=========== **THEORETICAL AND MATHEMATICAL PHYSICS** ===========

Rigorous and Approximate Methods for Modeling Wave Scattering from a Locally Perturbed Perfectly Conducting Surface

V. V. Zalipaev[1] and A. V. Kostin[2]

[1] *Steklov Institute of Mathematics, St. Petersburg Division, Russian Academy of Sciences,*
St. Petersburg, 191011 Russia
[2] *St. Petersburg State Institute of Precise Mechanics and Optics (Technical University),*
St. Petersburg, 197101 Russia

Abstract—Rigorous and approximate methods are considered for solving the problem of harmonic plane wave scattering from a plane surface arbitrarily perturbed along one dimension on a finite interval. This problem is treated using the Fredholm integral equations of the second kind and the Kirchhoff and Rayleigh approximations. The estimates of the computational efficiency of the integral equation method and the Rayleigh approximation are compared by calculating fields scattered from random rough surfaces in the resonance region (i.e., when the roughness height is comparable to or smaller than the incident wavelength) for an arbitrary incidence of a plane wave. Scattering patterns calculated using the integral equations and the Kirchhoff approximation are discussed in the case of large-scale random rough surface scattering. Particular attention is paid to scattering at near-grazing incidence. © 2000 MAIK "Nauka/Interperiodica".

INTRODUCTION

Electromagnetic and acoustic scattering from random rough surfaces has been extensively investigated (e.g., see references in [1–4]). The Kirchhoff and Rayleigh approximations and the rigorous approach based on the integral equation method are the best known methods for simulating wave scattering from a large-scale (the roughness height is large in terms of wavelength) or resonant (the roughness height is comparable to or smaller than the wavelength) rough surface. It is known that the Kirchhoff approximation is efficient for simulating scattering from large (in terms of the incident radiation wavelength) roughness asperities at small incidence angles. However, at near-grazing incidence, which is typical for many radar problems, the Kirchhoff approximation cannot be applied because of the surface shadowing and multiple scattering from the surface elements. A numerical method based on the Rayleigh approximation provides exact data for the fields scattered from surfaces with low slopes in the resonance region, i.e., when the incident wavelength and the roughness dimensions have the same order of magnitude. The integral equation method, whose applicability is not restricted within the framework of classical electrodynamics, is quite efficient for solving scattering problems in the resonance region but meets certain computational difficulties in the short-wavelength region.

The main purpose of this work is to find the domain of applicability of the methods mentioned above by analyzing the numerical solution to the problem of plane wave scattering from a perfectly conducting plane surface arbitrarily perturbed along one dimension on an interval of finite width. The roughness height and slope constraints are discussed in the case when the integral equations and the Rayleigh approximation are employed. We pay special attention to the plane wave scattering at near-grazing angles.

FORMULATION OF THE PROBLEM

Consider the two-dimensional scattering problem for the plane wave, incident on random rough surface Σ, $u_i(P) = \exp(i\alpha_0 x - i\beta_0 y)$, where $\alpha_0 = k\sin\Theta_0$, $\beta_0 = k\cos\Theta_0$, Θ_0 is the incidence angle, k is the wave number, and $P = (x, y)$ is an observation point. The analysis is performed in the frequency domain with $\exp(-i\omega t)$ as the time-dependent factor. Surface Σ is described by a
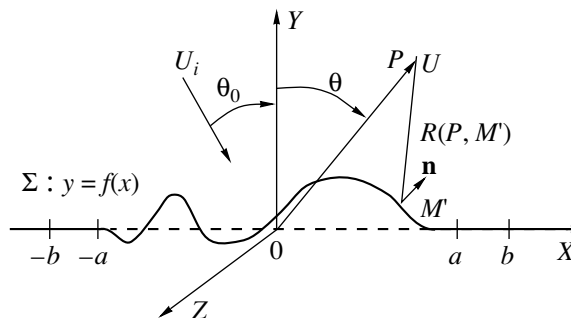

Fig. 1. Geometry of the scattering surface and notations.

smooth function $y = f(x)$ (Fig. 1). An arbitrarily perturbed surface fragment is located on the interval $x \in (-a, a)$, $a > 0$, i.e., $f(x) \equiv 0$ for $|x| > a$. Total wave field $u(P)$, which is the electric or magnetic z-component of a linearly polarized electromagnetic wave, is a solution to the Helmholtz equation

$$\Delta u + k^2 u = 0 \qquad (1)$$

and satisfies the Dirichlet,

$$u_\Sigma = 0, \qquad (2)$$

or Neumann,

$$\left.\frac{\partial u}{\partial n}\right|_\Sigma = 0, \qquad (3)$$

boundary condition in the case of electric or magnetic polarization, respectively. Represent total field u in the form

$$u = u_i + u_{rm} + u_s, \qquad (4)$$

where $u_{rm} = \varepsilon \exp(i\alpha_0 x + i\beta_0 y)$ is the plane wave regularly reflected from the plane $y = 0$; $\varepsilon = -1$ and $\varepsilon = 1$ for boundary conditions (2) and (3), respectively; and u_s is the scattered component of the total field.

THE INTEGRAL EQUATION METHOD

In the case of electric polarization, we look for component $u_s(P)$ in the form of the double layer potential with unknown density $\nu(M')$

$$u_s(P) = \int_\Sigma \nu(M')\frac{\partial}{\partial n'}H_0^{(1)}(kR(P, M'))ds \qquad (5)$$

satisfying the boundary condition

$$u_s|_\Sigma = 2i\exp(i\alpha_0 x)\sin(\beta_0 f(x)), \qquad (6)$$

and in the case of electric–magnetic polarization, we look for $u_s(P)$ in the form of the simple layer potential with density $\mu(M')$

$$u_s(P) = \int_\Sigma \mu(M')H_0^{(1)}(kR(P, M'))ds \qquad (7)$$

satisfying the boundary condition

$$\frac{\partial u_s}{\partial n_\Sigma} = (-2\exp(i\alpha_0 x)\{\beta_0\sin(\beta_0 f(x)) \qquad (8)$$

$$+ \alpha_0 if'(x)\cos(\beta_0 f(x))\})/(\sqrt{1 + f'(x)^2}).$$

In formulas (5) and (7), $M' = (x', y')$ is an integration point on scattering surface Σ, $\partial/\partial n'$ is the normal derivative calculated at M', $H_0^{(1)}(x)$ is the zero-order Hankel function of the first kind, $R(P, M') = \sqrt{(x - x')^2 + (y - f(x'))^2}$ is the distance between P and M', and ds is the differential of the arc length. Employ-

ing the known properties of the limiting values of the double layer potential and the normal derivative of the simple layer potential at $P \longrightarrow M \in \Sigma$ (see, for example, [5]), we obtain the following integral equations of the second kind for unknown functions $\nu(M)$ and $\mu(M)$:

$$\nu(M) = \nu_0(M) + \frac{1}{2i}\int_\Sigma \nu(M')\frac{\partial}{\partial n'}H_0^{(1)}(kR(M, M'))ds, \qquad (9)$$

$$M \in \Sigma,$$

$$\mu(M) = \mu_0(M) - \frac{1}{2i}\int_\Sigma \mu(M')\frac{\partial}{\partial n}H_0^{(1)}(kR(M, M'))ds, \qquad (10)$$

$$M \in \Sigma,$$

where the integrals are taken in the sense of their principal values and the right-hand sides are specified by

$$\nu_0(M) = -\exp(i\alpha_0 x)\sin(\beta_0 f(x)),$$

$$\mu_0(M)$$
$$= \frac{i\exp(i\alpha_0 x)\{\beta_0\sin(\beta_0 f(x)) + i\alpha_0 f'(x)\cos(\beta_0 f(x))\}}{\sqrt{1 + f'(x)^2}}.$$

For the convenience of the subsequent analysis, we represent integral equations (9) and (10) in a form where all the functions depend on the horizontal coordinate. Calculating the normal derivative in (9) and (10) yields the following integral equations for the electric and magnetic polarizations:

$$\nu(x) = \nu_0(x) + \int_{-\infty}^{+\infty} K_D(x, x')\nu(x')dx', \qquad (11)$$

$$\mu(x) = \mu_0(x) + \int_{-\infty}^{+\infty} K_N(x, x')\mu(x')dx', \qquad (12)$$

with the kernels

$$K_D(x, x') = \frac{kH_1^{(1)}(kR(x, x'))}{2iR(x, x')} \qquad (13)$$
$$\times [f(x) - f(x') - f'(x')(x - x')],$$

$$K_N(x, x') = \frac{kH_1^{(1)}(kR(x, x'))}{2iR(x, x')} \qquad (14)$$
$$\times [f(x) - f(x') - f'(x)(x - x')],$$

where $R(x, x') = \sqrt{(x - x')^2 + (f(x) - f(x'))^2}$ and $H_1^{(1)}(x)$ is the first-order Hankel function of the first kind.

When deriving (11) and (12), we used the expression for the differential of arc length $ds = \sqrt{1 + f'(x)^2}\,dx'$. One can see that kernels (13) and (14) of equations (11) and (12) are continuous functions.

Integral equations (11) and (12) have no solutions expressed in an explicit form. Since the integration domain in (11) and (12) is unbounded, applying numerical methods for solving these equations creates an obvious problem. In order to overcome this difficulty, we should take into account that the scattered field u_s is due to the perturbation of the horizontal plane $y = 0$ localized on the interval $x \in (-a, a)$. Therefore, the Fredholm integral equations of the second kind can be obtained on the interval $x, x' \in (-b, b)$, $b > a$ [6]

$$
\nu(x) = \nu_0(x) + \int_{-b}^{b} \left[K_D(x, x') \right.
$$

$$
\left. - \int_{-b}^{b} K_D(x, x'') K_D(x'', x') dx'' \right] \nu(x') dx',
$$

$$(15)$$

$$
\mu(x) = \mu_0(x) + \int_{-b}^{b} \left[K_N(x, x') \right.
$$

$$
\left. - \int_{-b}^{b} K_N(x, x'') K_N(x'', x') dx'' \right] \mu(x') dx'.
$$

$$(16)$$

Quantizing these equations with respect to x and x' results in a system of linear algebraic equations (SLAE), which is solved numerically. For $|x| > b$, the potential density can be computed by the formulas

$$
\nu(x) = -\int_{-a}^{a} K_D(x, x') \nu(x') dx',
$$

$$
\mu(x) = -\int_{-a}^{a} K_N(x, x') \mu(x') dx'.
$$

In some papers devoted to the problem under consideration, for example in [7], a singular integral equation of the first kind is used. However, this approach is hampered when calculating fields scattered by rough surfaces with high slopes, especially at grazing incidence [6]. On the other hand, the method based on Fredholm equations of the second kind (15) and (16) is free from these drawbacks.

Using the integral representation for the Hankel function

$$
H_0^{(1)}(k\sqrt{x^2 + y^2}) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \exp[i\alpha x + i\beta|y|]\frac{d\alpha}{\beta},
$$

where $\beta = \sqrt{k^2 - \alpha^2}$ and $\mathrm{Im}\,\alpha \geq 0$ for the scattered field at the observation points located above surface Σ, we obtain from (5) and (7) the scattered field in the form of

the integral decomposition in terms of plane waves

$$
u_s(P) = \int_{-\infty}^{+\infty} A(\alpha)\exp[i\alpha x + i\beta y]\frac{d\alpha}{\beta}, \quad (17)
$$

where amplitude spectrum $A(\alpha)$ in the integrand is specified by

$$
A(\alpha) = \frac{i}{\pi}\int_{-\infty}^{+\infty} \nu(x)[\alpha f'(x) - \beta]
$$

$$
\times \exp[-i\alpha x - i\beta f(x)]dx
$$

$$(18)$$

for the electric polarization and by

$$
A(\alpha) = \frac{1}{\pi}\int_{-\infty}^{+\infty} \mu(x)\sqrt{1 + f'(x)^2}\exp[-i\alpha x - i\beta f(x)]dx \quad (19)
$$

for the magnetic polarization. When potential densities $\nu(x)$ and $\mu(x)$ are determined by numerically solving integral equations (15) and (16), amplitude spectrum $A(\alpha)$ can be computed from (18) and (19).

In fact, uniform plane waves, for which $-k < \alpha < k$, are of practical interest. Nonuniform waves contribute to scattered field $u_s(P)$ only in the proximity of Σ. Taking into account only uniform waves, representation (17) can be recast as

$$
u_s(x, y) = \int_{-\pi/2}^{\pi/2} A(\Theta)\exp[ik(x\sin\Theta + y\cos\Theta)]d\Theta, \quad (20)
$$

where amplitude spectrum $A(\Theta)$ is given by

$$
A(\Theta) = \frac{ik}{\pi}\int_{-\infty}^{+\infty} \nu(x)[f'(x)\sin\Theta - \cos\Theta]
$$

$$
\times \exp[-ik(x\sin\Theta + \cos\Theta f(x))]dx
$$

$$(21)
$$

and

$$
A(\Theta) = \frac{k}{\pi}\int_{-\infty}^{+\infty} \mu(x)\sqrt{1 + f'(x)^2}
$$

$$
\times \exp[-ik(x\sin\Theta + \cos\Theta f(x))]dx
$$

$$(22)
$$

for the electric and magnetic polarizations, respectively. In these formulas, angle Θ is specified by the following relationships: $\alpha = k\sin\Theta$, $\beta = k\cos\Theta$, and $-\pi/2 < \Theta < \pi/2$.

The scattered field can be represented in another form which is rather efficient in the case when observation point $P(x, y)$ is far from the perturbed region. Using in formulas (5) and (7) the asymptotics of Hankel function $H_0^{(1)}(x)$ for large arguments

$$
H_0^{(1)}(x) = \sqrt{\frac{2}{\pi x}}\exp[ix - i\pi/4](1 + O(x^{-1})),
$$

and the asymptotic formula for the distance

$$R(P, M') = r - (x'\sin\Theta + f(x')\cos\Theta) + O(r^{-1}),$$
$$r \longrightarrow \infty,$$

where $x = r\cos\Theta$ and $y = r\sin\Theta$, yields the expression for the principal term in the representation of the scattered field $u_s(x, y)$ in the form of a cylindrical wave,

$$u_s(P) = \sqrt{\frac{2\pi}{kr}}\exp[ikr - i\pi/4]A(\Theta) + o((kr)^{-1/2}). \tag{23}$$

This formula implies that scattering pattern $S(\Theta)$, which is proportional to the scattered field intensity, is

$$S(\Theta) = |A(\Theta)|^2.$$

Thus, the values of $A(\Theta)$ calculated from the numerical solution of integral equations (15) and (16) determine the scattering pattern, the latter being a physical quantity of practical interest.

For an arbitrary roughness profile and angle of incidence Θ_0, the amplitude spectrum satisfies the relationship

$$\int_{-\pi/2}^{\pi/2} |A(\Theta)|^2 d\Theta = -2\varepsilon\,\mathrm{Re}\,A(\Theta_0), \tag{24}$$

which is known in scattering theory as the optical theorem (see, for example, [7]).

In order to control computations, it is convenient to employ this identity in the form

$$\Delta E(\Theta_0) \stackrel{\text{def}}{=} \left|1 + \frac{\displaystyle\int_{-\pi/2}^{\pi/2} |A(\Theta)|^2 d\Theta}{2\varepsilon\,\mathrm{Re}\,A(\Theta_0)}\right| \equiv 0. \tag{25}$$

Expression (25) is called the energy balance criterion (EBC). The deviation of $\Delta E(\Theta_0)$ from zero characterizes the accuracy of computations. Note that the condition $\Delta E(\Theta_0) \equiv 0$ is necessary but, rigorously speaking, it is insufficient for the correctness of computations of the amplitude spectrum.

