)].

Translated by Paul F. Schippnick

Focusing of electrons reflected from a crystal with loss of energy

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

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

A study is made of the features arising in the spatial distributions of reflected electrons as a result of a focusing effect. Experiments are conducted on single-crystal Mo (100) with primary electron energies of 0.5–2 keV and detection of electrons which lose fixed amounts of energy up to 300 eV. An analysis of the data establishes the dependence of the electron focusing efficiency on the amount of energy loss. It is shown that when electrons are reflected with single losses through plasmon excitation, the magnitude of the effect is determined mainly by the average number of scattering atoms encountered by an electron along its path to the surface. When the energy losses are high, defocusing owing to multiple elastic and inelastic scattering of the electrons is found to predominate. © *1998 American Institute of Physics.* [S1063-7842(98)02306-X]

INTRODUCTION

In recent years, two types of diffraction effects observed during bombardment of test objects by medium energy electrons have come into active use for the structural analysis of solid surfaces. The first is the diffraction of reflected electrons, which shows up in the form of peaks in their spatial distributions oriented along close-packed atomic rows and planes of the crystal,[1–3] and the second is the dependence of the intensity of electron reflection and Auger electron emission on the angle of incidence of the primary electrons relative to the crystal axes.[4,5] Here the emission peaks appear for the same orientations of the incident electrons at which the peaks in the reflected electron diffraction patterns are observed. Thus far, it has been established that electron focusing effects in the crystal play a dominant role in the formation of these diffraction features. However, not all aspects of this phenomenon have been studied adequately. For example, there have been almost no studies of the focusing behavior of electrons reflected with different energy losses. Only a few papers touch on this problem.[6–9] In the meantime, it is known that the probe depth, the contrast of the diffraction patterns, etc., depend on the magnitude of the energy loss. Thus, it seemed appropriate to us to make a systematic study of the spatial distributions of electrons reflected with different energy losses over a wide range of exit angles. Single crystal Mo (100), which has been used in earlier work, was chosen as the object of study.

EXPERIMENTAL TECHNIQUE

The measurements were made in a special ultrahigh-vacuum angle-resolved secondary electron emission spectrometer that has been described elsewhere.[10] The energy resolution of the modified-plane-mirror analyzer was 0.4%, and the angular resolution was about 1°. The electron beam of the spectrometer bombarded the sample with electrons along the normal to its surface. The energy analyzer could be rotated around the sample, making it possible to vary the polar angle θ of the escaping electrons. The azimuthal exit angle φ could be varied by rotating the sample about an axis perpendicular to its surface. This instrument design made it possible to measure the distributions $I(\varphi)$ of the reflected electrons with respect to the azimuthal exit angle φ for different polar angles θ and thereby to obtain almost complete diffraction patterns. The energy ΔE lost by the electrons upon reflection was varied as a parameter. The intensity of the electron flux was measured by modulating the current of the primary beam. Measurements were made for primary electron energies E_r of 0.5–2 keV. Principal attention was devoted to an energy of 1.25 keV, which is high enough for the focusing effect to show up but not so high that the energy resolution of the analyzer no longer allows the main peaks in the characteristic electron energy loss spectrum to be resolved.

The method of preparing the test sample has been described elsewhere.[11] The cleanliness of the surface of the molybdenum single crystal was monitored by electron Auger spectroscopy and the structure of its surface region, by low-energy electron diffraction (LEED). The measurements were made at room temperature in a vacuum of 5×10^{-10} Torr.

RESULTS OF THE MEASUREMENTS AND DISCUSSION

A general idea of the appearance of the electron focusing effect in the simplest case of quasielastic reflection ($\Delta E < 1$ eV) is provided in Fig. 1a, which shows a two dimensional map of the distribution of the intensity $I(\theta, \varphi)$ of the electrons over the polar and azimuthal exit angles, obtained by synthesis of a family of azimuthal scans measured at an energy $E_r = 2$ keV. The data are shown in a stereographic projection. The center of the circle corresponds to the normal to the surface of the sample, and the outer circle to emission of electrons along the surface. The distribution is shown in a linear scale of gray shadings in which the maximum reflection corresponds to white and the minimum, to black. Quasielastic electron reflection from single-crystal Mo (100) is
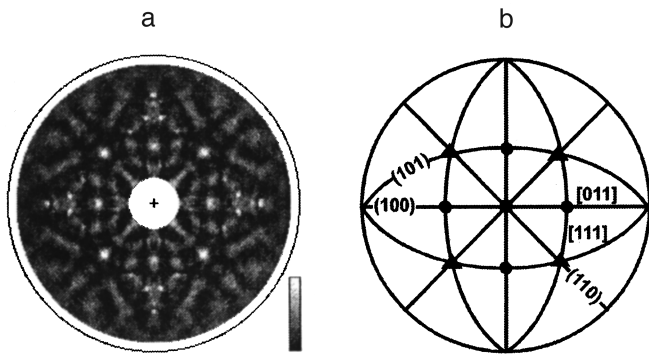
FIG. 1. A two dimensional map of the intensity distribution $I(\theta,\varphi)$ of quasielastic electron reflection over the polar and azimuthal exit angles obtained at an energy $E_r = 2$ keV for single crystal Mo (100) (a) and a stereographic projection of this face indicating the most close-packed planes and directions of the crystal (b).
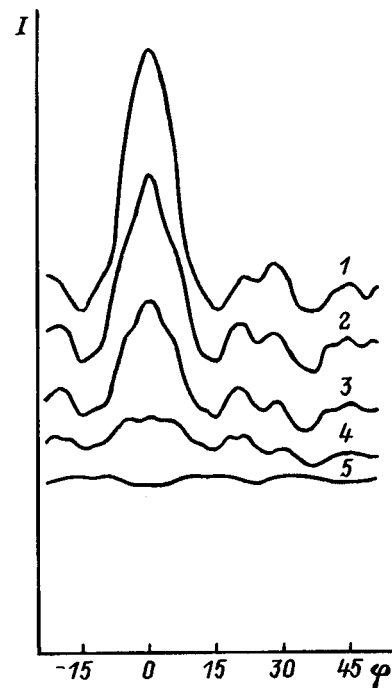

FIG. 2. Azimuthal angular distributions of electrons inelastically reflected from single crystal molybdenum, measured for a primary electron energy $E_r = 1.25$ keV and a polar angle of emission of 55° (from the normal to the sample surface). The energy losses ΔE for different groups of electrons: *1*—less than 1 eV (quasielastically reflected electrons), *2*—24 eV (electrons which have lost energy exciting a bulk plasmon), *3*—50, *4*—100, and *5*— 250 eV.

evidently highly anisotropic and has a distinct diffraction pattern. Its symmetry reflects the symmetry of the (100) face of the body-centered cubic crystal, as illustrated in Fig. 1b, which shows the corresponding stereographic projection. A comparison of the data of Fig. 1a with this projection makes it possible to identify the main peaks in this pattern at once. Clearly, they are caused by the escape of electrons along close-packed directions of the crystal, such as $\langle 111 \rangle$ and $\langle 110 \rangle$. These peaks were observed over the entire range of energies that was studied and are caused by focusing of electrons moving along strings of atoms with small interatomic distances. These results are all in good agreement with data on quasielastic scattering of electrons which we have obtained previously[11] for another face of molybdenum, Mo (110).

The focusing effect also shows up in the spatial distributions of the electrons reflected from the crystal with a loss of energy. As an illustration, Fig. 2 shows several typical azimuthal distributions of the electrons reflected with losses $\Delta E < 300$ eV. (In accordance with the symmetry of the Mo (100) face, the range of variation of the azimuth here has been limited to half a quadrant). These data were obtained for $E_r = 1.25$ keV and a polar emission angle of $\theta = 55°$. A distinct diffraction structure is visible in all the curves. The strongest feature is the peak observed at $\varphi = 0$, which is caused by focusing of electrons along the most close-packed crystal direction, $\langle 111 \rangle$. Weaker peaks are observed at azimuths of $\varphi = 18$ and 45°, corresponding to orientations of the escaping electrons along the (130) and (100) planes. The shape of the $I(\varphi)$ curves depends on the magnitude of the energy losses experienced by the electrons. As ΔE increases, there is a noticeable drop n the focusing peak for electrons along the $\langle 111 \rangle$ direction, which essentially vanishes by $\Delta \approx 200$ eV, and then is inverted to form a minimum. Such behavior is typical of the focusing peaks observed for electrons moving along other close-packed directions.

For a quantitative estimate of the focusing effect for electrons reflected along the low-index directions $\langle hkl \rangle$ with different energy losses we can use

$$\chi = [(I_{max} - I_{min})/I_{max}] \cdot 100\%,$$

where I_{max} and I_{min} are the intensities of the electron flux in the analyzed peak of the $I(\varphi)$ distribution and in its deepest minimum, respectively.

By analyzing a family of azimuthal distributions measured for a number of values of ΔE and θ, it is possible to determine the electron focusing efficiency $\chi(\Delta E)$ as a function of the magnitude of the energy losses for the major close-packed directions of the crystal. In order to obtain this information, the distributions $I(\varphi)$ were measured for 6 different angles θ with a step in ΔE amounting to 2 eV. These data are represented in Fig. 3. It is clear that for small energy losses (in the region $\Delta E < 50$ eV), the focusing is quite strong and χ can reach 75%. It is noteworthy that χ depends monotonically on ΔE. For most of the curves, nonmonotonicities are observed at the same values of ΔE. The fine structure of the $\chi(\Delta E)$ curves for small angles θ looks like it is superimposed on a horizontal straight-line background, and as the polar angle is increased, the background falls off with increasing ΔE. This is typical of the $\chi(\Delta E)$ curves for large energy losses ($\Delta E > 50$ eV), as well. However, even here the rate of decrease in the degree of focusing with rising ΔE depends on the polar angle, increasing with rising θ. For sufficiently large angles, the $\chi(\Delta E)$ curves are observed to settle into a region of negative values corresponding to the above-noted inversion in the diffraction structure of the angular distributions $I(\varphi)$ for electrons reflected with large energy losses.

In order to clarify the nature of the observed nonmonotonicities in $\chi(\Delta E)$, let us compare them with the character-
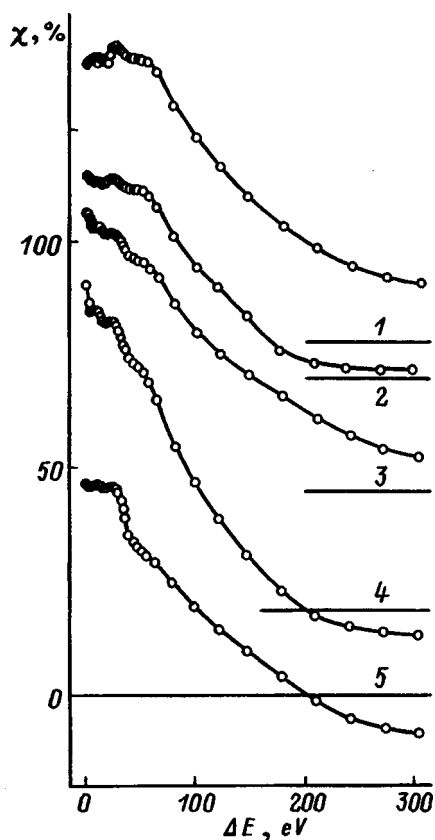
FIG. 3. Plots of $\chi(\Delta E)$, the electron focusing efficiency as a function of the energy lost by the electrons during reflection. These data were obtained for $E_r = 1.25$ keV and apply to the following close-packed directions of the crystal: *1*—$\langle 012\rangle$, *2*—$\langle 112\rangle$, *3*—$\langle 110\rangle$, *4*—$\langle 111\rangle$, and *5*—$\langle 122\rangle$. All the curves are plotted on the same scale as that indicated for curve *5*. For the other curves, only the lines corresponding to the shifted zero ordinate are indicated.

istic electron energy loss spectrum in molybdenum. One such spectrum is shown in Fig. 4. It was taken for $E_r = 0.5$ keV and a polar angle of 45°. There are three distinct maxima in this spectrum. The first, observed at ΔE
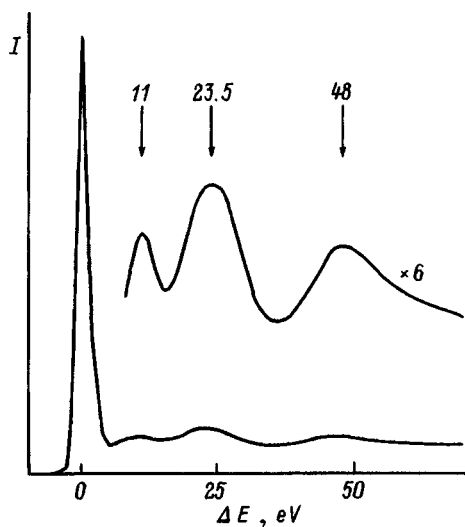

FIG. 4. Characteristic electron energy loss spectrum in molybdenum measured for $E_r = 500$ eV.

$= 11$ eV, corresponds to overlapping excitation peaks for a surface plasmon ($\hbar\omega_s = 9.5$ eV) and a low-energy bulk plasmon ($\hbar\omega_v = 10.4$ eV).[12] The second, larger loss at $\Delta E = 23.5$ eV is caused by the generation of the fundamental bulk plasmon of molybdenum. The third peak at $\Delta E = 48$ eV is usually attributed to dynamic polarization of electrons in the shallow core $4p$ level,[13] but some part of it may come from double excitation of bulk plasmons. Besides these peaks, the spectrum includes a noticeable background of multiple losses, which increases with rising ΔE. Comparing the data of Figs. 3 and 4 shows that electron focusing increases if the electrons experience single energy loss through excitation of bulk plasmons as they are reflected. For small polar escape angles, the anisotropy of the distributions for these electrons is greater than in the case of quasielastic scattering.

Let us analyze these data in terms of a simple model that we have used before to study the focusing of quasielastically scattered[14] and primary electrons.[15] This model is based on the assumption that the reflection of electrons from a solid is the result of single, large-angle electron–phonon scattering events. Although a rigorous solution of the problem will include multiple quasielastic scattering processes,[16] this model is entirely appropriate for describing the motion of electrons in a thin subsurface layer of the crystal. In this case the trajectory of the electrons inside the solid can be approximated by a broken line consisting of two straight segments. The first of these (l_1) corresponds to motion of the electron into the interior of the crystal (to a point where a quasielastic scattering event occurs) and the second segment (l_2), to the motion of the scattered electron toward the surface. Note that, since the focusing of the reflected electrons takes place along path l_2, this is the parameter in the problem which should have a controlling influence on the magnitude of the observed effect.

Estimating l_2 is simplest in the case of quasielastic scattering of the electrons, with $l_2 = \lambda/(1 + \cos\theta)$, where λ is the mean free path of the electron with respect to an inelastic interaction. It is clear that the length of the focusing string in this case becomes greater as θ increases. For inelastic reflection of electrons involving single excitation of plasmons, yet another elementary act is involved which may take place before or after the quasielastic scattering into the backward hemisphere. Since plasmon excitation takes place through a long-range Coulomb interaction with the electronic subsystem of the crystal, the probability of generating plasmons is essentially independent of electron focusing.[15] In addition, since the electron generates long-wavelength, small-momentum plasmons with a higher probability,[17] there is no significant change in the direction of motion of a 1 keV electron. Thus, the events in which plasmons are generated should have little effect on the focusing process. Thus, we assume that the observed differences in χ for electrons which excite plasmons and are scattered only quasielastically originate primarily in a difference in the average exit depths of these groups of electrons and, therefore, in the different lengths of the focusing strings. Since the exit depth is roughly twice as large for electrons reflected with plasmon excitation as for quasielastically scattered electrons, we may
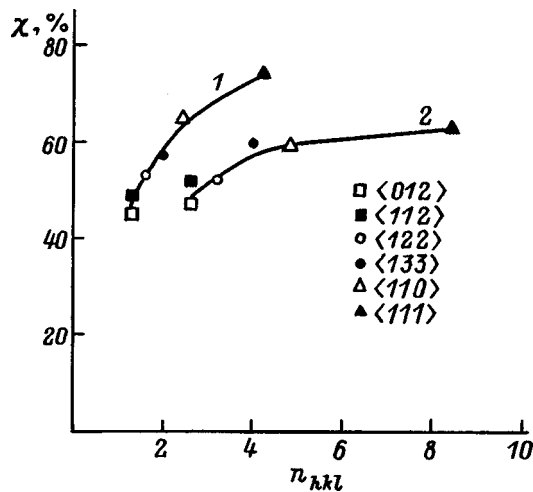
FIG. 5. Plots of the electron focusing efficiency $\chi(n)$ as a function of the number of scatterers encountered on the path of the electrons toward the surface for emission along different close-packed directions of the crystal. *1*—Quasielastic reflection, *2*—reflection with single excitation of a bulk plasmon; $E_r = 1.25$ keV.

assume that the length l_2 is also twice as large. Thus, by comparing the $\chi(\Delta E)$ curves obtained in the single energy loss region for different angles θ, we can understand how the focusing efficiency for electrons moving along different crystallographic directions depends on their path lengths.