THE RAYLEIGH APPROXIMATION METHOD

According to the Rayleigh hypothesis, which is well known in scattering theory, the integral decomposition of the scattered field in plane waves

$$u_s(P) = \int_{-\infty}^{\infty} A(\alpha)\exp(i\alpha x + i\beta y)d\alpha$$

is assumed to be valid everywhere up to the scattering boundary surface Σ. Then, the Fredholm integral equation of the second kind with continuous kernels can be obtained for the amplitude spectrum $A(\alpha, \alpha_0)$ of uni-

form waves [8]. For electric polarization, this equation has the form

$$A(\alpha, \alpha_0) = A_D^{(0)}(\alpha, \alpha_0)$$
$$+ \int_{-\pi/2}^{\pi/2} R_D(\alpha, \alpha')A(\alpha', \alpha_0)d\alpha', \tag{26}$$

where the right-hand side and the kernel are specified by

$$A_D^{(0)}(\alpha, \alpha_0) = \frac{1}{\pi}\int_{-a}^{a} \sin(\beta_0 f(x))[i\beta\cos(\beta f(x))$$
$$- \alpha f'(x)\sin(\beta f(x))]\exp[i(\alpha_0 - \alpha)x]dx,$$

$$R_D(\alpha, \alpha') = \frac{i}{2\pi}\int_{-a}^{a} \sin(\beta f(x))[\alpha'f'(x) - \beta'] \tag{27}$$
$$\times \exp[i(\alpha' - \alpha)x + i\beta'f(x)]dx,$$

$$\beta' = \sqrt{k^2 - \alpha'^2}.$$

Similarly, for magnetic polarization, we obtain

$$A(\alpha, \alpha_0) = A_N^{(0)}(\alpha, \alpha_0)$$
$$+ \int_{-\pi/2}^{\pi/2} R_N(\alpha, \alpha')A(\alpha', \alpha_0)d\alpha', \tag{28}$$

where

$$A_N^{(0)}(\alpha, \alpha_0) = \frac{1}{\pi}\int_{-a}^{a} \cos(\beta f(x))[\alpha_0 f'(x)\cos(\beta_0 f(x))$$
$$- i\beta_0\sin(\beta_0 f(x))]\exp[i(\alpha_0 - \alpha)x]dx,$$

$$R_N(\alpha, \alpha') = -\frac{1}{2\pi}\int_{-a}^{a} [\alpha f'(x)\cos(\beta f(x)) \tag{29}$$
$$+ i\beta'\sin(\beta f(x))]\exp[i(\alpha' - \alpha)x + i\beta'f(x)]dx.$$

THE KIRCHHOFF APPROXIMATION

The Kirchhoff approximation is the simplest and most widespread method for investigating scattering from large-scale undulations. According to this approximation, it is supposed that the field at any point of the scattering surface is the same as in the case when the surface at this point is replaced by the corresponding tangential plane. Thus, for electric polarization, the normal derivative of the total field is expressed by

$$\frac{\partial u}{\partial n} = 2\frac{\partial u_i}{\partial n} = -\frac{2i(\alpha_0 f'(x) + \beta_0)}{\sqrt{1 + f'(x)^2}}\exp[i\alpha_0 x - i\beta_0 f(x)].$$

Using Green's formula yields the explicit integral representation for the amplitude spectrum of the scattered field [9]

$$A(\Theta, \Theta_0) = \varepsilon \frac{k[1 + \cos(\Theta + \Theta_0)]}{2\pi(\cos\Theta + \cos\Theta_0)}$$

$$\times \int_{-a}^{a} \exp[i(\alpha_0 - \alpha)x]\{\exp[-i(\beta + \beta_0)f(x)] - 1\}dx. \tag{30}$$

In this approximation, the absolute value of the amplitude spectrum and, hence, the scattering pattern are independent of the incident wave polarization.

NUMERICAL IMPLEMENTATION AND RESULTS OF COMPUTATIONS

Solving the integral equations obtained above presents no principal problems. On the other hand, the application of numerical methods for solving integral equations arising in scattering analysis meets certain computational difficulties. The problem is that the kernels and right-hand sides of (15) and (16) are rapidly oscillating functions. Numerical simulations show that integral equations (15) and (16) should be discretized under the condition $N \geq 8(2b/\lambda)$, where N is the number of discretization points. Thus, even in order to calculate fields scattered from an undulation of the width $2a \approx 100\lambda$, one needs to solve a SLAE with a complex $10^3 \times 10^3$-matrix. Solving this SLAE by the Gauss technique requires $\approx 10^9$ operations with complex numbers and about 8 MB of a PC RAM for storing the matrix. In fact, actual technical problems involve scattering from surfaces with roughness regions (bright areas) of about $10^4\lambda$ in width. In this situation, the stored matrix occupies at least 95 GB of the RAM, and solving the corresponding SLAE requires approximately 10^{15} arithmetic complex operations, which is a challenge for even modern computers. In addition, one should keep in mind that numerous scattering problems are formulated statistically; i.e., it is necessary to simulate scattering using an ensemble of surfaces and, then, to compute the required scattering parameters. On the other hand, the obvious progress in computer technology allows us to hope that the difficulties mentioned above will be overcome in the near future.

The integral equations of the Rayleigh hypothesis are simpler for computational implementation than the integral equations of the rigorous method. The discretization interval for integral equations (26) and (28) of the Rayleigh approximation should be chosen depending primarily on the required resolution of the calculated scattering pattern. In addition, one should take into account that the time of computing one element of the corresponding matrix is proportional to $2a$. The most time-consuming operation involved in the numerical implementation of the Rayleigh approximation is calculating the matrices corresponding to the integral

equations. It should be noted that the frequency of the amplitude spectrum variations increases with the width of the illuminated region. Therefore, in order to calculate the amplitude spectrum on increasingly wider surfaces, it is necessary to increase N even if one does not need to provide a resolution as high as this.

The methods for calculating scattering patterns described above (the numerical solution of rigorous integral equations (9) and (10), and the direct calculation of the scattering pattern using equations (26) and (28) of the Rayleigh hypothesis and representation (30) of the Kirchhoff approximation) are implemented in computer codes for IBM PC/AT. We present the results of calculations for two typical situations in order to find the domains of applicability of the discussed methods, which is important for future numerical simulations of actual scattering problems.

The scattering pattern is calculated for the cylindrical surface described by the function

$$f(x) = A_0\cos(2\pi x/d)V(x), \tag{31}$$

where $V(x)$ is a patch function specified by

$$V(x) = \begin{cases} \dfrac{a+x}{a-c} - \dfrac{1}{2\pi}\sin\left(2\pi\dfrac{a+x}{a-c}\right), & \text{if } x \in (-a, -c) \\ 0 \leq c < a \\ 1, & \text{if } x \in (-c, -c) \\ \dfrac{a-x}{a-c} - \dfrac{1}{2\pi}\sin\left(2\pi\dfrac{a-x}{a-c}\right), & \text{if } x \in (c, a) \\ 0, & \text{if } x \notin (-a, a). \end{cases}$$

The corresponding values of $\Delta E(\Theta_0)$, which characterize the relative accuracy of computations (the extent to which the EBC is fulfilled) are given in the figure captions. Quantity $\Delta E(\Theta_0)$ is denoted by ΔE_D and ΔE_N for the *TE* and *TM* polarizations, respectively. All the geometric parameters of the problem are expressed in terms of wavelength.

In the first series of calculations, we investigate the domain of validity of integral equations (26) and (28) obtained in the Rayleigh approximation. As known, for the infinite periodic surface $f(x) = A_0\cos(2\pi x/d)$, the condition of applicability of this approximation has the form $2\pi A_0/d < 0.448$ [10]. Our computations show that this condition is necessary. When the EBC is fulfilled to within the maximum permissible error $\Delta E \leq 0.05$, A_0 in (31) cannot exceed 0.2. This is illustrated by Fig. 2, which shows the scattering pattern for profile (31) with the parameters $a = 50$, $c = 45$, $A_0 = 0.1$, and $d = 2$ for the incidence angle $\Theta_0 = 89°$. This pattern is calculated for both polarizations using integral equations (26) and (28) of the Rayleigh hypothesis and rigorous integral equations (15) and (16). The results obtained by both methods are in good agreement. In this case, a difference between two polarizations observed in scattering patterns for near-grazing incidence is manifested in dif-
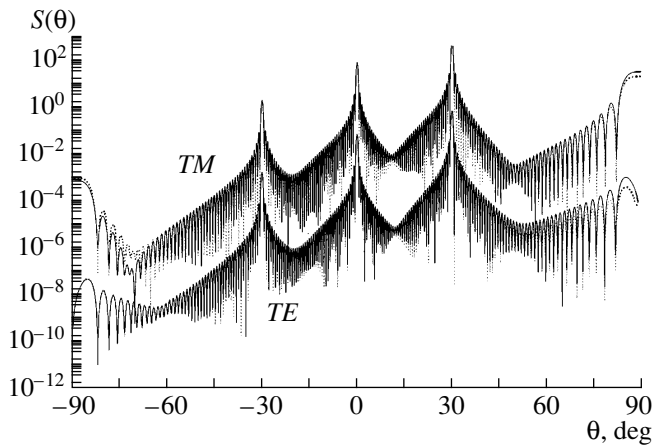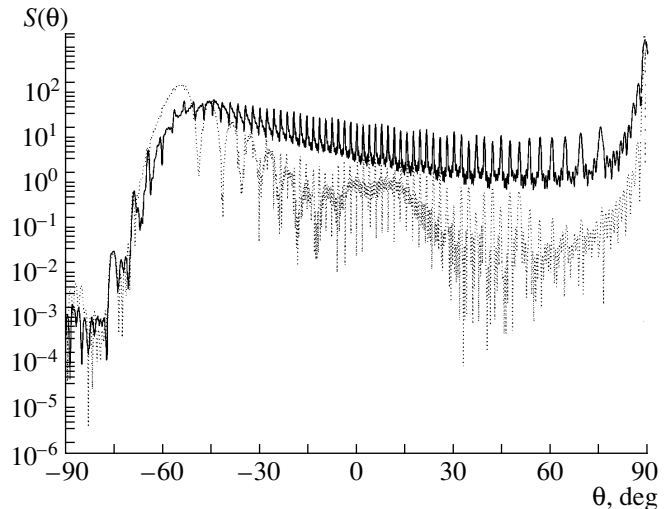
Fig. 2. Plane wave scattering pattern for profile (31) with the parameters $a = 50$, $c = 45$, $A_0 = 0.1$, and $d = 2$ calculated for the incidence angle $\Theta_0 = 89°$. Results obtained using integral equations (26) and (28) of the Rayleigh approximation (solid lines) and integral equations (15) and (16) of the rigorous method (dotted lines). EBC: $\Delta E_D = 4.9 \times 10^{-2}$, $\Delta E_N = 4.0 \times 10^{-2}$ (the Rayleigh approximation), $\Delta E_D = 1.1 \times 10^{-4}$, $\Delta E_N = 1.1 \times 10^{-5}$ (the rigorous integral equation method).
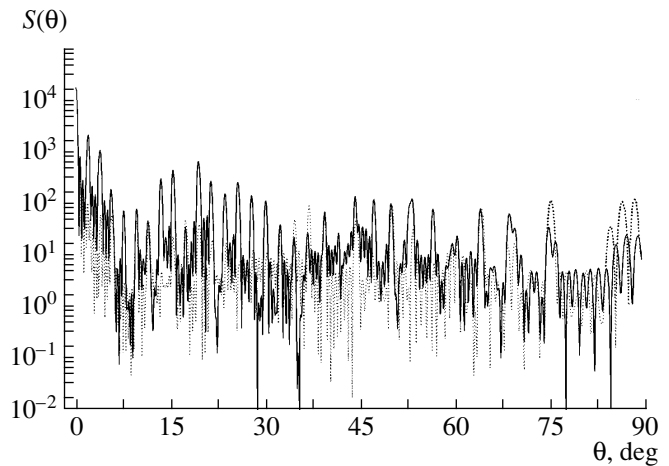
Fig. 3. Plane wave scattering pattern for profile (31) with the parameters $a = 70$, $b = 75$, $c = 65$, $A_0 = 15$, and $d = 30$ calculated for the incidence angle $\Theta_0 = 0$. Results obtained using rigorous integral equation (15) for the *TE* polarization (solid line) and the Kirchhoff approximation (30) (dotted line); EBC: $\Delta E_D = 3 \times 10^{-6}$ (the rigorous integral equation method), $\Delta E = 0.39$ (the Kirchhoff approximation).
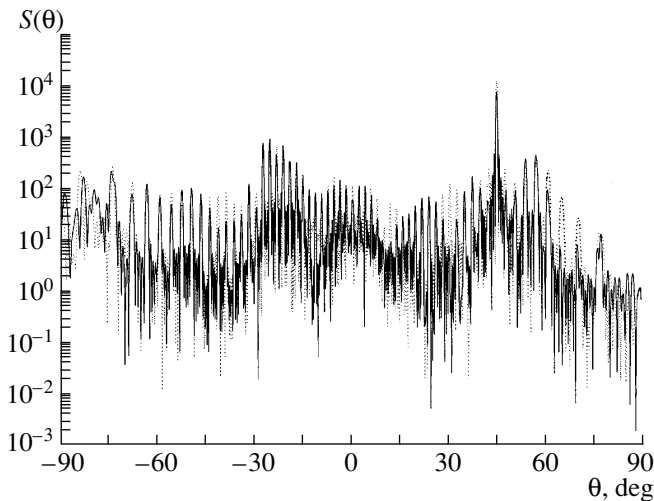
Fig. 4. Plane wave scattering pattern for profile (31) with the parameters $a = 70$, $b = 75$, $c = 65$, $A_0 = 15$, and $d = 30$ calculated using rigorous integral equations (15) and (16) at the incidence angle $\Theta_0 = 45°$ for the *TE* (solid line) and *TM* (dotted line) polarizations. EBC: $\Delta E_D = 4.5 \times 10^{-4}$, $\Delta E_N = 2.5 \times 10^{-4}$.

Fig. 5. Plane wave scattering pattern for profile (31) with the parameters $a = 70$, $b = 75$, $c = 65$, $A_0 = 15$, and $d = 30$ calculated for the incidence angle $\Theta_0 = 89°$. EBC: $\Delta E_D = 2.6 \times 10^{-4}$, $\Delta E_N = 2.5 \times 10^{-4}$.

ferent scale coefficients but not in pattern behavior. The Kirchhoff approximation, surely, cannot be applied to this geometry.

Thus, in order to calculate fields scattered from shallow rough surfaces with low slopes in the resonance region, integral equations (26) and (28) of the Rayleigh hypothesis can be used as a simpler (for computational implementation) alternative to the rigorous integral equation method.