The focusing properties of the atomic strings depend on the characteristic distance between the atoms, as well as on their length. Thus, it is appropriate to present the data of Fig. 3 as plots of $\chi(n)$, where n is the average number of scatterers encountered along an electron's path as it moves toward the surface along a given string. In terms of our model, this number can be estimated as follows: $n = l_2/d$, where d is the interatomic distance along the given direction. Plots of $\chi(n)$ obtained on the basis of data for quasielastic scattering and scattering with single excitation of a bulk plasmon are shown in Fig. 5 (curves *1* and *2*, respectively). They characterize the focusing of electrons by different strings of atoms oriented along 6 low-index directions of the crystal ($\langle 111 \rangle$, $\langle 110 \rangle$, $\langle 133 \rangle$, $\langle 012 \rangle$, $\langle 112 \rangle$, $\langle 122 \rangle$) with interatomic distances varying between 2.72 and 9.45 Å. Pairs of points referring to the same direction are denoted by the same symbol. It is clear from the figure that all the data fit fairly well on two monotonically increasing curves. This means that for each group of electrons, the number of scatterers in the string plays a controlling role in focusing them, while the interatomic distance is less important. The data for the $\langle 122 \rangle$ direction are especially characteristic in this regard. For the other atomic strings, an increase in the polar angle θ, leading to lengthening of l_2 and an increase in n, is accompanied by a monotonic rise in χ, but in this case, although the path length l_2 is quite large ($\theta = 70°$), the values of n are quite small (1.6 and 3.2) because of the large interatomic distance and, accordingly, the values of χ are also lower.

A monotonic growth in the focusing efficiency of quasielastically reflected electrons is observed at least up to four scatterers, and for electrons reflected in processes involving plasmon excitation, up to eight. At the same time,

both dependences are nonlinear and reveal a distinct slowing down in the growth of χ with increasing n. This is evidence of the onset of an electron defocusing process,[11] which arises because of multiple elastic scattering of electrons in sufficiently long atomic strings. The effect of defocusing becomes especially noticeable when points on curves *1* and *2* corresponding to the same direction are compared pair by pair. In fact, while a doubling of the length of strings having a small number of scatterers (going from $n = 1–2$ to $n \approx 3$) slightly increases the degree of focusing of electrons that have excited plasmons, in the case of the $\langle 111 \rangle$ and $\langle 110 \rangle$ directions, where the initial n are already quite large, a further increase in n has the contrary effect of making the points on curve *2* appear to lie below those on curve *1*.

The weakening of the diffraction structure of the distributions for high energy losses (the drop in the curves in Fig. 3 for $\Delta E > 50$ eV) is evidently also related to enhanced defocusing of the electrons. Here inelastic interactions of the electrons with the crystal play an important role, along with multiple elastic interactions. The point is that the focusing of reflected electrons, which causes them to escape manly along the atomic rows, leads to an increase in the electron density near the ion core of the crystal, and this should be accompanied by increased elastic, as well as inelastic, scattering of the electrons at large angles. As a result, the intensity of the flux of electrons moving toward the surface along the atomic rows falls off more rapidly than in other orientations and, in particular, in those for which minima of the $I(\varphi)$ curves are observed for quasielastic reflection. The influence of these processes increases as the path traversed by the electrons in the direction toward the surface becomes longer; this happens when the energy loss ΔE and angle θ increase. This factor also evidently explains the faster drop in the $\chi(\Delta E)$ curves seen upon an increase in the polar angle of escape of the electrons. The observed inversion in the structure of the spatial distributions of the electrons which have experienced large energy losses during reflection is of a similar nature.

Thus, an extension of the range of energy losses by the reflected electrons detected in medium energy electron diffraction greatly complicates the picture of the phenomena, since multiple elastic and inelastic scattering processes come into play which defocus the electrons. This, on one hand, makes it more difficult to analyze the patterns and model them numerically, and, on the other, should reduce the contrast in the measured distributions.

CONCLUSION

An experimental study of the focusing effect for electrons with energies of the order of 1 keV using an analysis of the spatial distributions of the electrons reflected from single-crystal Mo (100) with different energy losses ΔE has revealed the following behavior:

1. The effect shows up over a fairly wide range of electron energy losses up to roughly 200 eV. The size of this range depends on the polar angle of escape θ of the electrons.

2. For electrons reflected with small (single event) energy losses, the focusing efficiency is determined principally by the number of atoms n encountered by an electron along

its path as it moves toward the surface along a close-packed direction, and it increases with n. The spatial orientation of the exit directions and the packing density of the atoms in them are of less importance.

3. The focusing efficiency for electrons reflected with large energy losses declines, as a rule, with increasing ΔE, although the $\chi(\Delta E)$ curves are nonmonotonic and correlate with the energy loss spectrum. The weakening of the focusing is caused by multiple elastic and inelastic scattering of the electrons by the crystal (defocusing processes). The peaks of the distributions are damped most rapidly along the most close-packed directions of the crystal, for which an inversion of the diffraction patterns is observed at large ΔE and θ.

[1] S. A. Chambers, Surf. Sci. Rep. **16**, 261 (1992).

[2] M. Erbudak, M. Hochstrasser, and E. Wetli, Mod. Phys. Lett. B **8**, 1759 (1994).

[3] N. S. Faradzhev, M. V. Gomoyunova, and I. I. Pronin, Phys. Low-Dim. Struct. **314**, 93 (1997).

[4] S. Mroz, Prog. Surf. Sci. **48**, 157 (1995).

[5] S. Valeri and A. Di Boni, Surf. Rev. Lett. **4**, 141 (1997).

[6] A. Mosser, Ch. Burggraf, S. Goldsztaub, and Y. H. Ohtsuki, Surf. Sci. **54**, 580 (1976).

[7] M. V. Gomoyunova, I. I. Pronin, and S. L. Zaslavskiĭ, Fiz. Tverd. Tela (Leningrad) **24**, 2006 (1982) [Sov. Phys. Solid State **24**, 1145 (1982)].

[8] S. Hufner, J. Osterwalder, T. Greber, and L. Schlapbach, Phys. Rev. B **42**, 7350 (1990).

[9] M. Erbudak, M. Hochstrasser, and E. Wetli, Phys. Rev. B **50**, 12973 (1994).

[10] I. I. Proninn, M. V. Gomoyunova, D. P. Bernatskiĭ, and S. L. Zaslavskiĭ, Prib. Tekh. Éksp., **1**, 175–179 (1982).

[11] M. V. Gomoyunova, I. I. Pronin, and N. S. Faradzhev, Zh. Éksp. Teor. Fiz. **110**, 311 (1996) [JETP **83**, 168 (1996)].

[12] J. H. Weaver, D. W. Lynch, and C. Golson, Phys. Rev. B **10**, 501 (1974).

[13] M. V. Zharnikov, V. D. Gorobchenko, and I. L. Serpuchenko, Zh. Éksp. Teor. Fiz. **92**, 228 (1987) [Sov. Phys. JETP **65**, 128 (1987)].

[14] I. I. Pronin, M. V. Gomoyunova, and N. S. Faradzhev, Fiz. Tverd. Tela (St. Petersburg) **39**, 752 (1997) [Phys. Solid State **39**, 752 (1997)].

[15] M. V. Gomoyunova and I. I. Pronin, Zh. Tekh. Fiz. **67**(8), 117 (1997) [Tech. Phys. **42**, 961 (1997)].

[16] M. A. Vicente Alvares, H. Ascolani, and G. Zampieri, Phys. Rev. **53**, 7524 (1996).

[17] H. Raether, *Excitation of Plasmons and Interband Transitions by Electrons*. Springer Tracts in Modern Physics, Vol. 88. Springer-Verlag, Berlin, Heidelberg, New York (1980).

BRIEF COMMUNICATIONS

Mechanisms for formation of a population inversion in the levels of metal atoms and ions in a plasma jet

V. P. Starodub

Institute of Electron Physics, National Academy of Sciences of Ukraine, 294016 Uzhgorod, Ukraine
(Submitted May 5, 1997)
Zh. Tekh. Fiz. **68**, 134–135 (June 1998)

The mechanisms for formation of population inversions in plasma jets of lithium, sodium, cadmium, and strontium vapor are studied. The primary contribution to forming a population inversion over the transitions of the alkali atoms is found to be three-body electron–ion recombination, and for transitions between the ion levels of strontium and cadmium, inelastic collisions with the buffer gas play an important role. Using helium instead of argon as a buffer leads to a substantial increase in the magnitude of the inversion. © *1998 American Institute of Physics.* [S1063-7842(98)02406-4]

It is important to examine the conditions and analyze the mechanisms for population inversions in moving plasmas with rapidly varying density and temperature in order to ascertain the feasibility of plasmadynamic lasers. In this paper, a study is made of plasma jets formed from the vapors of lithium, sodium, cadmium, and strontium. The experiments were conducted on a device described elsewhere.[1] The source of the jet was a dc plasmatron with a 3-mm-diam acoustic nozzle. Argon and helium were used as buffer gases. The excited level populations were determined optically from the intensities of spectral lines, which in turn were determined by comparison with the intensity of a standard source which has a known spectral energy distribution.

These studies showed that population inversions develop between certain levels of lithium and sodium atoms and cadmium and strontium ions. These atoms and ions each have one electron in their outer shell, so they are similar. The data show, however, that the behaviors of the population inversions over these atoms and ions are not at all identical, either in terms of the initial conditions or along the jet.

Figure 1 shows plots of the magnitudes of the population inversions as functions of distance along the jet for the transitions of these elements which are the most promising from the standpoint of lasing. For lithium and sodium the maximum population inversion is observed near the exit section of the nozzle and falls off rapidly along the jet, following the variation in the electron density.[2,3] This indicates that processes involving electrons, in particular, three-body electron–ion recombination

$$A^+ + e + e = A^* + e, \tag{1}$$
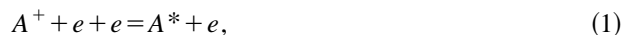
play an important role in creating the population inversion.

As opposed to the alkali metals, the peak inversion in the populations of the Sr^+ and Cd^+ ions is observed a significant distance away from the nozzle exit (Fig. 1, curves *3* and *4*). This is unambiguous evidence that in the cadmium and

strontium plasmas, reaction (1) does not make a significant contribution to the formation of a population inversion. More likely, a population inversion is created over the $6\,^2S_{1/2} \rightarrow 5\,^2P_{3/2}$ transition of strontium through processes involving the buffer gas He, as is indicated by the dependence of the magnitude of the inversion on the helium pressure (Fig. 2). Note that all the curves shown in Fig. 2 were obtained with a constant feed rate for the metals. An analysis of published data[4–7] and of the level diagrams for the strontium and helium atoms and the strontium ion shows that the following elementary processes may play an important role in populating these levels:
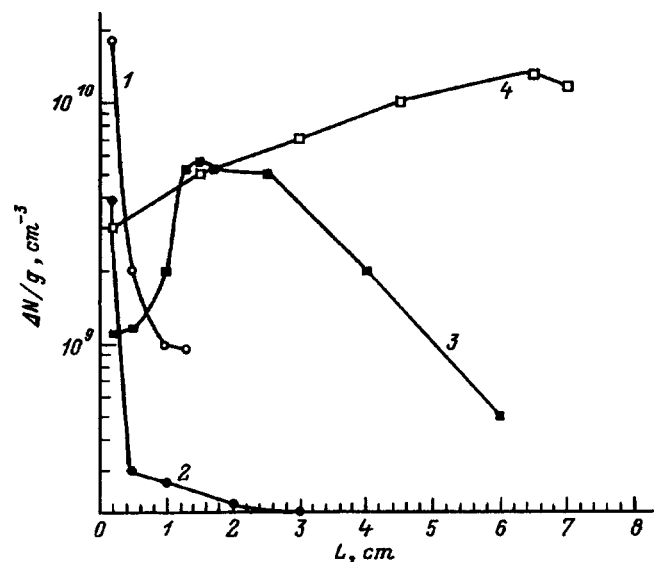

FIG. 1. The variations in the magnitudes of population inversions over transitions of Li I, Na I, Cd II, and Sr II along the jet: *1*—$3\,^2S_{1/2} \rightarrow 2\,^2P_{3/2}$ Li I, *2*—$5\,^2S_{1/2} \rightarrow 3\,^2P_{3/2}$ Na I, *3*—$6\,^2S_{1/2} \rightarrow 5\,^2P_{3/2}$ Sr II, and *4*—$5s\,^2D_{5/2} \rightarrow 5\,^2P_{3/2}$ Cd II.
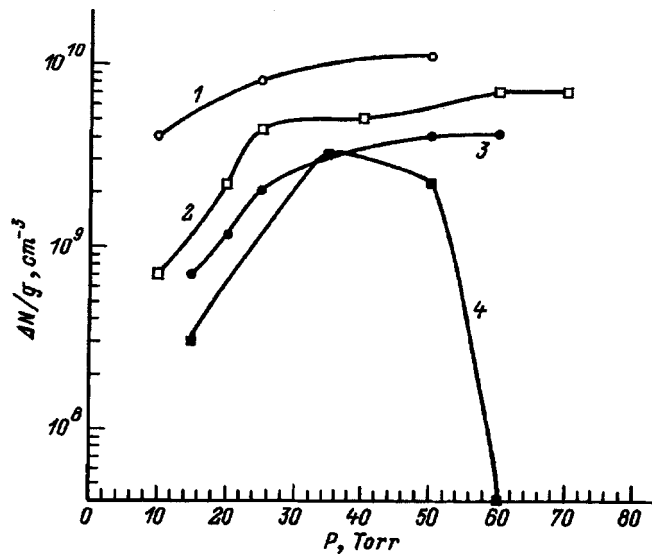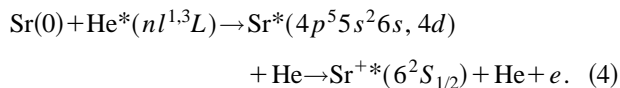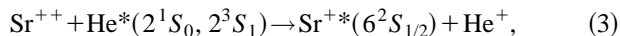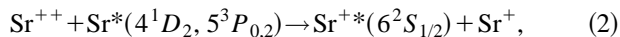
FIG. 2. The magnitudes of population inversions over transitions of Li I, Na I, Cd II, and Sr II as functions of the helium pressure: 1—$3\,^2S_{1/2} \rightarrow 2\,^2P_{3/2}$ Li I, 2—$5s^2\,^2D_{5/2} \rightarrow 5\,^2P_{3/2}$ Cd II, 3—$5\,^2S_{1/2} \rightarrow 3\,^2P_{3/2}$ Na I, and 4—$6\,^2S_{1/2} \rightarrow 5\,^2P_{3/2}$ Sr II.

$$Sr^{++} + Sr^*(4\,^1D_2, 5\,^3P_{0,2}) \rightarrow Sr^{+*}(6\,^2S_{1/2}) + Sr^+, \qquad (2)$$

$$Sr^{++} + He^*(2\,^1S_0, 2\,^3S_1) \rightarrow Sr^{+*}(6\,^2S_{1/2}) + He^+, \qquad (3)$$

$$Sr(0) + He^*(nl^{1,3}L) \rightarrow Sr^*(4p^55s^26s, 4d)$$
$$+ He \rightarrow Sr^{+*}(6\,^2S_{1/2}) + He + e. \quad (4)$$

Two of these reactions involve atoms of the buffer gas. The rate constants for reactions (2) and (3) are large and, according to the data of Ref. 5, are equal, respectively, to 1×10^{-8} cm^{-3} s^{-1} and 2×10^{-8} cm^{-3} s^{-1}. It was not possible to find the rate for reaction (4), but there is some indication[4] that above 19–25 eV there are some autoionizing levels of the strontium atom which can decay to form a strontium ion in the $6\,^2S_{1/2}$ state. The cross section for this kind of reaction for different levels of Sr* I is estimated to be in the range 10^{-14}–10^{-17} cm^2.

The inversion does not exist for low concentrations of helium (Fig. 2), since the helium concentration is too low for pumping the $6\,^2S_{1/2}$ level (reactions (3) and (4)). As the helium concentration is raised, a population inversion appears, reaches a maximum, and then falls off. The decrease in the magnitude of the inversion is explained by the fact that, as the helium concentration is raised, relaxation processes in the plasma are faster and the metastable states of strontium and helium are destroyed more rapidly.

In order to clarify the role of processes (1)–(4) we have performed an additional experiment. Some cesium, which is easily ionized, was added to the He–Sr plasma jet. The cesium density was 10^{13}–10^{14} cm^{-3} at most. In this experiment, the cesium impurity enhanced the electron density by roughly a factor of two, without affecting the electron temperature significantly. If the $6\,^2S_{1/2}$ level is populated by recombination (reaction (1)), then the intensity of the corresponding line should have increased. If processes (2)–(4) predominate in filling the level, then the intensities of the 430.5 and 416.2 nm lines should have decreased, since in the first case $\Delta N \propto N_e$ and in the second case $\Delta N \propto 1/N_e$. Introducing a cesium impurity into the He–Sr plasma jet did not increase the intensities of these lines but decreased them by $\sim 20\%$. It can thus be inferred that reactions (2)–(4) make a significant contribution to filling the $6\,^2S_{1/2}$ state of Sr II. To clarify the role of the buffer gas in creating a population inversion over the cadmium ion lines, helium was replaced by argon. It was found that in an Ar–Cd plasma the inversion is 6–8 times smaller than in the He–Cd plasma jet. This is explained by the fact that the cross section for the endothermic charge exchange on the Cd$^+$ level is roughly an order of magnitude smaller than the corresponding cross section for the endothermic charge exchange in He–Cd. In addition, a Penning reaction takes place in the He–Cd plasma, but not in argon.