In the second series of calculations, we investigated the applicability of the rigorous integral equation method for computing fields scattered from large-scale rough surfaces. As mentioned above, the key problem typical for these computations is the rapidly (proportionally to $(2a)^3$) increasing number of arithmetic operations necessary for solving the integral equations. In addition, the integral equations should be discretized using a more refined mesh as the slopes of irregularities

on the scattering surface increase. However, the drawbacks of the rigorous integral equation method are compensated for by high accuracy and reliability of results. Figures 3–5 show scattering patterns for the large-scale surface with high slopes characterized by the following parameters: $a = 70$, $b = 75$, $c = 65$, $A_0 = 15$, and $d = 30$. Figure 3 displays the scattering patterns calculated for the incidence angle $\Theta_0 = 0°$ using the rigorous integral equation method for the *TE* polarization and the Kirchhoff approximation. Due to the symmetric geometry under consideration, the scattering pattern is an even function of the scattering angle Θ for normal incidence, and, therefore, we show the pattern only in the range $0° \le \Theta \le 90°$. One can see that even in the case of normal incidence, the Kirchhoff approximation provides sufficiently accurate results only in the proximity of the symmetric reflection direction. This fact can be attributed to the existence of multiple scattering from surface elements, which is ignored in the Kirchhoff approximation. Figures 4 and 5 demonstrate the corresponding scattering patterns for the *TE* and *TM* polarizations when the incidence angles are $\Theta_0 = 45°$ and $89°$. The accuracy of these computations remains very high including the grazing incidence when a long shadow region is formed at $x > a$. The scattering patterns affected by the presence of this region can be correctly determined by reducing initial integral equations (11) and (12) with the infinite integration domain to equations (15) and (16) involving integrals over a finite interval. The latter equations take into account the effect of the perturbed surface fragment $(-a, a)$ on the plane part of the surface $|x| > b$.

CONCLUSION

The results of computations demonstrate that the Rayleigh approximation is efficient for determining fields scattered from shallow rough surfaces with low slopes. The accuracy of these computations proves to be sufficiently high for an arbitrary incidence of the plane wave including near-grazing angles. However, the computation error abruptly increases with the roughness height, namely, at $A_0 > 0.2$, when the Rayleigh approximation becomes invalid.

Comparing the results of computations performed using integral equations (9) and (10) of the potential theory and the Kirchhoff approximation, we can see that, in the proximity of the symmetric reflection direction, the Kirchhoff approximation rather accurately describes the fields scattered from large-scale rough surfaces with low slopes in the case of incidence far from grazing. Computations show that the Kirchhoff approximation cannot even qualitatively describe scattering from surfaces with large roughness height and high slopes or scattering in the case of near-grazing incidence. At the same time, the rigorous approach based on the Fredholm integral equations of the second kind is the only one among tested methods that solves this complicated problem, providing a high accuracy of computations in the whole range of the incidence and scattering angles.

REFERENCES

1. J. M. Soto-Crespo and M. Nieto-Vesperinas, J. Opt. Soc. Am. A **6** (3), 367 (1989).

2. J. A. Sánchez-Gil, M. Nieto-Vesperinas, F. Morino, *et al.*, J. Opt. Soc. Am. A **10** (12), 2628 (1993).

3. H. D. Ngo and C. L. Rino, Radio Sci. **29** (6), 1365 (1994).

4. D. E. Barrick, Radio Sci. **30** (3), 563 (1995).

5. V. S. Vladimirov, *Mathematical Physics Equations* (Nauka, Moscow, 1981).

6. Yu. A. Baloshin and A. V. Kostin, in *Int. Meth. Sci. and Eng. Vol. 1. Analytic Methods* (Addison Wesley Longman, London, 1997), pp. 32–37.

7. D. Maystre, IEEE Trans. Antennas Propag. **AP-31** (6), 885 (1983).

8. Yu. A. Baloshin, V. V. Zalipaev, A. V. Kostin, *et al.*, Opt. Zh. **63** (12), 24 (1996).

9. D. Maystre, O. Mata Mendes, and A. Roger, Opt. Acta **30** (12), 1707 (1983).

10. R. F. Millar, Proc. Cambridge Philos. Soc. **65** Part 3, 773 (1969).

Translated by I. G. Efimova

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Screw Extruder for Solid Hydrogen

I. V. Vinyar and A. Ya. Lukin

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

Received May 21, 1998

Abstract—The first screw extruder capable of continuously forming macroscopic solid hydrogen particles to be injected into the plasma of thermonuclear devices was created and successfully tested. A model of nonisothermal hydrogen flow in the screw channel was developed and used to calculate the extruder efficiency as a function of the system parameters. The results of calculations and experimental data are presented for the extrusion of 500 cm^3 of solid hydrogen using an extruder with a working volume of 8 cm^3 operating at a rate of 63 mm^3/s. Temperature dependences of the viscosity and shear strength of solid hydrogen were studied in the temperature interval from 10 to 13 K, where solid hydrogen behaves like a Bingham fluid. © *2000 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

A fuel can be supplied to the plasma of thermonuclear setups either by admitting gaseous hydrogen or by injecting macroparticles of solidified hydrogen isotopes. The injection of particles into a reactor operating in the stationary regime must be performed at a frequency of 1–10 kHz [1]. A device recently suggested for this purpose is based on the periodic formation of macroparticles by rapid freezing of liquid drops delivered into the injector channel through a porous sleeve [2, 3]. An alternative injector system employs three sequentially operating piston extruders for processing preliminarily frozen hydrogen into a rod, followed by cutting this rod into macroparticles [4]. The system of fuel supply by macroparticle injection used in the ITEP reactor has to meet high requirements with respect to the reliability of operation (99%) [5]. This circumstance stimulates the search for the most simple and, hence, reliable methods of the hydrogen macroparticle formation. In 1990–1994, we studied the process of fuel extrusion under the action of pressure developed in the gas phase [6]. Experimental data gained in that work were used to develop a new method for the macroparticle formation using the principle of a screw extruder.

EXTRUSION PROCESS: DESCRIPTION AND MODEL

In a conventional screw extruder, a material to be processed is supplied to the screw channel in the solid state. As the material is pressed and transferred to the discharge head, the temperature is increased so as to provide for a decrease in the shear strength and viscosity necessary for the extrusion process. In the case of hydrogen extrusion, the temperature must be decreased on approaching the head in order to provide for the hydrogen condensation and solidification in the screw

channel. A decrease in the temperature is accompanied by increase in the strength and viscosity of hydrogen, but the compression at a pressure exceeding 5 MPa (necessary for the hydrogen extrusion [3, 6]) leads to a considerable heat evolution. This phenomenon frequently causes local heating and melting of hydrogen, which results in cessation of the extrusion process [7, 8]. Therefore, a mathematical model used to describe the process must allow for a nonisothermal character of flow and compression in solid hydrogen and take into account the temperature dependence of hydrogen viscosity and shear stress.

For determining the rheological characteristics of solid hydrogen, we used the data obtained previously [6] on the behavior of hydrogen extruded through a thin narrowing channel. Our analysis of numerous models describing the flow of non-Newtonian fluids showed that the flow of plastic hydrogen in the temperature interval 10–13 K under a pressure of 2.5–10 MPa can be described using a model of viscoplastic Bingham's fluid. According to this model, the material behaves as an absolutely hard body unless the shear stress τ exceeds a certain threshold τ_0, and as a viscous fluid, when the stress is above this threshold level:

$$\dot{\gamma} = 0 \ \text{ for } \tau \leq \tau_0, \quad \mu\dot{\gamma} = \tau - \tau_0 \ \text{ for } \ \tau > \tau_0.$$

Here and below $\dot{\gamma}$ denotes the shear strain rate, μ is the plastic viscosity, and τ_0 is the threshold (shear onset) stress. The isothermal steady-state flow of a viscoplastic medium along the z-axis in a straight tube with a variable radius R is described by the following equation:

$$\frac{dP}{dz} = \frac{1}{r}\frac{d(r\tau_{rz})}{dr}, \tag{1}$$

where P is the pressure in a cross-section with the coordinate z (P is assumed to be constant across the tube,

i.e., independent of the radial coordinate r) and τ_{rz} is the strain tensor component. Assuming small conicity of the tube, equation (1) can be used in all cross sections with the current values of radius R. Expressing dP/dz through a constant flow rate Q and numerically integrating with respect to z, we obtain a relationship between P and Q. In the case of a tube with constant cross section, this relationship reduces to the well-known Buckingham formula [9]:

$$Q = \frac{\pi R^4}{8\mu}\frac{\Delta P}{l}\left[1 - \frac{4}{3}\left(\frac{2\tau_0 l}{R\Delta P}\right) + \frac{1}{3}\left(\frac{2\tau_0 l}{R\Delta P}\right)^4\right], \quad (2)$$

where ΔP is the pressure drop along the path l. Taking the results of experiments on the hydrogen extrusion [6], which are depicted by points in Fig. 1, and using the procedure of sequential iterations, we may calculate the variation of the shear onset stress τ_0 and the viscosity μ as functions of the temperature. The initial approximation for τ_0 was determined by the pressure corresponding to a zero extrusion velocity, and the initial μ value was estimated from the maximum flow rate. The iterative process leads to the following approximate relationships:

$$\begin{aligned}\tau_0 &= 0.026\exp[0.28(T_s - T)]\ \text{MPa},\\ \mu &= 0.0027\exp[0.44(T_s - T)]\ \text{MPa s},\end{aligned} \quad (3)$$

where T is the current temperatures of hydrogen in the interval 10–13 K, and T_s is the temperature corresponding to a triple point.

Solid curves in Fig. 1 show the plots of calculated extrusion velocity versus temperature for various pressures. The good coincidence of theory and experiment is evidence that the flow in solid hydrogen during extrusion at temperatures in the 10–13 K interval is satisfactorily described by a model of the Bingham fluid type with the parameters indicated above. It should be noted that the shear onset stresses calculated by formula (3) virtually coincide with the yield points of solid hydrogen determined in the 10–12 K interval and exceed these values by 30% on approaching 13 K [10]. The second formula (3) presents for the first time an estimate of the plastic viscosity coefficient not reported previously for solid hydrogen [10, 11].

We will use the above estimates for τ_0 and μ in modeling the hydrogen flow in a screw extruder. The extrusion process is described using a system of coordinates related to the screw. In this system, the immobile screw is situated inside a cylinder rotating in the reverse direction and solid hydrogen moves along the screw channel toward the discharge head. The screw channel is represented by a developable surface with neglect of its curvature. The position of a point in the channel is characterized by a set of coordinates (Fig. 2): x and y are the coordinates in a cross section of the screw channel ($0 \leq x \leq w$; $0 \leq y \leq h$) having width w and depth h, and z is the coordinate along the channel. As is known [12], a screw with the diameter $D = 2R$ and $h \ll Dw$ is satisfac-
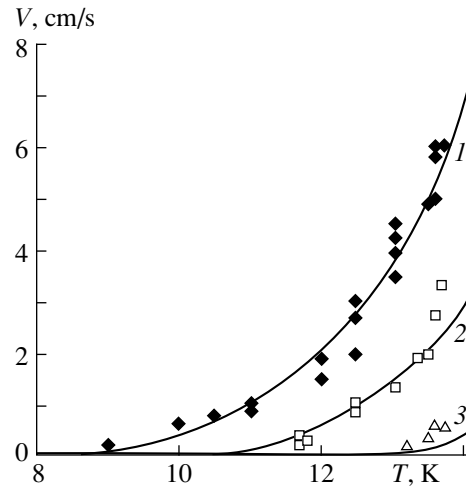

Fig. 1. The plots of hydrogen extrusion velocity versus temperature at various pressures (MPa): (*1*) 10; (*2*) 5.5; (*3*) 2.5; symbols represent experimental points, solid curves show the results of calculation.
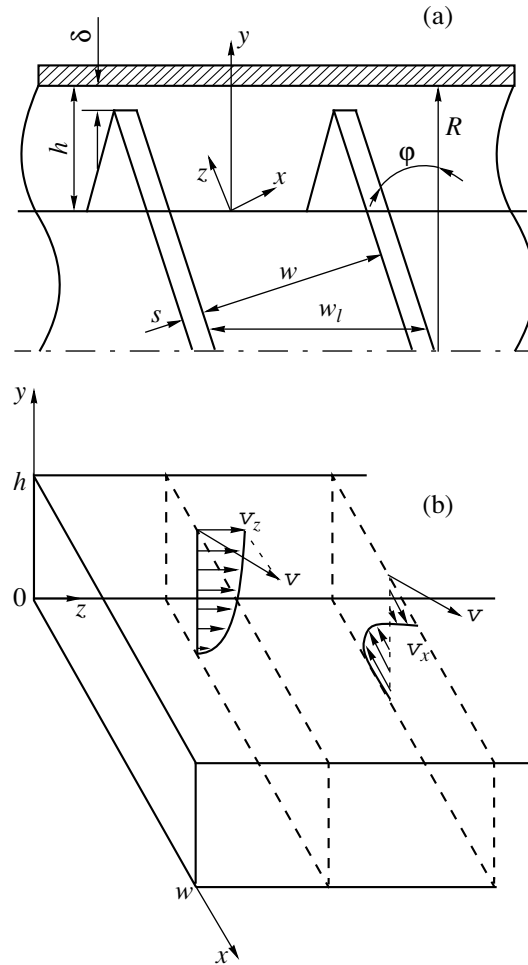

Fig. 2. Schematic diagrams of (a) a screw extruder and (b) a developable screw channel surface with extrusion velocity projections.

torily described with neglect of both the end effects and the curvature. In this approximation, the motion of a substance in the screw channel is equivalent to a flow between two planes moving relative to one another. The equation of motion and the corresponding boundary conditions (formulated assuming $v_y = 0$ in the whole cross-section) are as follows:

$$\frac{\partial \tau_{xy}}{\partial y} = \frac{\partial P}{\partial x}, \quad \frac{\partial \tau_{yz}}{\partial y} = \frac{\partial P}{\partial z} \text{ for } 0 < x < w, \quad 0 < y < h,$$

$$\mu \dot{\gamma}_{xy} = \tau_{xy}(1 - \tau_0/\tau), \quad \mu \dot{\gamma}_{yz} = \tau_{yz}(1 - \tau_0/\tau) \text{ for } \tau > \tau_0,$$

$$\dot{\gamma} = 0 \text{ for } \tau \le \tau_0, \quad \tau = \sqrt{\tau_{xy}^2 + \tau_{yz}^2}, \quad (4)$$

$$v_x = \int_0^y \dot{\gamma}_{xy} dy, \quad v_z = \int_0^y \dot{\gamma}_{yz} dy,$$

$$v_x(h) = \overline{\omega} R \sin\varphi, \quad v_z(h) = \overline{\omega} R \cos\varphi.$$

Here and below, $\overline{\omega}$ is the angular velocity of the screw rotation and φ is the screw thread angle. The efficiency (or the volume hydrogen consumption rate) of the extruder Q is given by the formulas

$$Q = \frac{w}{\sin\varphi}(q_z \sin\varphi - q_x \cos\varphi),$$

$$q_x = \int_0^h v_x dy, \quad q_z = \int_0^h v_z dy, \quad (5)$$

where q_x and q_z are the components of hydrogen flux per unit width w of the screw channel. In the absence of losses through a gap between the screw ridges and extruder cylinder walls, the flux component q_x is zero and the pressure developed by every turn of the screw can be calculated from the system of equations (4) and (5). Considering a pressure drop between two neighboring turns, the loss flux can be approximately determined using the following system of equations:

$$\frac{\partial \tau_{xy}}{\partial y} = \frac{\partial P}{\partial x} \text{ for } w < x < w + s, \quad h - \delta < y < h,$$

$$\mu \dot{\gamma}_{xy} = \tau_{xy}(1 - \tau_0/\tau) \text{ for } \tau > \tau_0,$$

$$q_{x\delta} = \int_{h-\delta}^h v_x dy, \quad v_x = \int_{h-\delta}^h \dot{\gamma}_{xy} dy, \quad (6)$$

$$\dot{\gamma} = 0 \text{ for } \tau \le \tau_0,$$

where s is the width of the screw ridge and δ is the gap width between screw and cylinder, which typically amounts to $(0.001–0.002)R$ [13]. The hydrogen flux component along the z-axis in the gap does not significantly contribute to the loss flux and, hence, can be ignored. In order to make a correction for the loss flux

in the extruder efficiency, the finite flux $q_x = q_{x\delta}$ has to be used instead of $q_x = 0$ in equations (5).