Based on data obtained from a study of metal vapor plasma jets, therefore, it has been found that three-body electron–ion recombination predominates in the formation of population inversions over transitions of the alkali atoms, while inelastic collisions with the buffer gas play a significant role for transitions between levels of the strontium and cadmium ions, and if helium is used instead of argon as a buffer gas, there is a substantial increase in the magnitude of the inversion.

[1] S. P. Bogacheva, M. F. Veresh, I. P. Zapesochnyĭ *et al.*, Ukr. Fiz. Zh. **30**, 186 (1995).

[2] S. P. Bogacheva, L. V. Voronyuk, I. P. Zapesochnyĭ *et al.*, Zh. Prikl. Mekh. Tekh. Fiz., No. 6, 10 (1984).

[3] M. F. Veresh, I. P. Zapesochnyĭ, and V. P. Starodub, Zh. Tekh. Fiz. **57**, 572 (1987) [Sov. Phys. Tech. Phys. **32**, 347 (1987)].

[4] A. A. Borovik, I. S. Aleksakhin, V. F. Bratsov, and A. B. Kuplyauskene, Opt. Spektrosk. **53**, 976 (1982) [Opt. Spectrosc. (USSR) **53**, 583 (1982)].

[5] B. M. Smirnov, *Excited Atoms* [in Russian], Énergoizdat, Moscow (1982), 232 pp.

[6] V. N. Kondrat'ev and E. E. Nikitin, *Kinetics and Mechanism of Gas-Phase Reactions* [in Russian], Nauka, Moscow (1975), 397 pp.

[7] B. M. Smirnov, *Asymptotic Methods in Collision Theory* [in Russian], Atomizdat, Moscow (1973), 294 pp.

Translated by D. H. McNeill

Mathematical modeling of the nonstationary operation of thermoelectric current sources

Yu. I. Dudarev and M. Z. Maksimov

Sukhumi Physicotechnical Institute, Academy of Sciences of the Abkhaz Republic, Sukhumi, Abkhazia
(Submitted April 14, 1997)
Zh. Tekh. Fiz. **68**, 136–137 (June 1998)

Thermoelectric energy conversion based on the Seebeck, Thomson, and Peltier effects has been used extensively for several decades in various areas of science, technology, and low-level power generation. In refrigeration technology, as well as in thermoelectric generators, stationary operating regimes have traditionally been used. Studies of nonstationary cooling and current generation began only much later (Stil'bans and Fedorovich,[1] E. K. Iordanishvili and co-workers,[2] and others) and certain advantages of these operating regimes were identified at once, especially in the initial stage of operation, when the difference in the time constants of the thermal and electrical processes are important.

Thermoelectric power supplies are multilayer structures in which the temperature distributions are determined by solving a system of nonlinear differential heat conduction equations with the corresponding boundary conditions for the heat sources and sinks. In their general form in the one dimensional approximation, these equations are

$$\frac{\partial}{\partial x}\left[\varkappa_i(T)\,\frac{\partial T_i}{\partial x}\right]+j_i(t)\,\frac{\partial \alpha_i(T)}{\partial T_i}\,\frac{\partial T_i T_i}{\partial x}+j_i^2(t)\rho_i(T)$$

$$=C_i(T)\,\frac{\partial T_i}{\partial t},$$

$$\frac{\partial T_1}{\partial x}-\frac{Q_1(t)}{\varkappa_0(0)}+b_1(T-T_1(t))=f_1(t,T),\quad x=0,$$

$$\frac{\partial T_L}{\partial x}-\frac{Q_2(t)}{\varkappa_0(L)}+b_2(T-T_2(t))=f_2(t,T),\quad x=L,$$

$$T(x,0)=T_H(x),\tag{1}$$

where j is the current density, $\varkappa$ and α are the thermal conductivity and thermopower (Seebeck) coefficient, ρ is the electrical resistivity, C is the specific heat, and i is the layer number.

Usually the thermal characteristics of the material in each layer can be represented in the form

$$\sigma(x,T)=\sigma_0(x)+\Delta\sigma(x,T),\quad \frac{\Delta\sigma}{\sigma_0}<1.\tag{2}$$

The system of Eqs. (1) must be supplemented by equations for determining the current density j, which depend on the electrical circuit joining the thermal elements. In accordance with the method outlined previously,[3] the solution of this problem is conveniently represented in the form

$$T(x,t)=T_0(x,t)+U(x,t),\tag{3}$$

where $T_0(x,t)$ is the temperature of the system in neglect of the dependence of the thermal properties and thermal sources on the temperature and current effects.

Going from a multilayer system to an effective homogeneous system using the WKB method,[4] i.e., making the substitution

$$x\to\Pi(x)=\int_0^x\frac{dx}{\sqrt{a_0(x)}},\quad a_0=\frac{\varkappa_0}{C_0},\tag{4}$$

and using a Green function formalism,[4] we obtain

$$T_0(x,t)=\int_0^{\Pi(L)}T_H(\Pi')G(\Pi,t;\Pi',0)d\Pi'$$

$$-\int_0^{t^+}\left\{\left(\left(\frac{Q_1\sqrt{a_0(0)}}{\varkappa_0(0)}\right.\right.\right.$$

$$\left.+b_1\sqrt{a_0(0)}T_1(\tau)\right)G(\Pi,t;0,\tau)-\left(\frac{Q_2\sqrt{a_0(L)}}{\varkappa_0(L)}\right.$$

$$\left.\left.+b_2\sqrt{a_0(L)}T_2(\tau)\right)G(\Pi,t;\Pi(L),\tau)\right\}d\tau.\tag{5}$$

To determine $U(x,t)$ we have the integral equation

$$U(x,t)=-\int_0^{t^+}\left\{\int_0^{\Pi(L)}\Psi(\Pi',\tau)G(\Pi,t;\Pi',\tau)d\Pi'\right.$$

$$+\sqrt{a_0(0)}f_1 G(\Pi,t;0,\tau)$$

$$\left.-\sqrt{a_0(L)}f_2 G(\Pi,t;\Pi(L),\tau)\right\}d\tau,\tag{6}$$

where

$$\Psi=-\frac{\partial}{\partial x}\left[\Delta\varkappa(x,T)\,\frac{\partial T}{\partial x}\right]+\Delta C(x,T)\,\frac{\partial T}{\partial t}$$

$$-j(t)\,\frac{\partial\alpha}{\partial T}\,\frac{\partial T}{\partial x}\,T-j^2(t)\rho(T),$$

$$f_k=-\frac{\Delta\varkappa_k}{\varkappa_{0k}}\,\frac{\partial T}{\partial x}+\omega_k(t,T),\quad k=1,2,$$

Q_k and ω_k are known functions; the upper limit t^+ means that the integration with respect to τ is taken to $t+\varepsilon$ followed by a transition to the limit $\varepsilon\to 0$; $G(\Pi(x),t;\Pi(x'),\tau)$ is the Green function for the system of equations (1), which is found approximately using the WKB method as

$$G(\Pi,t;\Pi',\tau) = \frac{1}{2\pi i} \int_{\delta-i\infty}^{\delta+i\infty} N(\Pi,\Pi',P)$$

$$\times \exp[p(t-\tau)]dp, \qquad (7)$$

where

$$N \simeq \frac{\left[\cosh\sqrt{p}\,\Pi' - \dfrac{b_1\sqrt{a_0(0)}}{\sqrt{p}}\sinh\sqrt{p}\,\Pi'\right]}{R}$$

$$\times \left\{ \frac{b_2\sqrt{a_0(L)}}{\sqrt{p}} \sinh\sqrt{p}\,[\Pi(L)-\Pi] \right.$$

$$\left. + \cosh\sqrt{p}\,[\Pi(L)-\Pi] \right\},$$

$$R = \sqrt{p}\left[\left(\frac{b_2\sqrt{a_0(L)}}{\sqrt{p}} - \frac{b_1\sqrt{a_0(0)}}{\sqrt{p}} \right) \cosh\sqrt{p}\,\Pi(L) \right.$$

$$\left. + \left(1 - \frac{b_1 b_2 \sqrt{a_0(0)a_0(L)}}{\sqrt{p}} \right) \sinh\sqrt{p}\,\Pi(L) \right], \qquad (8)$$

and $\Pi > \Pi'$.