Equations (4)–(6) describe a relationship between Q and dP/dz. Figure 2b shows a typical pattern of distribution of the velocities v_x and v_z along the channel height y. According to a special feature of the Bingham fluid, the flow contains an immobile (or moving as a whole) core (with $\tau < \tau_0$) and one or two zones of viscous friction.

The motion of a viscoplastic medium in the plastic strain region is accompanied by heat evolution at a local rate given by the formula

$$W = \mu[(\dot{\gamma}_{xy})^2 + (\dot{\gamma}_{yz})^2]. \quad (7)$$

The extruder is cooled by a flow of helium with the temperature T_1, the heat exchange coefficient α, and the heat transfer rate $g = c_p G$, G and c_p being the helium flow rate and heat capacity at constant pressure, respectively. Because the extruder cylinder has thin copper walls, we may assume the temperature to be constant across the cylinder and equal to the hydrogen temperature in the same cross section. The temperature of hydrogen along the y-axis is also assumed constant due to intensive mixing. The thermal balance equations written with an allowance for the heat conduction in the extruder material, the enthalpy transfer with the flow of hydrogen, and the heat exchange are as follows:

$$\frac{d}{dl}\left(-S\lambda \frac{dT}{dl} + \rho QH\right) = 2\pi R\left[\int_0^h W dy + \alpha(T_1 - T)\right], \quad (8)$$

$$\frac{d}{dl}(gT_1) = 2\pi R\alpha(T - T_1).$$

Here, S is the wall cross-section area; λ is the coefficient of thermal conductivity of the cylinder material (copper); ρ is the hydrogen density; H is the enthalpy of hydrogen with an allowance for the heat of crystallization; and T is the temperature in a given cross section of the cylinder. It is suggested that hydrogen is supplied to the extruder in a liquid state at the temperature of crystallization.

Consistent solution of the system of equations (4)–(8) allows the extruder efficiency to be calculated for various system geometries and process parameters.

RESULTS OF CALCULATIONS AND EXPERIMENTS

The main characteristic of a screw extruder represents a relationship between its efficiency and a pressure developed by the screw. Figure 3 (curves *1–4*) shows the extruder characteristics for various screw rotation speeds. The calculations were performed for the following set of parameters: screw radius $R = 6$ mm; screw channel depth $h = 2$ mm; screw thread pitch $w_l = 12$ mm; liquid helium flow rate $g = 5$ l/h; helium tem-

perature at the input of the extruder heat exchanger $T_1 = 5$ K; gap width between screw and cylinder $\delta = 0.05$ mm; screw ridge width $s = 2$ mm. As the screw rotation speed increases from 10 to 25 rpm, the maximum extruder efficiency grows, but the hydrogen compression induced by the screw drops. If the output hole diameter in the discharge head equals the internal diameter of the cylinder, then $P = 0$ and the extruder efficiency reaches 160 mm^3/s. As the output hole diameter decreases, the pressure produced by the screw grows to reach 25–60 MPa at small Q values. The working point of the extruder is determined by a discharge head resistance that can be calculated using equation (2). Lines a and b in Fig. 3 (and in the other figures) represent the characteristics of a discharge head with an output hole diameter of 2 mm and a channel length of 4 mm operating at a temperature of 10 and 11 K, respectively. The points of intersection of these lines with the extruder characteristics determine the efficiency of an extruder equipped with this discharge head and the hydrogen pressure developed by the screw. Since the process of hydrogen ice extrusion at 10–11 K requires a pressure exceeding 10 MPa [6], the working parameters of the extruder should be selected for the intersection points occurring above this value. As seen from Fig. 3, a steady-state regime of extrusion with an efficiency of 60–115 mm^3/s with respect to solid hydrogen can be provided by rotating the screw at a speed of 10–25 rpm and cooling the discharge head to a temperature of 10–11 K. As the screw rotation speed increases above the interval indicated, the extrusion process slows down and eventually ceases. This effect is explained by increased hydrogen heating as a result of intensive mixing, which leads to a decrease in the medium viscosity and a drop of the pressure developed by the screw. The excess heat supplied to helium causes more rapid heating of the coolant and renders it uncapable of cooling the same amount of hydrogen in the extruder from melting point to the preset discharge head temperature. In terms of the working extruder length (which is the distance at which hydrogen carried by the screw cools from the melting point to the discharge head temperature), an increase in the screw rotation speed decreases the working extruder length and the effective number of screw turns compressing solid hydrogen (Fig. 4). This number of turns is insufficient to produce a pressure level required for maintaining the extrusion process; moreover, liquid hydrogen filling a part of the extruder above the working screw turns may lead to instability in the extruder operation.

Figure 5 shows an example illustrating the influence of the cooling rate on the efficiency of an extruder operated at a screw rotation speed of 15 rpm. An increase of the helium flow rate in the heat exchanger shifts the working point of the extruder toward greater efficiency. This is also accompanied by increase in the hydrogen compression induced by the screw, which ensures stability of the extrusion process within broader range of conditions.
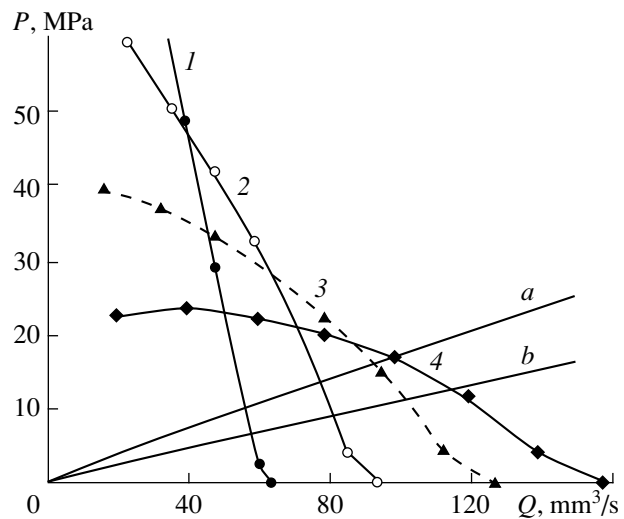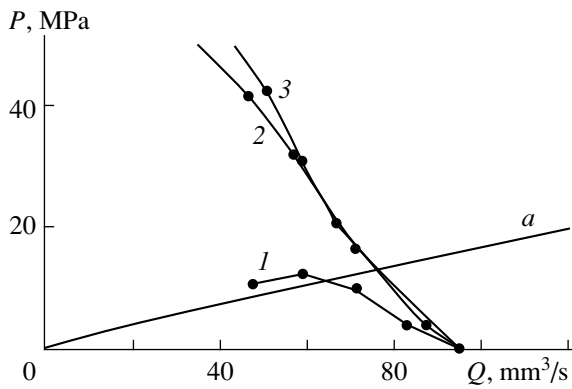
Fig. 3. The characteristics of (a, b) a discharge head operated at a hydrogen temperature of (a) 10 K and (b) 11 K and (*1–4*) a screw extruder operated at various screw rotation speeds ω (rpm): (*1*) 10; (*2*) 15; (*3*) 20; (*4*) 25.
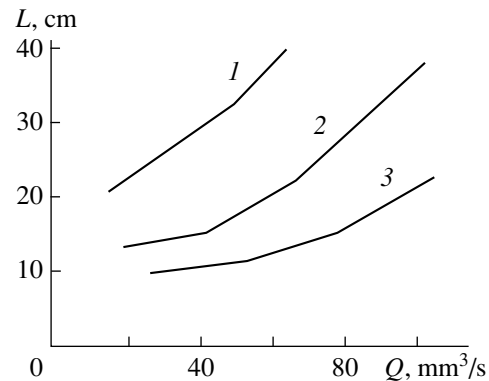

Fig. 4. The plots of working extruder length L versus efficiency Q at various screw rotation speeds ω (rpm): (*1*) 15; (*2*) 20; (*3*) 25.
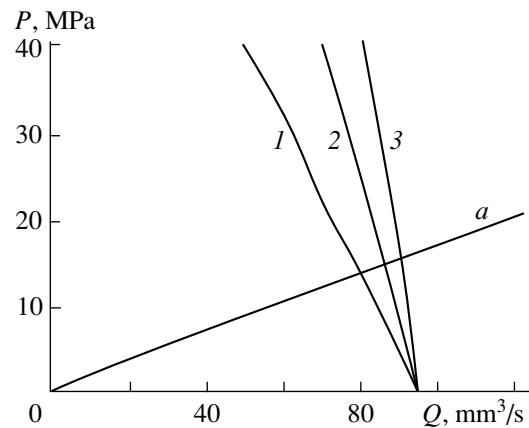
Fig. 5. The characteristics of (a) a discharge head operated at a hydrogen temperature of 10 K and (*1–3*) a screw extruder operated at various helium flow rates in a heat exchanger (l/h): (*1*) 5; (*2*) 10; (*3*) 15.

Fig. 6. The characteristics of (*a*) a discharge head operated at a hydrogen temperature of 10 K and (*1–3*) screw extruders with various screw channel depths (mm): (*1*) 1; (*2*) 2; (*3*) 3.
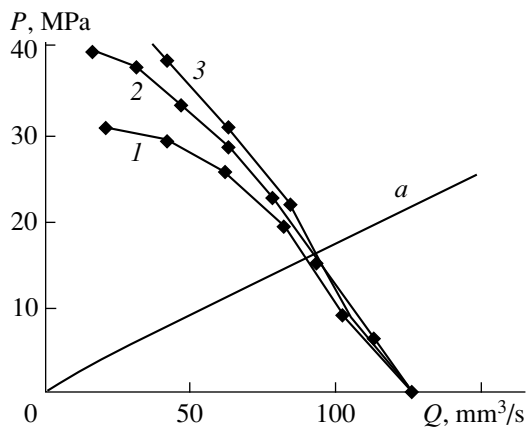

Fig. 7. The characteristics of (*a*) a discharge head operated at a hydrogen temperature of 10 K and (*1–3*) screw extruders with various screw radii (mm): (*1*) 5; (*2*) 6; (*3*) 7.

Figures 6 and 7 illustrate the effects of the channel depth and radius of the screw. To make the results of calculations comparable for various screw channel depths h, we maintained constant the product $h\overline{\omega} = 3$ mm/s, while the data for various screw radii R were compared under the conditions of constant $w_l R \overline{\omega} = 36$ mm^2/s and $w_l = 2R$. During the extrusion under isothermal conditions, a pressure developed by the screw in hydrogen increases with decreasing channel depth. In a real nonisothermal process, a decrease in the channel depth is accompanied by a sharp growth in the heat evolution, which leads to a decrease in hydrogen viscosity and a drop in the pressure. As seen from Fig. 6, the screws with 2- and 3-mm-deep channels produce approximately the same pressures at an efficiency of $Q = 75$ mm^3/s, but the screw with a smaller channel depth is advantageous by providing a more homogeneous temperature distribution along the channel height and a decrease in the amount of hydrogen (or tritium in the future systems) in the extruder.

An increase in the screw radius improves the conditions of heat exchange from hydrogen to helium and leads to a growth in the pressure developed in the extruder (Fig. 7), These factors favor an increase in stability of the extruder operation.

Based on the results of calculations performed for a 2-mm-diam rod of solid hydrogen extruded at a velocity of 20 mm/s, we have designed a screw extruder with the following set of parameters: screw radius $R = 6$ mm; screw channel depth $h = 2$ mm; screw thread pitch $w_l = 12$ mm; extruder length $L = 150$ mm; gap width between screw and cylinder $\delta = 0.05$ mm; screw ridge width $s = 2$ mm; output hole diameter in discharge head 2 mm; head length 4 mm; head temperature 10–11 K; efficiency $Q = 60$–115 mm^3/s; extrusion pressure $P = 10$–20 MPa, screw rotation speed $\overline{\omega} = 10$–25 rpm; hydrogen volume in the extruder < 8 cm^3. The vapors of liquid helium, upon cooling the extruder processing solid hydrogen, were used in an additional heat exchanger to convert hydrogen from gaseous to liquid state prior to charging it into the extruder. The extruder design and the experimental methods used to study the device operation were described elsewhere [14]. In these experiments, the extrusion process was suspended and renewed without delay by switching off/on a motor rotating the screw. The extruder head temperature was maintained on a level of 10–11 K throughout the experiment. Gaseous hydrogen was supplied from a gas container into the extruder chamber. After freezing and extrusion, evaporated hydrogen was evacuated from the system by a roughing vacuum pump. The duration of continuous extrusion exceeded one hour and was limited only by volume (100 cm^3) of the chamber to which solid hydrogen was admitted upon extrusion. The chamber was periodically purged by pulsed admitting of room- temperature helium, which accelerated the evaporation of extruded solid hydrogen and increased the working chamber volume for discharge of a new portion of solid hydrogen. The extruded hydrogen rod was transparent, which indicated a good quality of the hydrogen ice. The total volume of hydrogen ice extruded in a continuous regime exceeded 500 cm^3 without any visible evidence of deterioration of the ice quality. Since the total volume of hydrogen ice present in the extruder at one instant did not exceed 8 cm^3, we may speak of attaining a steady-state continuous regime of hydrogen freezing and extrusion.

The extruded hydrogen rod was periodically cut with a sharp edge of a thin-wall tube to obtain hydrogen macroparticles. These particles were accelerated by a flow of compressed helium supplied to the same chamber. In this way, a total of several thousand macroparticles with an average size of about 2 mm were successfully formed at a frequency of 1 to 5 Hz and accelerated to a velocity of 0.3–0.6 km/s. Upon acceleration, the particles were photographed in a diagnostic chamber. Analysis of the photographs showed that the quality of extruded hydrogen ice and macroparticles formed upon
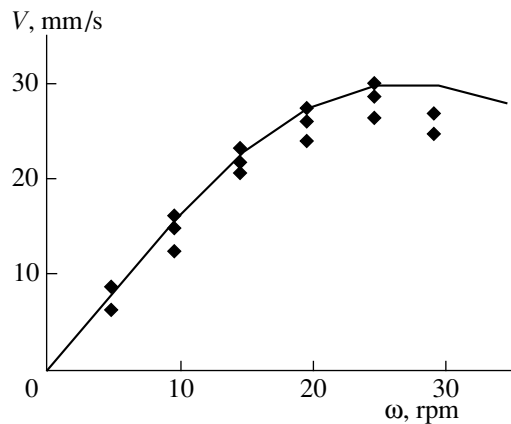
Fig. 8. The plot of extrusion velocity versus screw rotation speed; symbols represent experimental points, solid curve shows the results of calculation.

extrusion did not adversely change with time during continuous operation of the extruder.

Figure 8 shows a plot of data (points) on the extrusion velocity experimentally measured for various values of the screw rotation speed. A solid curve in the same figure represents the theoretical extrusion velocity calculated as a function of the screw rotation speed for a hydrogen temperature of 10 K at the extruder output. As the screw rotation speed increases above 25 rpm, the extrusion velocity grows at a slower rate and then even begins to decrease. These results confirm the above conclusion concerning a significant effect of the heat evolution during the compression of hydrogen upon the pressure produced by the screw and upon the extrusion velocity.

The most stable regime of extruder operation was observed for a screw rotation speed of 15 rpm, whereby the hydrogen extrusion velocity was 20 ± 4 mm/s and the average efficiency was 63 ± 13 mm^3/s. As seen from Fig. 3, this efficiency falls within the region of extruder parameters determined by calculation.

CONCLUSION

A method of the fuel macroparticle production based on the screw extruder can be used in the injectors of all types used for the fuel supply to thermonuclear devices. The proposed extruder model allows the system parameters to be calculated proceeding from a pre-

set geometry and the macroparticle injection frequency. The results obtained for the production of a 2-mm-diam hydrogen rod in the screw extruder operated at an efficiency of 63 mm^3/s demonstrated a good agreement between calculated and experimental data.

REFERENCES

1. S. Milora, W. Houlberg, L. Lenguel, and V. Mertens, Nucl. Fusion **35**, 657 (1995).
2. I. V. Viniar, RF Inventor's Certificate No. 1611139; Byul. Izobret., No. 15, 223 (1997).
3. I. Viniar and S. Subo, Rev. Sci. Instrum. **68**, 1444 (1997).
4. S. Combs, Rev. Sci. Instrum. **64**, 1679 (1993).
5. M. Gouge, K. Onge, S. Minora, *et al.*, Fusion. Eng. Des. **19**, 53 (1992).
6. I. V. Viniar, B. V. Kuteev, S. V. Skoblikov, *et al.*, Zh. Tekh. Fiz. **65** (7), 167 (1995).
7. L. Towle, J. Phys. Chem. Solids **26**, 659 (1965).
8. H. Rahman, E. Ruden, Ströhmaier, *et al.*, Rev. Sci. Instrum. **67**, 3533 (1996).
9. L. Loitsyanskii, *Mechanics of Fluids and Gases* [in Russian] (Nauka, Moscow, 1987).
10. *Hydrogen: Properties, Production, Storage, Transport, and Applications*, Ed. by D. Yu. Gamburg and N. F. Dubovkin (Khimiya, Moscow, 1989).
11. P. Souers, *Hydrogen Properties for Fusion Energy* (California University Press, Berkeley, 1986).
12. D. Ryabinin and Yu. Lukach, *Screw Apparatuses for the Processing of Plastics and Rubber Blends* (Mashinostroenie, Moscow, 1965).
13. *Equipment for Plastic Processing Plants*, Ed. by V. K. Zavgorodniĭ (Khimiya, Leningrad, 1972).
14. I. V. Viniar, S. V. Skoblikov, and P. Yu. Koblents, Zh. Tekh. Fiz. **68** (5), 117 (1998).

Translated by P.P. Pozdeev

=== **EXPERIMENTAL INSTRUMENTS AND TECHNIQUES** ===

Optical Multiplexing of Microcavity Sensors Based on a Fiber Laser

V. D. Burkov*, F. A. Egorov**, Ya. V. Malkov*, and V. T. Potapov**

* Moscow State University of Forest, Mytishchi, Moscow oblast, 141001 Russia
** Institute of Radio Engineering and Electronics, Russian Academy of Sciences,
Fryazino, Moscow oblast, 141120 Russia
Received August 13, 1998

Abstract—A new principle of creating multiplex systems of active oscillator microcavity fiber-optic sensors is considered. The results of experimental studies of a multiplex system with two measuring channels are presented. © 2000 MAIK "Nauka/Interperiodica".

As demonstrated by Burkov *et al.* [1, 2], fiber-optic active oscillator systems employing the resonant interaction of radiation produced by an erbium fiber laser (EFL) with optically excited micromechanical cavities (*MC*s) provides an opportunity to create microcavity fiber-optic transducers (MFOTs) of physical quantities with frequency data encoding. The necessary condition of resonant interaction in an EFL–*MC* system can be represented as $f_{\text{rel}} \approx f/n$, where f_{rel} is the frequency of relaxation oscillations in the EFL; f is the natural frequency of the microcavity; and $n = 1, 2, 3, \ldots$. The EFL self-modulation frequency is given by $F \approx f/n$. Under these conditions, active oscillations usually arise within discrete areas in the space of parameters characterizing optical, acoustic, and thermoelastic properties of the EFL and the *MC*. Lasers with complex optical cavities may be characterized by sets of frequencies of relaxation oscillations corresponding to different mode groups [3]. Consequently, we can assume that resonance conditions, $f_{\text{rel}\,i} \approx f_i$, can be satisfied for each microcavity MC_i in a complex EFL optical cavity consisting of a multimirror system based on microcavities. This circumstance allows a multiplexed system of $MFOT_i$ to be created with the use of EFLs.

As is well known [4], the frequency of relaxation oscillations of EFL radiation intensity can be written as

$$f_{\text{rel}} \approx [(r-1)/\tau_{\text{ph}}\tau_M(1 + \sigma Nc\tau_{\text{ph}}l/L)]^{1/2}, \qquad (1)$$

where r is the relative pump level of the EFL active medium, τ is the photon lifetime in the EFL fiber-optic cavity with the length of the fiber equal to L and the length of the active fiber equal to l, τ_M is the lifetime of the metastable level of active particles, c is the speed of light in the optical fiber, and N and σ are the concentration and the cross section of radiative transitions between the working levels of active particles.

Note that relationship (1) was derived with an assumption that the active section of the fiber is uniformly pumped and the fiber cavity is isotropic. The photon lifetime in this case is $\tau_{\text{ph}} \approx 2L(\alpha - \ln R)$, where $\alpha - \ln R$ is the coefficient of nonresonant optical losses per single pass of the fiber cavity with allowance for the coefficients of losses and transmission coefficients of laser mirrors. As can be seen from expression (1), the quantity $f_{\text{rel}}(r, L)$ in the case of an active medium with given parameters and a cavity with fixed optical parameters can be controlled by varying the pump level r and the cavity length. We will consider the principle of the proposed multiplexing method by analyzing the operation of a two-channel measuring system (Fig. 1), which can be employed as a basis for differential MFOT schemes. In the case under consideration, the EFL cavity, including a single-mode fiber-optic *X*-coupler, is a complex (three-cavity) system consisting of sections MXM_1 and MXM_2, which represent partial cavities with
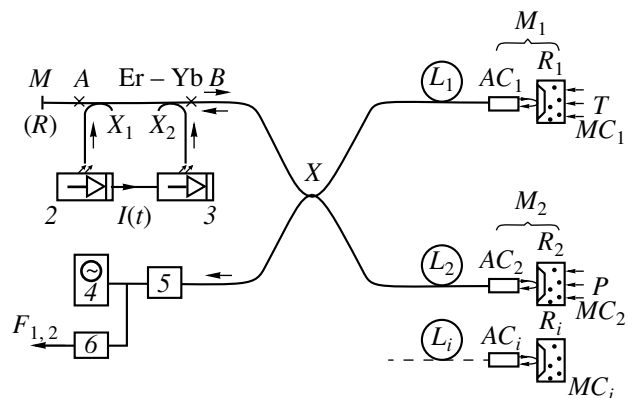
Fig. 1. Diagram of a two-channel multiplex system: (*1*) erbium fiber laser; (*2*) DL_1; (*3*) DL_2; $I(t)$, $DL_{1,2}$ injection current; X, single-mode fiber-optic coupler; L_i, segments of single-mode fibers; AC_i, fiber-optic autocollimators; MC_i, microcavities; (*4*) spectrum analyzer; (*5*) photodetector; (*6*) frequency meter; M_1, M_2, mirrors of the composite fiber-optic EFL cavity.

lengths $L_{1,2}$ and losses $\alpha_{1,2}$. As is well known [1, 2], resonant self-modulation may arise in different regimes of optical feedback (OF) between the EFL and the *MC*. For example, a Fabry–Perot interferometer and an auto-collimator (AC) can be used for this purpose. Consequently, both Fabry–Perot interferometers and autocol-limators with microcavities may serve as mirrors $M_{1,2}$ (with reflection coefficients $R_{1,2}$) in the considered scheme. Since MFOTs with autocollimators are characterized by a high stability with respect to various destabilizing factors and are very promising from the practical point of view, we will restrict our consideration below to MFOTs based on autocollimators. Reflecting surfaces of microcavities $MC_{1,2}$ characterized by natural frequencies $f_{1,2}$ play the role of mirrors $M_{1,2}$ in this case. When the sections $MXM_{1,2}$ have close optical parameters and microcavities $MC_{1,2}$ are characterized by close efficiencies of optical excitation, one can expect that self-excited oscillations should arise in the EFL–$MC_{1,2}$ system either at the frequency $F_1 \approx f_1$ or at the frequency $F_2 \approx f_2$ when the resonance conditions (1) $f_{\text{rel}1} \approx f_1$ or (2) $f_{\text{rel}2} \approx f_2$ are satisfied and when the relative EFL pump levels $r_{1,2}$ fall within the ranges corresponding to the discrete areas where self-excited oscillations exist, $r_1 \in O_1$ and $r_2 \in O_2$. Thus, we can establish a one-to-one correspondence between excitation states of microcavities $MC_{1,2}$ and the pump levels of the EFL in the regime of self-excited oscillations: self-excited oscillations arise in MC_1 when $r_1 \in O_1 = (r_1'$; $r_1'')$, and MC_2 is excited when $r_2 \in O_2 = (r_2'$; $r_2'')$. Figure 2 illustrates the evolution of the states of the considered active oscillator system corresponding to a gradual increase in the relative pump level $r(t)$. As the level r_1' is achieved at the moment of time t_1', self-excited oscillations with the frequency $F_1 \approx f_1$ arise in the system in the "soft" regime. These oscillations are quenched at the moment of time t_1'', when the level r_1'' is achieved. The system resides in the nonexcited state within the time interval $(t_1'' - t_2')$. Then, when the level r_2'' is achieved at the moment of time t_2', self-excited oscillations with the frequency $F_2 \approx f_2$ arise. These oscillations are quenched at the moment of time t_2'', when the pump level r_2'' is achieved. Next, the pump is decreased down to the initial level, and the above-described cycle can be repeated. Note that there are several factors limiting the scanning rate of the EFL pump level. First, with an abrupt increase in the EFL pump level with a duration of the leading edge of the pump pulse less than 10 ms, the dynamics of lasing in an EFL becomes rather complicated [5]. Second, time intervals $(t_1'$; $t_1'')$ and $(t_2'$; $t_2'')$ when the pump levels correspond to the excitation areas $O_1(O_2)$ should exceed the duration of the time interval required for the
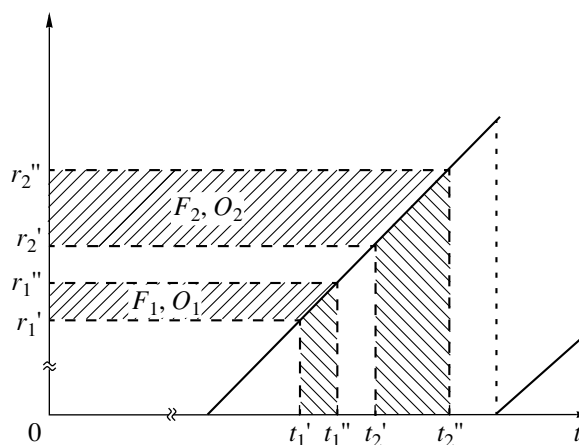

Fig. 2. Time dependence of the relative EFL pump level: $O_{1,2}$ show areas of self-excited oscillations with resonance frequencies of microcavity structures, $F_{1,2} \approx f_{1,2}$.

establishment of steady-state self-excited oscillations in the system. Consequently, the inequalities $(t_1'' - t_1') \gg Q_1/f_1$ and $(t_2'' - t_2') \gg Q_2/f_2$ should be satisfied for $MC_{1,2}$ with acoustic finesse values equal to $Q_{1,2}$. Third, measurement (averaging) time exceeding certain minimum time interval is required to ensure the adequate accuracy of measurements of oscillation frequencies. Our estimates show that, for microcavities with typical parameters $Q_{1,2} = 100$ and $f_{1,2} > 50$ kHz, the pump increase rate dr/dt is mainly limited by the first factor, which imposes the restriction $dr/dt \ll 10^2$ s^{-1}. Thus, scanning the pump level of a fiber laser, we can sequentially excite oscillations in the system and determine the frequencies of these oscillations, which permits the multiplexing of MFOTs.

The possibility of optical multiplexing of MFOTs was investigated with the use of the experimental setup shown in Fig. 1. We were able to vary the mean output radiation power of the EFL ($\lambda = 1.54$ μm) pumped with a diode laser (DL, $\lambda_p \approx 0.98$ μm) within the range of 0–40 mV. An *AB* segment of a single-mode fiber doped with erbium (Er^{+3}) and ytterbium (Yb^{+3}), where ytterbium played the role of an effective activator, served as an active medium in the EFL. The active medium was optically pumped by counterpropagating radiation fluxes of diode lasers $DL_{1,2}$ ($\lambda_p = 0.98$ μm), which were coupled into the active fiber with the use of fiber-optic couplers $X_{1,2}$. This symmetric pump scheme ensures a higher uniformity degree of pump radiation intensity distribution along the considered segment of the active fiber. The segment of the active fiber has a length $l = 5$ m and the concentration of erbium $N_{\text{Er}^{+3}} \approx 300$ ppm. The total length of the EFL fiber-optic cavity can be varied within a broad range by attaching segments of passive single-mode fibers to the ends of the active segment at points *A* and *B* by means of arc welding. Typi-
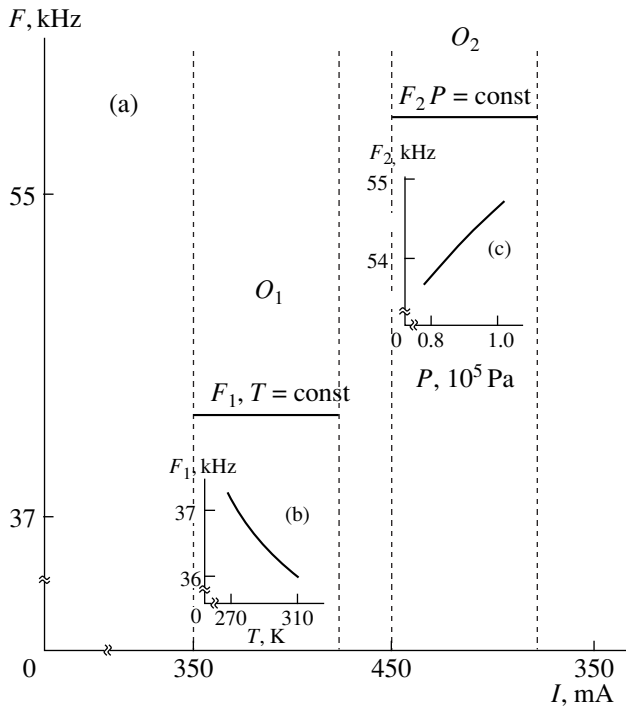
Fig. 3. (a) Dependece of the frequency of the signal at the output of the detector on the EFL injection current and transform functions for (b) temperature and (c) pressure.