For $\Pi < \Pi'$, Π and Π' must be switched. Equation (6) is easily solved by successive approximations, with $T_0(x,t)$ taken as a zeroth approximation.

It should be noted that using asymptotic methods,[5] in particular the reduced matching technique,[6] makes it possible to simplify the computational formulas greatly. For short times

$$\frac{t}{\Pi_0^2(L)} < \frac{1}{4}$$

and low current densities

$$j < 0.1 \frac{\varkappa}{\alpha L},$$

the influence of current effects on the temperature can be neglected, and this substantially shortens the computational procedure. The effective characteristics for multilayer systems can be obtained using the sum rule for the eigenvalues of the original problems.[7]

Figure 1 shows the time variation in the relative electrical power of a thermoelectric generator, W/W_T^{max}, obtained through calculations employing the method proposed here (curve *1*), an exact solution (curve *3*), and experimentally (curve *2*). The experimental data were obtained by A. A. Sokolov (Sukhumi Physicotechnical Institute). A comparison
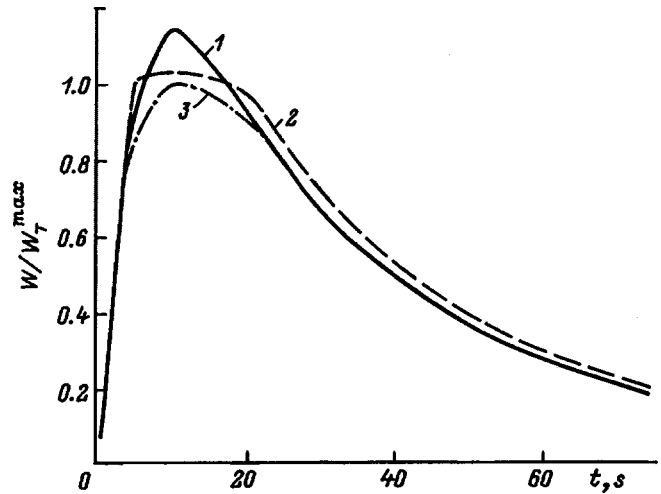

FIG. 1. W/W_t^{max} as a function of time: *1*—calculated by the method proposed here, *2*—experiment, *3*—exact solution.

of these results shows that their are in satisfactory agreement, so the method developed here can be recommended for calculating and analyzing the characteristics of specific thermoelectric devices.

Mathematical modeling of the nonstationary operation of thermoelectric generators shows that, in order to accelerate the approach of a thermoelectric generator to a given power production regime, it is necessary to reduce the thickness of the transition layers on the hot-junction side and to use materials with a high thermal diffusivity in them. In order to sustain prolonged operation of the system, its heat capacity must be increased and the temperature or calorific capacity of the heat source must be raised. The amplitude of the output electrical signal also depends significantly on these factors.

[1] L. S. Stil'bans and N. A. Fedorovich, Zh. Tekh. Fiz. **28**, 489 (1958) [*sic*].
[2] E. K. Iordanishvili and V. P. Babin, *Nonstationary Processes in Thermoelectric and Thermomagnetic Energy Conversion Systems* [in Russian], Nauka, Moscow (1983), 216 pp.
[3] Yu. I. Dudarev and M. Z. Maksimov, Teplofiz. Vys. Temp. **26**, 824 (1988).
[4] F. M. Morse and H. Feshbach, *Methods of Mathematical Physics*, McGraw-Hill, New York (1953); IL, Moscow (1958), Vols. 1 and 2.
[5] Yu. I. Dudarev, A. P. Kashin, V. I. Lozbin, and O. V. Marchenko, Inzh. Fiz. Zh. **42**, 492 (1982).
[6] A. P. Kashin, T. M. Kvaratskheliya, M. Z. Maksimov, and Z. E. Chikovani, Teor. Mat. Fiz. **78**, 392 (1989).
[7] Yu. I. Dudarev, A. P. Kashin, and M. Z. Maksimov, Inzh. Fiz. Zh. **48**, 333 (1985).

Translated by D. H. McNeill

On the problem of estimating the relaxation time of local deformations on metal surfaces

Yu. I. Dudarev, A. V. Kazakov, and M. Z. Maksimov

Sukhumi Physicotechnical Institute, Academy of Sciences of the Abkhaz Republic, Sukhumi, Abkhazia
(Submitted April 28, 1997)
Zh. Tekh. Fiz. **68**, 138–139 (June 1998)

Asymptotic methods for evaluating the change in the relief of local surface formations are used
to obtain simple and fairly exact equations for the profile of the relief and characteristic
relaxation times of local deformations on metal surfaces in kinetic and diffusion models. © *1998
American Institute of Physics.* [S1063-7842(98)02606-3]

1. In creating ultrahigh-capacity information storage systems based on solids, information is coded through the action of the tip of a scanning tunneling microscope made of a refractory metal. The resulting changes in the relief (prominences and indentations) can be associated with bits of information.[1] Self-diffusion of the material of the storage medium can, however, cause a gradual smoothing of the relief and reduce the bit lifetime. Studies of the relaxation kinetics and dynamics of surface structures are therefore of some interest for predicting the operating lifetime of this type of device. This question is the subject of a detailed paper.[1] We believe, however, that the solutions of the equations obtained there for the function $f(r,t)$ which describes the surface relief were not examined analytically to a sufficient extent. We shall discuss this in more detail. First of all, these equations can be written in a unified form

$$\frac{\partial f(r,t)}{\partial t} = -K_n(-1)^n L^{n-1} \frac{1}{r} \frac{\partial}{\partial r}\left(\frac{rf_r}{\sqrt{1+f_r^2}}\right),$$

$$L = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right), \quad f_r = \frac{\partial f}{\partial r}, \tag{1}$$

where L is the radial Laplacian operator on the plane, $n=1$ corresponds to the kinetic model ($K_1 \equiv K$) and $n=2$, to the diffusion model ($K_2 \equiv \lambda$).

Here it is assumed that the initial profile of the depressions on the surface of the metal are described by a Gaussian distribution

$$f(r,0) \equiv -b_0 f_0(r) = -b_0 \exp\left(-\frac{\pi b_0 r^2}{V_0}\right), \tag{2}$$

where b_0 is the initial depth of the depression and V_0 is its volume with a root-mean-square radius of the equivalent cylinder of

$$\langle r^2 \rangle = R^2 = \frac{V_0}{\pi b_0} = b_0 \mu^2, \quad \mu^2 = \frac{V_0}{\pi b^3}. \tag{3}$$

This representation of the initial conditions for the profile is entirely natural, since it conforms with the overall distribution of roughness on the surface.[2-4] In the general case, the solution of the system of Eqs. (1) and (2) is difficult; however, the horizontal dimensions $2R$ of the pits which develop on metallic surfaces owing to the interaction with the tip of a tunneling microscope are much larger[1,3] than their vertical dimensions, similarly to natural surface asperities.[2-4] In this case,

$$|f_r| \sim \frac{r}{R}\frac{b_0}{R} < 1,$$

so that Eq. (1) can be linearized to take the compact form

$$\frac{\partial f}{\partial t} = K_n(-1)^n L^n f. \tag{4}$$

2. Here it should be pointed out that describing a roughness profile by the system of equations (2) and (4) is fully equivalent to the problem of the time evolution of the distribution function $f_0(r)$ in studies of the kinetics of fine pulverization and other technological processes for working ores and raw materials.[5,6] This makes the analysis of the corresponding solutions much easier. In fact, according to Refs. 5 and 6, the solution of the system of Eqs. (2) and (3) can be written in the following integral form:

$$f_n(r,t) = -b_0 \int_0^\infty f_0(r_0) G_n(\mathbf{r},t|\mathbf{r}_0,0)\,d\mathbf{r}_0, \tag{5}$$

where $G_n(\mathbf{r},t|\mathbf{r}_0,0)$ is the Green function of the operator (4), whose Fourier representation has the form

$$G_n(\mathbf{r},t|\mathbf{r}_0,t) = \frac{1}{(2\pi)^2}\int d\mathbf{q}\,\exp[i\mathbf{q}(\mathbf{r}-\mathbf{r}_0) - K_n t q^{2n}]. \tag{6}$$

It is easy to confirm that, as $t\to 0$,

$$G_n(\mathbf{r},t|\mathbf{r}_0,0) = \delta(\mathbf{r}-\mathbf{r}_0), \tag{7}$$

and on the basis of Eq. (5) we immediately obtain the result (2) for any $f_0(r)$. Furthermore, since the main contribution to the integral (6) is from finite $q \sim q_0$, by making the substitution $K_n q^{2n} t = \rho$ it is easy to obtain a second leading asymptotic term of G_n as $t\to\infty$

$$G_n(\mathbf{r},t|\mathbf{r},0) = \frac{1}{2}\Gamma\left(1+\frac{1}{n}\right)(K_n t)^{-1/n}, \tag{8}$$

where $\Gamma(z)$ is the gamma function.