cally, the segments of passive fibers have a numerical aperture $NA \approx 0.15$, the core diameter $d_c \approx 6.5$ μm, the diameter of quartz cladding $D \approx 125$ μm, the cutoff wavelength $\lambda_{cut} \approx 1.2$ μm, and the magnitude of losses at the working wavelength ($\lambda = 1.54$ μm) of no higher than 0.5 dB/km. The threshold injection current of $DL_{1, 2}$ was $I_{th} \approx 110$ mA. Lasing in the EFL was observed under these conditions with currents $I > I_g \approx$ 200 mA, and the dependence of the output optical power of $DL_{1, 2}$ on the injection current was close to a linear function. We were able to vary the current I within the range of 0–600 mA. The corresponding range of variation in the relative EFL pump level was estimated in accordance with the formula $r \approx (I - I_{th})/(I_g - I_{th})$ as 0–5. To measure the dependence $r(t)$ shown in Fig. 2, we defined the corresponding waveform of the $DL_{1, 2}$ injection current $I(t)$. In these experiments, we employed MFOTs with an autocollimator based on graded-index rod lenses, which shaped collimated Gaussian beams with an effective diameter of 400–500 μm and an angular divergence of $(2–3) \times 10^{-3}$ rad. The magnitude of losses in the symmetric X-coupler did not exceed 0.5. The length of the MXM_1 section was $L_1 = 12$ m, and the length of the MXM_2 section was varied from 14 to 30 m. Experiments were performed with microcavities of different topologies (micromembranes, microcantilevers, and microbridges) manufactured by means of anisotropic etching of single-crystal silicon with natural frequencies of

transverse oscillations $f = 20–170$ kHz and finesse values $Q = 50–200$ (in air) [2]. Thin films of different metals (Al, Ni, Ti, etc.) 20–300 nm thick coated on *MC* surfaces by means of magnetron deposition served as mirrors of composite microcavities. Angular and linear alignment devices were employed to ensure the required initial orientation of *MC*s with respect to the collimated beam. Note that self-excited oscillations with frequencies $F \approx f = 170$ kHz were observed in our experiments for EFL–*MC* systems with microcavities with a sufficiently large thickness, $h \approx 20$ μm, in which decay length for the temperature wave is $\delta = 13$ μm, which corresponds to the ratio $h/\delta \approx 1.6$. This finding confirms the high efficiency of the excitation of *MC* oscillations based on the resonant interaction with a fiber laser. The nonlinearity of the considered system may give rise to the interference of measuring channels due to the presence of ultra- and subharmonics of microcavity resonance frequencies in the spectrum of EFL intensity modulation. One can expect that, in the case of considerable detunings of $MC_{1, 2}$ natural frequencies meeting the condition $|f_2 - mf_1| \gg (f_2Q_2 + mf_1Q_1)$, this interference is insignificant. Figure 3 presents the results obtained in experiments with EFL–$MC_{1, 2}$ systems where the microcavities $MC_{1, 2}$ had a topology of a microbridge and a microbridge on a micromembrane with frequencies $f_1 \approx 37$ and $f_2 \approx$ 55 kHz. Dependences of the frequency of the signal at the output of a photodetector on the DL injection current shown in Fig. 3 were measured with scanning rates of the injection current not exceeding 1 A/s. For $I \in$ 350–420 mA, self-excited oscillations with a frequency $F_1 \approx f_1$ arise in the system under study ($L_1 = 12$ m, $L_2 =$ 14 m). These oscillations are quenched with the current $I = 420$ mA. Self-excited oscillations are not observed within the range of 420–450 mA. The further increase in the EFL pump dives rise to the establishment of self-excited oscillations with the MC_2 natural frequency $F_2 \approx f_2$. These oscillations are quenched with $I = 530$ A. Thus, in accordance with the scenario considered above, the scanning of the EFL pump sequentially leads to the excitation of oscillations with the natural frequencies of microcavities. We should emphasize that the signal-to-noise ratio at the output of a photodetector within the band of 100 kHz for the considered EFL–$MC_{1, 2}$ systems in the regime of self-excited oscillations usually ranges from 20 to 30 dB, which is less than the signal-to-noise ratio characteristic of EFL–*MC* systems with two-mirror cavities [1]. More detailed studies are required to reveal the mechanisms responsible for the lowering of noise. However, one of the factors that leads to the increase in the noise level is obviously associated with a strong interaction of mode groups in partial cavities through the active medium of an EFL.

A remarkable feature of the considered EFL–$MC_{1, 2}$ systems is associated with the fact that, when microcavities are subject to some external perturbations changing the natural frequencies of the microcavities

within some limited ranges $\Delta f_{1,2}$ ($|\Delta f_{1,2}/f_{1,2}| < 0.1$), relations $F_1 \approx f_1$ and $F_2 \approx f_2$ are satisfied for both independent action on $MC_{1,2}$ and simultaneous action keeping intact the transfer functions of individual measuring channels, $\Delta F_{1,2}/F_{1,2} \approx \Delta f_{1,2}/f_{1,2}$ (Fig. 3), with a relative level of frequency fluctuation $\langle \Delta F_{1,2}/F_{1,2} \rangle < 2 \times 10^{-4}$. This property of the considered systems allows us to propose them as a basis for the development of multiplex MC fiber detectors of sensors of physical quantities. Note that an important advantage of the proposed multiplexing method is that this technique makes it possible to increase the number of measuring channels in the system. For this purpose, one has to optimize parameters of the elements in the multiplex system (EFL, MC, autocollimating system, and single-mode fibers) in such a way as to ensure the maximum number of areas where self-excited oscillations of individual MCs may exist for a given range of acceptable changes in the EFL pump level.

REFERENCES

1. V. D. Burkov, F. A. Egorov, and V. T. Potapov, Pis'ma Zh. Tekh. Fiz. **23** (6), 33 (1997) [Tech. Phys. Lett. **23**, 224 (1997)].

2. V. D. Burkov, F. A. Egorov, Ya. V. Malkov, *et al.*, Radiotekhnika, No. 3, 36 (1998).

3. K. Otsuka, Proc. SPIE **2039**, 182 (1993).

4. I. R. Salcedo, I. M. Sousa, and V. V. Kusmin, Appl. Phys. B **62**, 83 (1996).

5. R. Rangel-Rojo and M. Mohebi, Opt. Commun. **137**, 98 (1997).

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Diagnostics of an H⁻ Ion Beam by Light Emission from the Drift Chamber

A. S. Artemov, G. F. Astrakharchik, Yu. K. Baĭgachev, and A. K. Gevorkov[†]

Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia
Received September 24, 1997; in final form, March 26, 1998

Abstract—The possibility of diagnosing an ion beam by light emission from the drift chamber is demonstrated using a 2-MeV H⁻ ion beam as an example. For a local gas puffing and negligible beam losses, spatial characteristics of the beam and the time behavior of the current pulse were monitored and the falling of a small number of ions onto the vacuum-chamber wall was recorded. A profilometer for recording the emission from individual layers of the observed region is described. © 2000 MAIK "Nauka/Interperiodica".

1. INTRODUCTION

Nonperturbative diagnostics of charged-particle beams often use either photons emitted from the components of a residual gas or a probing beam of a corpuscular target (see, e.g., [1–7]). From this point of view, of great interest is light emission, which provides the possibility of obtaining information on the beam parameters in inaccessible regions of the drift chamber or in the presence of external electromagnetic fields. The diagnostics are based on determining the beam profiles by the spatial distribution of the photon flux or the time behavior of light emission. In particular, when this distribution is recorded in several planes crossing the beam at different places along the beam, it is possible to reconstruct the transverse and longitudinal emittances of the beam [1–3, 5]. The maximum accuracy of such diagnostics is achieved when the recorded photons come from thin (about 5 μm in diameter) radiating filaments moving across the beam at a known velocity; these filaments barely impair the beam parameters in the course of measurements [6]. Depending on the ion energy, the diagnostics use the luminescence, transient, or Cherenkov radiation. The total number of the emitted photons substantially exceeds the number of other particles produced during the interaction of ions with the filament. In low-intensity beams, the efficiency of measurements with the use of radiating filaments and foils is low. At the same time, with high-power beams, problems of inadmissible ion losses and radiative heating and destruction of the probing targets arise. In this case, of special interest are less accurate methods based on the two-dimensional tomography of charged-particle beams with the use of light emission from the components of a residual gas [1–3, 5]. The two-dimensional density distribution of the ion current over the cross section is reconstructed by the inverse-problem method from several (usually 3 or 4) one-dimensional profiles

of the beam images measured at some different angles in the cross-sectional plane.

One of the main difficulties in using the above methods for diagnosing the H⁻ ion beam is the lack of published theoretical and experimental information on the mechanisms for emission from a gas target interacting with weakly bound negative ions. However, it is the mechanism for the target excitation that determines the reliability of the information on the beam parameters and provides the gas density required to diagnose the beam. In particular, the beam image can be a result of either the direct interaction of ions with the gas components in the vacuum chamber or the gas excitation by the secondary particles generated in the course of this interaction. For example, if the optical image of the beam is produced due to the target ionization by electrons (this process has a substantially higher excitation cross section than other processes), the photon flux is proportional to the gas pressure squared and the image is usually distorted by the spatial charge of ions.

In this paper, we experimentally study the method for nonperturbative diagnostics of a charged-particle beam by light emission from the drift chamber, using a 2-MeV H⁻ ion beam as an example.

2. THE NATURE OF LIGHT EMISSION FROM GAS TARGETS AND ITS USE TO MONITOR THE PARAMETERS OF AN H⁻ ION BEAM

In the course of experiments carried out in 1989 in Sukhumi at the Institute for Physics and Technology, the working gas (N_2, H_2, or He) was puffed locally for the purpose of monitoring. The pressure used was $<10^{-3}$ torr so that the ion loss for stripping was negligible.

N_2 molecules are one of the main components of the residual gas in the accelerator. In addition, the photon emission in the wavelength range corresponding to the


[†] Deceased.

first negative (1^-) band system of the N_2^+ ion (transition $B^2\Sigma_u^+ \longrightarrow X^2\Sigma_g^+$, $\lambda = 3580\text{–}5230$ Å, $\tau_\gamma = 63$ ns) is best suited for the diagnostics. This is associated with a relatively large radiation cross section of these bands for the interaction of N_2 with high-energy charged particles, short radiation time, and the fact that these bands fall within the sensitivity region of most types of photocathodes ($\lambda_p \approx 3000\text{–}7500$ Å) and are sufficiently well isolated from the slow components (1^+ and 2^- band systems of nitrogen, $\lambda = 5974\text{–}9183$ Å, $\tau_\gamma = 14$ µs), which have comparable radiation cross sections. Analysis of the background conditions shows that the radiation of CO molecules of the residual gas (transitions $B^2\Sigma^+ \longrightarrow A^2\Pi_i$, $\tau_\gamma = 52$ ns and $A^2\Pi_i \longrightarrow X^2\Sigma^+$, $\tau_\gamma = 4$ µs of the CO^+ ion) and H_2O molecules (transition $A^2\Sigma^+ \longrightarrow X^2\Pi$ of the OH radical, $\tau_\gamma = 0.8$ µs), whose excitation cross sections are close to that of the mentioned 1^- system, also falls in the wavelength range of this system. The relative intensity of this background radiation is governed by the partial pressures of CO, H_2O, and N_2.

Hydrogen and helium, which are also components of the residual gas, are of interest by themselves from the point of view of the formation of supersonic jet targets with a high density and good spatial localization. The wavelength region λ_p in question contains the following lines of these gases: the Balmer series of hydrogen ($\lambda = 3880$, 3970, 4100, 4340, 4861, and 6563 Å, $\tau_\gamma = 5\text{–}160$ ns), the radiation of the Fuhler molecular series and dissociative continuum of H_2 with the lifetime of radiative levels $\tau_\gamma = (3\text{–}4) \times 10^{-8}$ s, and intense helium lines ($\lambda = 3965$, 5016, 5876, and 6678 Å, $\tau_\gamma = 10^{-8}\text{–}10^{-7}$ s).

The schematic of the experiment is shown in Figs. 1a and 1b. An H⁻ ion beam from the outlet of the accelerator (*1*) with high-frequency quadruple focusing entered the drift chamber either directly or through a shaping aperture (*2*). The beam current was monitored by an inductive gauge (*3*) and a Faraday cup (*4*), in which the secondary emission was suppressed by the magnetic field. The current value was ≤30 mA. The working gas was supplied (up to the necessary pressure in the observation region) through a valve (*5*) of the measurement unit (*6*) with differential pumping on the ends. To provide a good uniformity of the gas density in the observation region, the gas-puff valve was located at some distance from this region, so that the pressure changed slowly with time. The differential-pumping facilities were located at a distance substantially exceeding the characteristic size of the measurement region. The operation of the gas-puff valve and arrangements for displacement and replacement of the shaping apertures was controlled remotely. The light emission was recorded by two interchangeable detectors (*7*, *8*) located along the horizontal (*y*) and vertical (*x*) directions. One of them (FEU-110) served to moni-
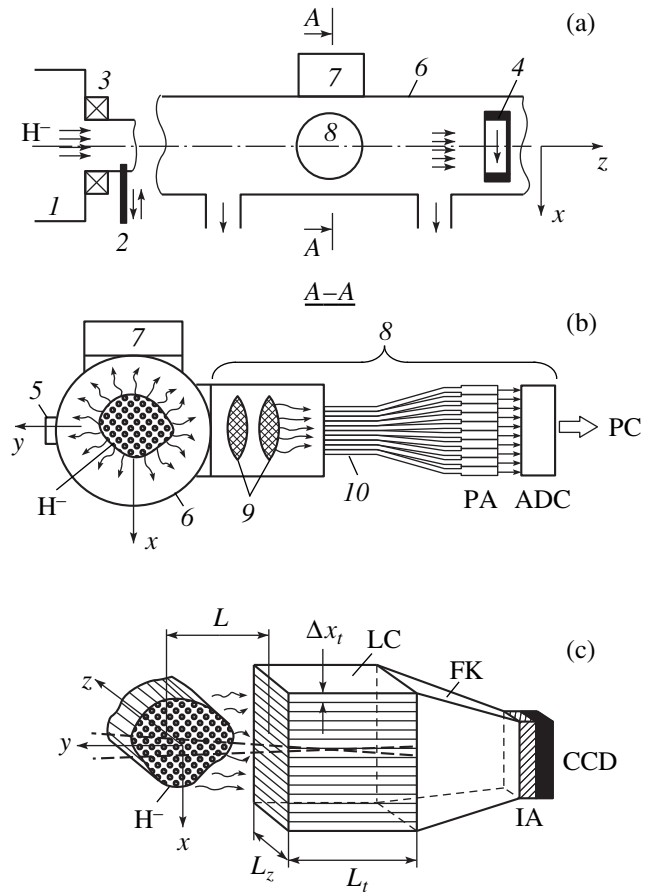
Fig. 1. (a, b) Schematic of experiments on the use of light emission from the drift chamber for diagnosing an H⁻ ion beam. (c) The general view of the profilometer based on the photon recording.
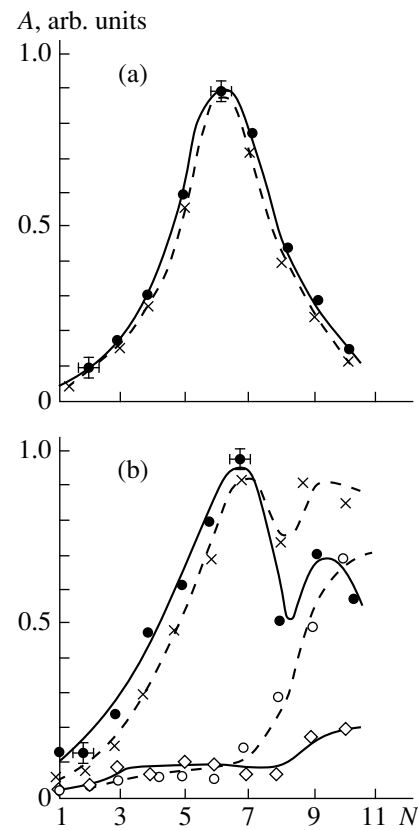
tor the time behavior of the photon flux, the other (*8*) measured the cross-section distribution of the flux along, e.g., the *x* coordinate. The image of the emitting region of the drift chamber was reduced by the optical system (*9*) and projected onto the input of a multichannel fiber array (*10*) oriented along the *x* coordinate, each channel being connected to an individual FEU-128 photomultiplier with a PA amplifier. At a given time instant, information on the pressure in the measurement unit and amplified signals from all of the radiation and ion-beam detectors (*3*, *4*) were led to a personal computer (PC) through analog-to-digital converters. The signals were digitized every one microsecond. The adjustment and calibration of the detector *8* was carried out using a luminous filament parallel to the *z*-axis and moving along the *x* and *y* coordinates. The spatial resolution measured near the center of the drift chamber with $\Delta Y = \pm2$-cm displacement was $\Delta X = 2\text{–}3$ mm.