Substituting Eq. (8) in Eq. (5) yields

$$f(r,t)_{t\to\infty} = -\frac{V_0}{4\pi}\,\Gamma\left(1+\frac{1}{n}\right)(K_n t)^{-1/n}, \qquad (9)$$

since

$$2\pi b_0 \int_0^\infty f(r_0) r_0 \, dr_0 = V_0. \qquad (10)$$

In the case of the kinetic model ($n=1$, $K_1=K$) and the Gaussian initial condition (2), all the calculations become simpler,[1] since the corresponding integrals of the Bessel functions in Eqs. (5) and (6) are complete Weber integrals.[7] We have

$$f_1(r,t) = -b_1(t)\exp(-\pi b_1(t)r^2/V_0), \qquad (11)$$

so that

$$f_1(0,t) = -b_1(t) = -b_0(1+4\pi K b_0 t/V_0)^{-1}. \qquad (12)$$

For arbitrary n and, in particular, for the diffusion model ($n=2$, $K_2=\lambda$), it is not possible to obtain such simple equations for $f_n(r,t)$. However, for approximate estimates we can use the reduced matching method[8] in the parameter t. According to the procedure of this method, for the leading asymptotic terms of the function $[-f(r,t)]^n$ we have from Eqs. (2) and (9)

$$\left(-\frac{f(r,t)}{b_0}\right)^n = \begin{cases} [f_0(r)]^n & t\to 0 \\ \tau_n/t & t\to\infty, \end{cases} \qquad (13)$$

which, when matched, give

$$f_n(r,t) = -b_0(t/\tau_n + 1/f_0^n(r))^{-1/n}, \qquad (14)$$

and

$$\tau_n K_n = \left(\frac{V_0\Gamma(1+1/n)}{4\pi b_0}\right)^n. \qquad (15)$$

From this, first of all for the kinetic model ($n=1$) and the initial condition (2) we obtain

$$f_1(r,t) \approx -b_0(\exp(r^2/R^2)+t/\tau_1), \quad \tau_1=R^2/4K, \qquad (16)$$

where Eq. (3) has been used.

Here, on the one hand, the function (16) satisfies the initial conditions, and its series expansion and that of expression (11), in powers of r^2/R^2 give the same values for the first few coefficients, and, on the other hand, for $b_1(t) = f_1(0,t)$ we obtain the correct result, Eqs. (11) and (12). Thus Eqs. (14) and (15) can be used with sufficient accuracy for other r and n. If here we introduce the degree of preservation of a dip in the surface, $\xi = -f(r,t)/b_0 < f_0(r)$, then for the time t for filling it from the initial $f(r)$ to a given level ξ, we find

$$t_n = \tau_n(\xi^{-n} - f_0^{-n}(r)), \quad \xi \leqslant f_0(r). \qquad (17)$$

Here it should be noted that none of the above estimates are very sensitive to the shape of the initial distribution $f_0(r)$, because, as in Ref. 1, it is assumed that condition (10), normalization to volume V_0 for a given depth $f_0=b_0$, and condition (3), for the average radius, are satisfied. All together, this makes it possible to choose an initial distribution of the simplest form, close to rectangular, for which the transition

from Eq. (1) to Eq. (4) becomes better justified. This is confirmed by numerical calculations[1] of two approximations for t_1 in the kinetic model. In these cases, we can also assume $f_0 \simeq 1$; then Eq. (17) for the kinetic ($n=1$) and diffusion ($n=2$) models becomes

$$t_{\text{kin}} \simeq \tau_1(1/\xi-1), \quad \tau_1=R^2/4K; \qquad (18)$$

$$t_{\text{dif}} \simeq \tau_2(1/\xi^2-1), \quad \tau_2=\pi R^4/64\lambda. \qquad (19)$$

Their ratio is

$$t_{\text{dif}}/t_{\text{kin}} = (1/\xi+1)\tau_2/\tau_1, \qquad (20)$$

i.e., for other conditions the same, it depends on the degree of filling. For example, for $\xi=0.5$ we have

$$t_{\text{dif}}/t_{\text{kin}} = 3\tau_2/\tau_1. \qquad (21)$$

Next, for $\xi\to 0$, Eq. (19) for the diffusion model gives the correct result $\lambda t_{\text{dif}}/\pi b_0^4\mu^4 = 1/64\xi^2$, as opposed to the estimates of Ref. 1, which gave $1/96\xi$.

3. By using a rigorous analysis of the initial kinetic equations and asymptotic methods for the change in the relief of local surface formations, we have obtained simple and fairly accurate equations for the relief profile and for the characteristic relaxation times of local deformations on metal surfaces in the kinetic and diffusion models. All of this, together with methods for calculating the energy parameters and transport coefficients,[1] makes it much easier to predict the operating lifetime of memory structures, to determine the conditions for their reliable operation, to choose the storage-medium material, and other tasks which are indispensable steps in nanotechnologies.

[1] A. M. Dobrotvorskiĭ and V. K. Adamchuk, Zh. Tekh. Fiz. **64**(8), 132 (1994) [Tech. Phys. **39**, 816 (1994)].

[2] V. I. Trofimov and V. A. Osadchenko, Poverkhnost', No. 9, 5 (1987).

[3] V. I. Vettegren', S. Sh. Rakhimov, and V. N. Svetlov, Fiz. Tverd. Tela (St. Petersburg) **37**, 913 (1995) [Phys. Solid State **37**, 495 (1995)].

[4] V. V. Boĭko, A. P. Kashin, M. Z. Maksimov, and Z. E. Chikovani, Poverkhnost', No. 11, 40 (1993).

[5] E. A. Nepomnyashchiĭ, in *Theoretical Foundations of Chemical Technology* [in Russian], Vol. 7, No. 5, p. 754 (1973); *ibid.*, Vol. 11, No. 3, p. 477 (1977); *ibid.*, Vol. 12, No. 4, p. 576.

[6] A. P. Kashin, M. Z. Maksimov, and Z. E. Chikovani, Ukr. Fiz. Zh. **36**, 973 (1991).

[7] M. M. Agrest and M. Z. Maksimov, *Theory of Incomplete Bessel Functions and their Applications*, Springer-Verlag, Berlin (1971).

[8] A. P. Kashin, T. M. Kvaratskheliya, M. Z. Maksimov, and Z. E. Chikovani, Teor. Mat. Fiz. **78**, 392 (1989).

Translated by D. H. McNeill

Possible design for a cryogenic ferromagnetic gyroscope

L. A. Levin and S. L. Levin

Three of the simplest schemes for the design of a new physical apparatus based on the Barnett effect are examined. It is shown that the most realistic is a design with a special type of electrically open-circuited, superconducting magnetic shield. The design of this device is described. © *1998 American Institute of Physics.* [S1063-7842(98)02706-8]

The possibility, in principle, of creating a new type of cryogenic ferromagnetic gyroscope, an angular velocity sensor, has been discussed.[1] The cryogenic ferromagnetic gyroscope is based on the Barnett effect,[2] in which a magnetic field is generated when a ferromagnetic body is rotated. It has been shown[1] that there are three physical phenomena for exciting a magnetic field when different bodies are rotated. These are the magnetoresonance effect in ^{3}He, the London moment in superconductors, and the Barnett effect. In order to obtain a magnetic induction in the cryogenic ferromagnetic gyroscope equal to the induction produced in the nuclear gyroscope with ^{3}He developed so long ago,[3] it would be necessary for the relative dimensionless magnetic permeability of the ferromagnetic body to be $\geqslant 800$. It was shown that the cryogenic ferromagnetic gyroscope should be much simpler than the nuclear gyroscope.

In this paper we examine some design versions of a cryogenic ferromagnetic gyroscope. It was pointed out previously[1] that the simplest scheme for a cryogenic ferromagnetic gyroscope is a ferromagnetic rod located inside a cylindrical superconducting, electrically open-circuited shield. A superconducting sensor winding is placed on the rod and connected to a superconducting short-circuiting loop incorporating the input winding of a SQUID. This scheme, however, has two disadvantages. First, a simple cylindrical open-circuited magnetic shield gives a low shielding coefficient and, second, a winding placed directly on the ferromagnetic rod has a high inductance, which reduces the current in the input winding of the SQUID and makes it difficult to match the winding to the input of the SQUID. In principle, it is possible to mount the ferromagnetic rod in a cylindrical, closed-circuited superconducting shield, but in this case the overall size of the superconducting shield becomes much larger owing to the compensating magnetic field which develops inside the shield, or the dimensions of the rod must be reduced, which results in a lower sensitivity for the device. In sum, the two versions with simple, open- or closed-circuited shields are poorly suited to making a realistic cryogenic ferromagnetic gyroscope.