Experiments showed the linear dependence of the ratio of the photomultiplier signal to the signal from the collector of H⁻ ions on the working-gas pressure in the interaction chamber over the entire pressure range
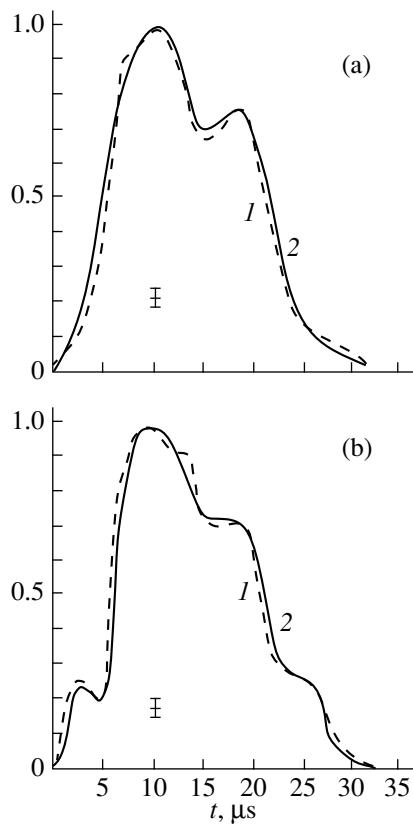
Fig. 2. Time dependences of (*1*) the H⁻ ion beam current signals and (*2*) the photomultiplier signals recorded by light emission from the drift chamber (a) at the background gas pressure $P_b = 3 \times 10^{-6}$ torr and (b) with an H_2 gas target at $P_H = 2 \times 10^{-4}$ torr. The signals are normalized to the amplitude.

Fig. 3. Distribution of the amplitudes of the photomultiplier signals over the recording channels (with the number *N*) during the passage of an H⁻ ion beam through the region of the drift chamber with a residual gas at the pressure $P_b = 3 \times 10^{-6}$ torr (open rhombuses and circles) and with a working gas (N_2) at the pressure $P_N = 3 \times 10^{-4}$ torr (closed circles and crosses) (a) without interaction with the wall and (b) when an insignificant part of the peripheral region of the beam falls onto the wall for $H = 0$ (solid curves) and $H = 200$ Oe (dashed curves).

(10^{-6}–10^{-3} torr). The shape of the photomultiplier signals corresponded well to the time behavior of the beam current (Fig. 2). A weak transverse magnetic field ($H \approx 100$ Oe) imposed in the observation direction had no effect on the luminous region of the working gas in the beam at different pressures (see, e.g., Fig. 3a); but, if a small portion of ions fell on the vacuum-chamber wall, this field essentially changed the intensity distribution throughout the channels of the fiber array (Fig. 3b). Special experiments with an ion beam of limited aperture that fell on the chamber wall showed that the spatial distribution and radiation intensity were independent of the working-gas pressure in the entire pressure range and were determined by the wall material and the value and sign of the applied magnetic field. With fixed magnetic field and residual-gas pressure, the subtraction of the background radiation from the corresponding emission distributions during the gas puffing allowed us to single out the region of the luminous gas in the beam and to obtain dependences similar to those

presented in Fig. 3a. Taking into account the spatial resolution ΔX of the detector channel, the characteristic cross section of the luminous region agrees well with that obtained by the direct measurement of the beam profile with the multilamella current receivers. Since we did not measure the gas density distribution in the measurement unit directly, such agreement points to the appropriate choice of the unit design and the existence of a fairly uniform target in the observation region. From the above results, a conclusion can be made that the optical image of an H⁻ ion beam obtained with the use of H_2, He, and N_2 gas targets in the pressure range under study is produced as a result of the direct interaction of the target atoms or molecules with the beam ions. Consequently, in the region with a sufficiently uniform gas density, this image can be associated with the corresponding spatial distribution of the beam current.

We present typical examples of the response of the detector *8* to the change in the beam characteristics. Figure 4a shows the distributions of the emission inten-

sity from the chamber for N_2 used as a working gas for different positions of a blind, which was displaced in the x-direction at the accelerator outlet and cut off some portion of the beam. It was found that a negligibly small portion of ions (too small to be recorded by the collector *4*) fell onto the upper wall of the vacuum chamber in the observation region of the drift channel. When a weak perpendicular magnetic field ($H = 100$ Oe) with one or the other polarity along the y-axis was applied to the drift region, the above effect was either completely eliminated or enhanced without the change in the recorded ion current. To simulate the shift of the ion flux in the transverse direction, we singled out different regions of the beam at the accelerator outlet with the use of a diaphragm. The corresponding distributions of radiation are presented in Fig. 4b for the residual (background) gas pressure $P_b = 3 \times 10^{-6}$ torr and nitrogen puffing up to the pressure $P_N = 8 \times 10^{-4}$ torr.

For all of the working gases used in experiment, we obtained results similar to those presented in Figs. 2–4. Both the absence of a lengthened trailing edge of N_2 radiation and experiments with the filter that cuts off the long-lived radiation components of 1^+ and 2^- band systems of nitrogen point to the small contribution of these components to the photomultiplier signal for the ion energy $E_i = 2$ MeV. The same experiments with the filter allow us to conclude that the contribution of H_α radiation from fast hydrogen atoms born as a result of stripping of a small portion of H⁻ ions by the gas target is small. Thus, fairly simple experiments showed that, for the local gas puffing and negligible beam losses, the photomultiplier signals and the distribution of their amplitudes over the channels of the fiber array give an adequate pattern of the time evolution of the H⁻ ion beam current, the shift of the beam in the transverse direction, and the redistribution of the current over the beam cross section. These signals also permit a high sensitivity of the recording of an insignificant fraction of ions falling onto the wall of the drift chamber. The results obtained show that the light emission can be used for the on-line monitoring of the ion loss and optimal matching of the individual sections of the accelerator where the beam energy is still small.

3. BEAM PROFILOMETRY BY LIGHT EMISSION FROM THE DRIFT CHAMBER

The upper boundary of the ion energy measured by light emission from the beam is determined by the beam current, the density of the gas target, and the sensitivity of the detection system. The number of photons producing the optical image of the beam on the detector surface for the time Δt is estimated as

$$N_\gamma \approx 2 \times 10^{35} I_i \sigma_\gamma P_b \tau_i f_i \Delta t L_\gamma \Omega_\gamma \eta_t / 4\pi, \qquad (1)$$

where I_i is the beam current [A], τ_i and f are the duration [s] and the frequency [Hz] of pulses, σ_γ is the radi-
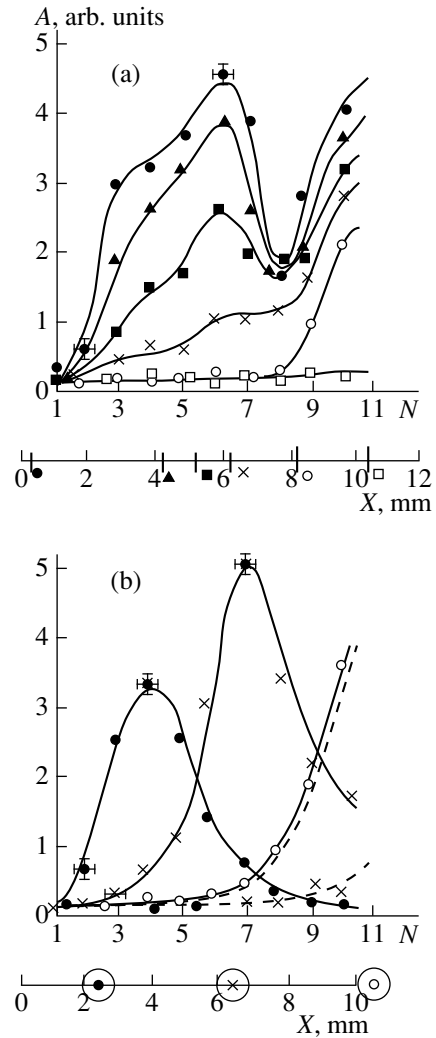
Fig. 4. Distribution of light-emission intensity over the channels of the fiber array (N) for different positions of (a) the blind and (b) the aperture along the x-axis at the accelerator outlet for $H = 0$ at $P_b = 3 \times 10^{-6}$ torr (dashed curves) and at $P_N =$ (a) 10^{-4} and (b) 8×10^{-4} torr (solid curves).

ation cross section of photons in the measured spectral range [cm²], P_b is the background gas pressure in the interval in question [torr], L_γ is the effective length of the observed interaction region [cm], Ω_γ is the average solid angle of light collection, and η_t is the transmittance of the optical system.

A relatively high light-collection power and high spatial resolution permit the profilometer to record photons from individual layers of the beam (Fig. 1c). The necessary photon separation is achieved with the use of a multilayer collimator (LC) consisting of thin transparent plates with rough absorbing surfaces. The filtered photon flux passes through the collimator, is transported by a focone (FK) to the image amplifier (IA), and is recorded by a charge-coupled-device (CCD) array. Spatial resolution of the profilometer at

the distance L from the collimator input is evaluated as

$$\Delta X \approx \Delta X_t(1 + 2Ln_t/L_t), \qquad (2)$$

where ΔX_t, n_t, and L_t are the thickness of the plate, its refractive index, and the length along the y-axis, respectively.

If the plate width along the z-axis is L_z, we obtain $L_\gamma = L_z[1 + 2n_tL\sqrt{L_t^2 - L_z^2(n_t^2 - 1)}$. In the general case, the average solid angle of the light collection from the observed beam layer is described by a cumbersome expression inconvenient to analyze. To make estimates, we can use the lower bound for L_γ and Ω_γ, which correspond to $n_t = 1$:

$$\Omega_{\gamma|min} = \Omega_\gamma(n_t = 1)$$

$$= \frac{2\Delta X_t}{A_2(L + L_t)}[A_1^{-1} - A_2^{-2}(A_1^{-1} - L_z/2(L + L_t)A_3) \qquad (3)$$

$$- 4L(1 + L/L_t)L_z^{-1}A_2^{-2}(A_3^{-1} - A_1^{-1}L_t/L_z)],$$

where $A_1 = \sqrt{1 + (L_t/L_z)^2}$, $A_2 = 1 + 2L/L_t$, and $A_3 = \sqrt{1 + [L_z/2(L + L_t)]^2}$.

For $L_z = 5$ cm, $L = L_t = 10$ cm, $n_t = 1.4$, and $\Delta X = 1$ mm, we obtain $\Delta X_t \approx 250$ μm, $L_{\gamma|min} \approx 15$ cm, and $\Omega_{\gamma|min} \approx 2 \times 10^{-4}$ sr. Contemporary detectors allow high-accuracy measurements of the one-dimensional spatial distribution of photons for a number down to $N_\gamma \approx 10^4$. In this case, for $\eta_t = 0.8$, $P_b = 5 \times 10^{-7}$ torr, and the ion beam with an average current of $\bar{I}_i = 5 \times 10^{-4}$ A, the potentialities of the profilometer in question are evaluated by relation (1) as $\sigma_\gamma\Delta t \approx 10^{-18}$ cm^2 s. We will assume that, in the spectral range of 1^- band system of N_2^+ nitrogen ions, for nonrelativistic H$^-$ ions with energy E_i [keV], the dependence $\sigma_\gamma \approx 6.4 \times 10^{-15}E_i^{-1}\ln(4.7 \times 10^{-2}E_i)$ (cm^2) is valid. Then the nonperturbative measurements of the beam profile during one second is possible up to the energy $E_i \approx 50$ MeV. The expression used for σ_γ is based on the averaged dependence for the protons with energy $E_i = 100$ keV–1 MeV (this dependence is obtained by taking the data from different experiments [8] and extending them to greater energies) and on the measured cross-section ratio $\sigma_\gamma(H^-)/\sigma_\gamma(H^+) \approx 3$ for $E_i = 100$ keV–2 MeV [9].

The estimate presented for the upper bound on the H$^-$ ion energy is somewhat speculative and changes with the use of different functional dependences for σ_γ at $E_i > 2$ MeV. A certain arbitrariness is also associated with a large scatter in the known data on the proton cross sections; the scatter in data is approximately one-and-a-half, because different procedures were used for calibrating the radiation detectors in experiments. It should be also noted that the ratio $\sigma_\gamma(H^-)/\sigma_\gamma(H^+) = 1.30 \pm 0.26$ obtained in [10] for $E_i = 1$ MeV differs from that used in our paper. In our opinion, the ratio obtained in [10] is unreliable, because the measurement procedure used there leads to the dependence of this ratio and the photon cross sections on the nitrogen pressure when the nitrogen interacts with H$^-$ ions and H^0 atoms. The stripping cross sections for these particles, which are presented in that paper, differ substantially from the well-known cross sections used in the literature.

REFERENCES

1. J. S. Fraser, IEEE Trans. Nucl. Sci. **26**, 1641 (1979).
2. J. S. Fraser, IEEE Trans. Nucl. Sci. **28**, 2137 (1981).
3. D. D. Chamberlin, G. N. Minerbo, *et al.*, IEEE Trans. Nucl. Sci. **28**, 2347 (1981).
4. J. Bosser, J. Mann, *et al.*, Nucl. Instr. Methods A. **238**, 45 (1985).
5. V. I. Artemov, S. A. Dobromirov *et al.*, Prib. Tekh. Éksp., No. 4, 42 (1989).
6. V. P. Novikov, E. V. Serga, and A. V. Kharlamov, in *Proceedings of the Second European Conference on Particle Acceleration* (1990), Vol. 1, p. 765.
7. G. G. Gul'bekyan, V. N. Mel'nikov, and O. Yu. Tsybin, Report No. P9-91-558, OIYaI (Joint Inst. for Nuclear Research, Dubna, Russia, 1991).
8. E. W. Thomas, *Excitation in Heavy Particle Collisions* (Willey, New York, 1972), p. 436.
9. A. S. Artemov, L. V. Arinin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **53**, 533 (1991) [JETP Lett. **53**, 558 (1991)].
10. R. F. Holland, D. D. Cobb, *et al.*, Phys. Rev. A. **41**, 2429 (1990).

Translated by A.D. Smirnova

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Self-Assembly of Cluster Protein Films
(Allotropic Nonequilibrium Noncrystalline Modification)
during the Process or Their Condensation

E. G. Rapis

Sourasky Medical Center, Tel-Aviv University, Tel-Aviv, 64239 Israel
Received December 30, 1998

Abstract—A new nonequilibrium allotropic protein modification (for brevity, hereafter called a *protos*) obtained by *in vitro* self-assembly of cluster protein films and identified with *in vivo* protein *folding*[1] has been experimentally observed and described in detail [1–7]. *In vitro* visualization of the dynamics of protein condensation in an open protein–water system under the nonequilibrium conditions (at a relatively fast water evaporation) both on the micro- and macrolevels provided the establishment of the chaotic nonlinear dynamics of the process. The protein film has no long-range order and is characterized by a morphologically new highly structurized order at the nano- and macrolevels. A number of new structural, electric, and magnetic properties of a protein considerably different from the properties observed for crystalline proteins formed under the equilibrium conditions have been established. © *2000 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

In the recent decade, we witness the rapid development of a new direction in science—self-assembly of molecular cluster films under the nonequilibrium thermodynamic conditions of matter condensation [8, 9]. It should be emphasized that such films acquire the same characteristic properties irrespectively of their composition (organic or inorganic) [10]. This process is interpreted in terms of the concept of biological self-organization or the so-called biomimetics. On the other hand, the still unsolved key problem of protein folding [11] underlying this concept seems to be aside of this new direction and the nonequilibrium dynamics itself, although the importance of this theoretical concept for establishing the folding mechanism is quite clear, so that the problem should be solved within this field of modern physics (the science of self-assembly of thin films) [12, 13]. At the same time, the functioning of three-dimensional protein structures is still studied in biology solely under the equilibrium conditions providing protein crystallization [14] despite the fact that the theory of self-assembly (i.e., nonequilibrium thermodynamics) requires the creation of strongly nonequilibrium conditions [15–17]. This is also true for self-assembly of cluster films [8, 9]. Therefore, it is very important to perform *in vitro* experiments on growth of protein films in the simplest protein–water systems under nonequilibrium conditions, i.e., conditions close to those set by the theory of self-assembly. Also, this would allow to compare self-assembly of protein films with self-assembly of cluster films of other materials, which is more important in view of the fact that all living biological structures can exist only under nonequilibrium conditions [15–17]. The morphology of the films (plates) of living organisms is an essential characteristic [8, 9, 12], e.g., for elastic vessel and mitochondrial membranes.