In order to create a functional cryogenic ferromagnetic gyroscope we propose using a special electrically open-circuited superconducting shield. It is an essentially cylindrical shield, slit along a generator of the cylinder, with a large overlapping zone at the site of the slit (for example, in the arrangement[4] of a ''snake swallowing its tail''). If the ratio of the height of the gap to the length of the gap overlap zone is kept at 1:10, then the shielding coefficient will be of the order of $10^5 - 10^6$. A further reduction in this ratio will increase the shielding coefficient. This scheme has been used by us in the device to be discussed below.[5]

An example of such a shield is a cylinder with a very narrow slit, of the order of 0.01–0.1 mm, along a generator and with a large wall thickness of 2–3 mm. Besides increasing the shielding coefficient, a shield of this sort makes it possible to shield the sensor winding placed on it from the ferromagnetic material and to reduce its inductance by roughly a factor of μ.

The ferromagnetic rod is mounted inside the shield. The diameter of the rod and the inner diameter of the shield must be close in value ($D_c \approx D_e$), so that a field close to the Barnett induction B_B acts on the inner surface of the shield. The current created on the inner side of the shield to compensate the field B_B flows over the outer surface of the shield, transferring the field B_B to the surface of the shield. This field excites a current in the superconducting sensor winding, which is wound on the surface of the shield and connected to a short-circuiting superconducting loop incorporating the input winding of a SQUID. In order to prevent current leakage, the width of the shield must be somewhat less than the length of the ferromagnetic rod.

The current excited in the sensor winding, which is proportional to the rotational speed of the apparatus, flows along the input winding of the SQUID, creating a flux which is measured by the SQUID. An electrically open-circuited shield can obviously be used in place of a sensor winding.

The magnetic flux created in the ferromagnetic rod by the Barnett effect is

$$\Phi_B = B_B S = \frac{\mu \Omega}{\gamma_B} S, \tag{1}$$

where μ is the relative dimensionless magnetic permeability of the rod material, Ω is the angular velocity of rotation, S is the transverse cross section of the rod, and γ_B is the gyromagnetic ratio ($\gamma_B = 1.7 \times 10^{11}$ A·s·kg^{-1}).

When $D_c = D_e$, the flux at the shield is

$$\Phi_e = l_e L_e = \Phi_B = \frac{\mu \Omega}{\gamma_B} S, \tag{2}$$

where I_e and L_e are the current and inductance of the shield.

The magnetic flux at the input winding of the SQUID is

$$\Phi_c = I_0 L_c = \frac{\mu \Omega S}{\gamma_B} \frac{L_c}{L_0 + L_c}, \tag{3}$$

where I_0 is the current in the short-circuited superconducting circuit flowing through the input winding of a SQUID with inductance L_c, and L_0 is the inductance of the sensor winding.

We refer to $L_c/(L_0 + L_c) = K$ as the flux transfer coefficient. In complicated circuits this coefficient is more cumbersome than in Eq. (3), but, as a rule, it consists of relationships among the inductances and mutual inductances of the components in the device. In practice, $K \sim 10^{-1} - 10^{-3}$. The Barnett magnetic flux can be measured if $\Phi_c \geq \Phi_{nc}$, where Φ_{nc} is the SQUID noise. Then

$$\Phi_c = \Phi_{nc} = \frac{\mu \cdot \Omega \cdot S \cdot K}{\gamma_B}, \tag{4}$$

which implies that the sensitivity of the cryogenic ferromagnetic gyroscope to angular velocity will be

$$f = \frac{\Phi_{nc} \gamma_B}{2\pi \cdot \mu \cdot S \cdot K}. \tag{5}$$

Equation (5) implies that in order to raise the sensitivity of the cryogenic ferromagnetic gyroscope, it is necessary to increase μ, S, and K. For example, for $\Phi_{nc} = 10^{-5}\Phi_0$ ($\Phi_0 = 2 \times 10^{-15}$ Wb is the quantum of magnetic flux), $\mu = 800$, $S = 10^{-3}$ m^2, and $K = 10^{-2}$, we obtain for f a value of the order of 10^{-7} s^{-1}.

A simplified sketch of the cryogenic ferromagnetic gyroscope is shown in Fig. 1. Inside the case *1* are a ferromagnetic rod *2* and a superconducting magnetic shield *3*, both rigidly connected to the case; possible shapes of the shield are illustrated to the right and left in the figure. A superconducting sensor winding *4* is wound on the shield *2* and connected to the input winding *5* of a squid *6*. Superconducting end-cap shields *7* are mounted on the ends to improve the screening coefficient. The ratio of the height of the gaps to the overlap lengths of the shields *3* and *7* is 1:10.

It is known that during rotation of a superconducting body, such as a solid or hollow closed cylinder, a magnetic field (the London moment) develops in it.[1] A London moment does not develop in the electrically open-circuited

shield *3*, To eliminate the influence of the London moment on the measurement of the Barnett field, it is necessary that the end-cap shields *7* be slit appropriately. Here the ratio of the gap height to the overlap zone must be $\leq 1:10$, i.e., the cap must be comparatively thick and the slit very narrow.

There are still some details involved in making and using a cryogenic ferromagnetic gyroscope. Cooling the apparatus to the working temperature to eliminate trapping of the magnetic flux by superconducting components should be done under conditions of essentially zero magnetic field, especially near the superconducting transition temperature of the material.

It is desirable to use a single crystal for the ferromagnetic rod material and have the easy axis of magnetization coincide with the axis of the rod.

Ferromagnetic materials are noisy even at extremely low temperatures. In order to reduce the level of magnetic noise it is appropriate to coat the surface of the ferromagnetic rod with a nonmagnetic material such as pure copper that has a low Ohmic resistance at the operating temperature.

The noise level of SQUIDs increases sharply below a certain frequency $f_n \approx 0.1 - 1.0$ Hz. To eliminate this effect it is appropriate to modulate a higher frequency signal and demodulate at the output of the SQUID.

If an additional winding (or windings) is placed on the open shield, then the prospects for the cryogenic ferromagnetic gyroscope are greatly expanded; with a suitable control system it can measure angular acceleration and rotation angle.

We have proposed a design for a new physical apparatus, the cryogenic ferromagnetic gyroscope. The sensitive element of the cryogenic ferromagnetic gyroscope (a rod with a shield) has no moving parts and does not require any additional energy to operate; it does not release heat. Thus, it is very convenient for cryogenic devices and its operating lifetime is unlimited.

A comparison with the known designs for cryogenic gyroscopes—angular velocity sensors (the cryogenic nuclear gyroscope)[3] and the cryogenic ferromagnetic gyroscope—shows that the cryogenic ferromagnetic gyroscope design is simpler and more reliable. In terms of sensitivity, the cryogenic nuclear gyroscope and the cryogenic ferromagnetic gyroscope are similar. There is a possibility, in principle, of increasing the sensitivity of the cryogenic ferromagnetic gyroscope by employing a ferromagnetic material with a higher μ.
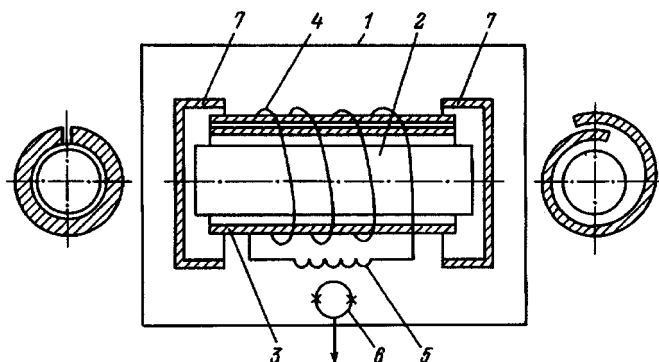

FIG. 1. A design sketch of the cryogenic ferromagnetic gyroscope.

[1] L. A. Levin, Zh. Tekh. Fiz. **66**(4), 192 (1996) [Tech. Phys. **41**, 399 (1996)].

[2] S. J. Barnett, Rev. Mod. Phys. **7**, 129 (1935).

[3] K. F. Woodman, P. W. Franks, and M. D. Richards, Rev. J. Navigation **40**, 336 (1987).

[4] D. B. Sullivan and R. F. Dziuba, Rev. Sci. Instrum. **45**, 517 (1974).

[5] A. P. Buravlev, B. E. Landau, L. A. Levin, and S. L. Levin, ''Cryogenic ferromagnetic gyroscope'' [in Russian], Inventor's Certificate No. 2084825; Publ. Byull. Izobret., No. 20 (1997).

Translated by D. H. McNeill