EXPERIMENTAL

In this work, we visually studied the dynamics of protein condensation and phase transitions in an open protein–water system under nonequilibrium conditions, i.e., far from thermodynamic equilibrium. We used proteins of different chemical compositions such as albumin, globulin, hemoglobin, cytochrome, lisozyme, crystallin, and some other human and animal proteins.

In the experiments on solid substrates (glass, plastic, wood, and iron) performed under atmospheric pressure at room temperature, water readily evaporated from the solid substrates, thus facilitating protein precipitation. Altogether, 8500 experiments were performed by various scientists. The process was visualized by various micro- and macroscopic techniques including the optical and polarization methods, scanning electron microscopy (a JEOL electron microscope, confocal scanning laser, and electron microscopes). The specimens were also subjected to X-ray diffraction study [7]. We determined some other characteristics of the specimens such as their magnetic sensitivity [6], current–voltage characteristics of liquid and solid proteins under nonequilibrium conditions, and the ability of

---

[1] Protein "folding" usually means spontaneous self-organization, where three-dimensional spiral structures of operating biologic machines arise from peptides inoperable in solutions.

Fig. 1. (*1*, *2*) Macrolevel (magnification 20×10); (*3*, *4*) microlevel (a JEOL scanning electron microscope): (*3*) magnification 1×100.38 μm, (*4*) magnification 2×430.38 μm.

proteins toward laser fluorescence (in a confocal microscope) [7].

RESULTS

Under qualitatively close experimental conditions, the reproducibility of the results was very high. The experiments allowed us to follow the dynamics of pro-tein condensation. During the experiments, we observed the continuous motion of the front of self-maintained oscillations (auto-oscillations) with the alternating zones of different protein densities [17]. Experimentally, this manifested itself in the optical phenomenon of the formation of three-color interfer-ence bands (rings) reflecting the existence of the zones of different protein densities corresponding to auto-
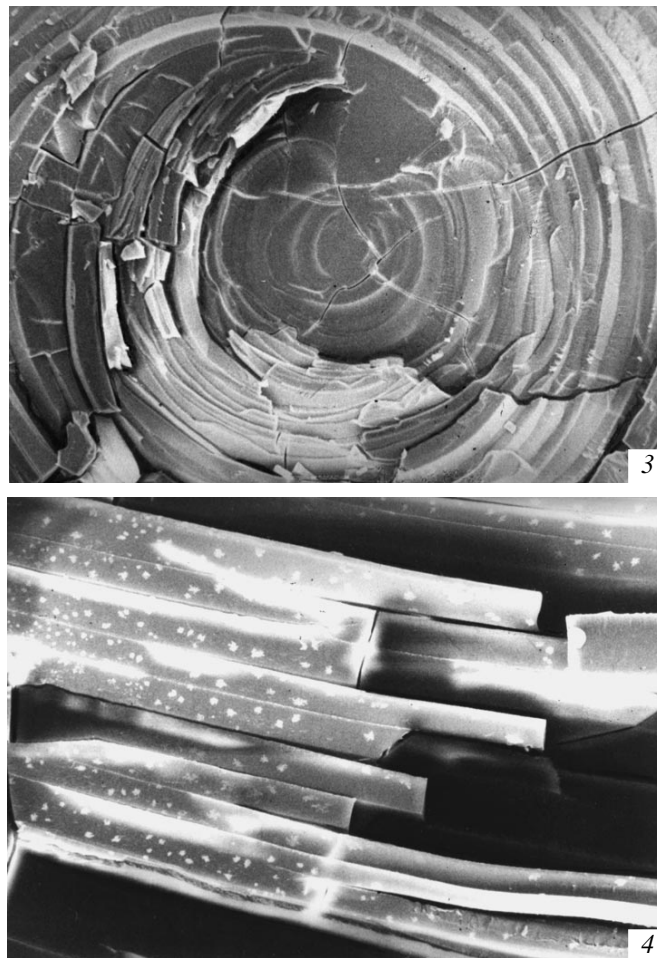
Fig. 1. (Contd.)

oscillations. The liquid-crystal phase of the protein changed color (red, yellow, green) in a jumpwise manner similar to the well-known oscillating changes in films or the variations in the elasticity characteristics of biopolymer gels [18, 19]. At the critical protein density, we observed the formation of structural defects caused by nucleation phenomena. In turn, this effect resulted in the formation of spontaneous radial, helical, and chiral symmetry and asymmetry in three-dimensional cluster protein films. In this case, the morphological cellular structures or domains with nuclei (one nucleus per domain) were formed. These observations indicate spontaneous nucleation in the film (Fig. 1). Around the nucleus, thin three-dimensional rolling films are formed with helical defects (similar to grain boundaries) in between, and also some vortices. Later, some of the nuclei are divided into smaller ones [20]. It is seen from Fig. 1, (*1*) and (*2*) that a nucleus of such a vortex is "cut" by linear defects into two or four smaller "daughter vortices" *B*, which, in turn, consist of even smaller rolled soliton-type films (*2*, *G*). One can also distiguish the regions where individual fragments are

linked (indicated by arrows). At the macrolevel, it is seen that these vortices have stepped "faces" built by repeating films. Similar patterns are also observed at the microlevel [7]. The corresponding scanning electron micrographs revealed films in the shape of parallelepipeds with heights of 10–20 μm with pronounced periodicity and oriented layers forming vortices (Fig. 1, (*3*) the view from above). These formations are reminiscent of macrolevels (the specimen wholeness is partly disrupted). One can also see rolled films, fine striation, and a relief with the nucleus in the center. The side view (the cleavage) of the same specimen is shown in Fig. 1, (*4*). One can see piles of similar films cut into equal parts by dislocation-type defects and smooth steps with islandlike aggregates in the bulk (accumulations of white dots and "snowflakes").

Morphologically, we managed to identify some blocks (multilayer piles of thin films of various dimensions) with smooth edges in certain end zones. As a rule, the external parts of the films were cut by treelike defects forming new structures having similar shapes but on a smaller scale, which seems to reflect the fractal
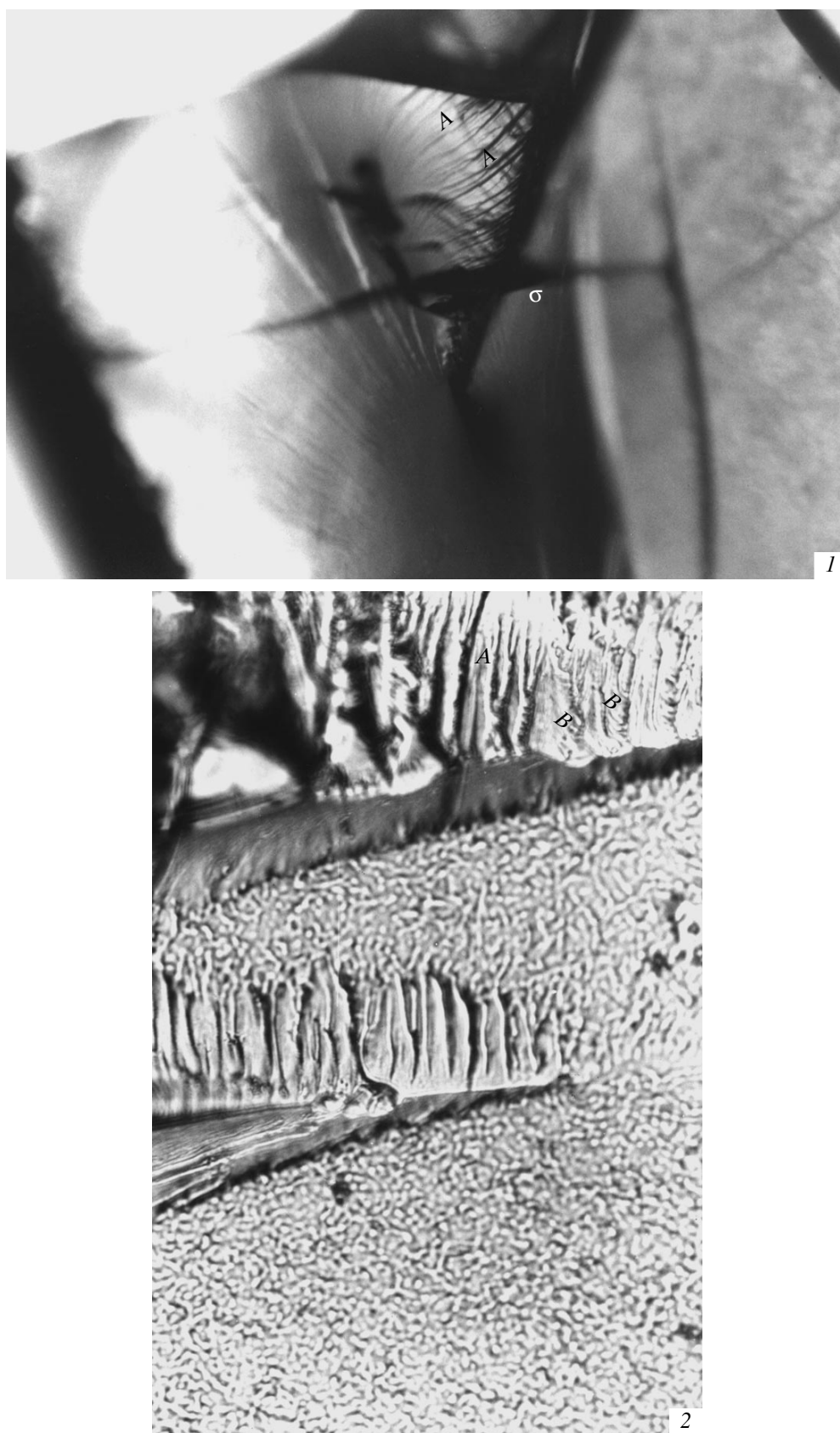
Fig. 2. (*1–3*) Macrolevel (magnification 7 × 40), (*4*) (magnification 10 × 40), (*5*) microlevel (magnification 1 × 1000.36 μm).
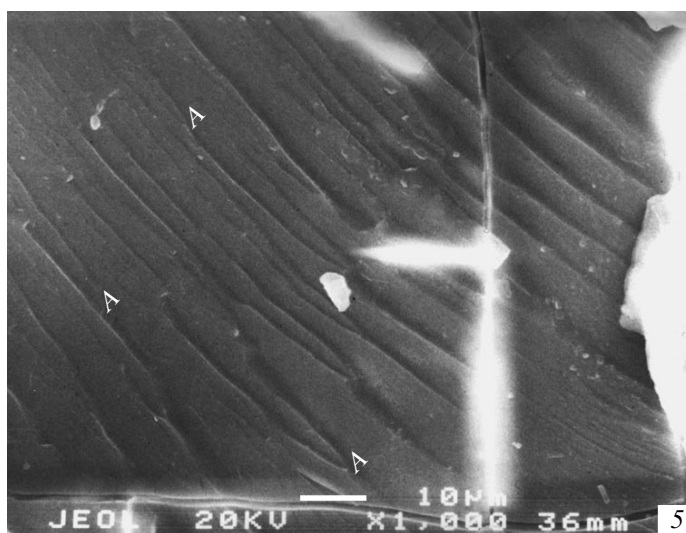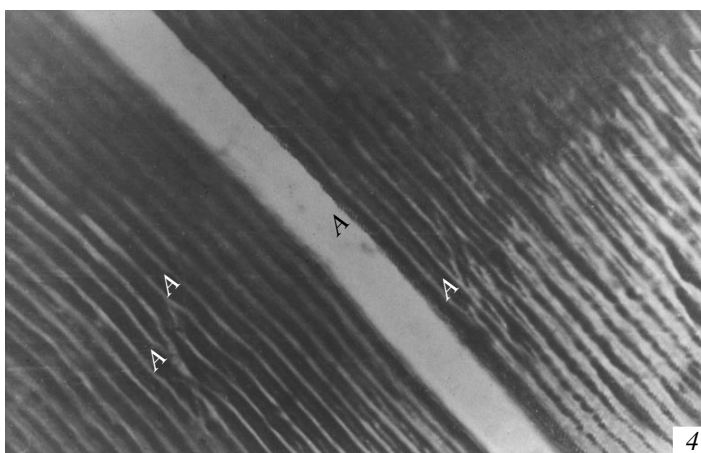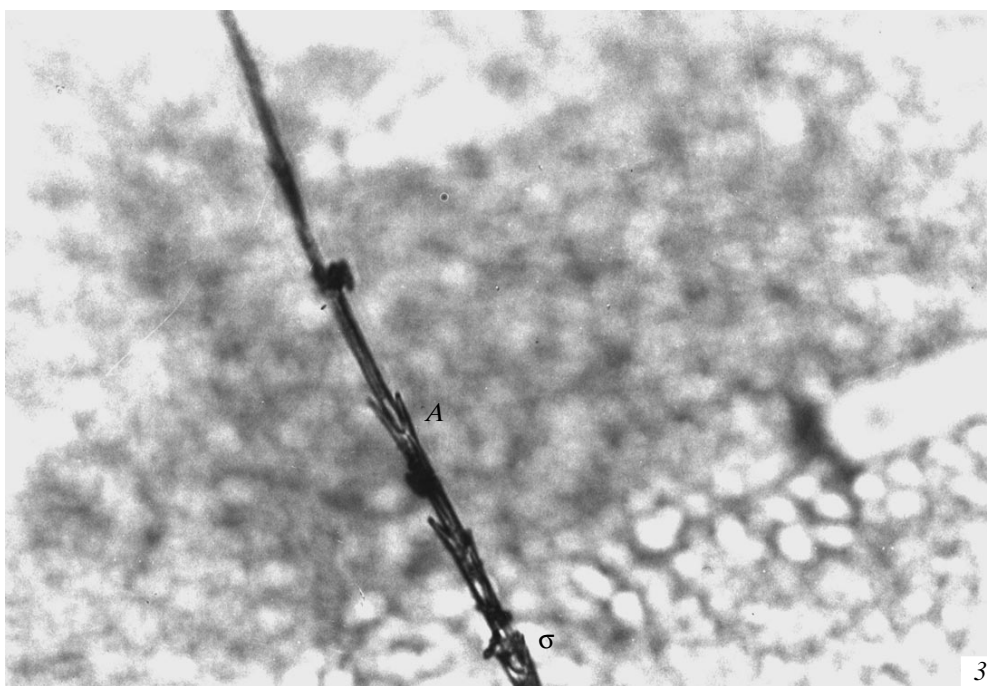
Fig. 2. (Contd.)
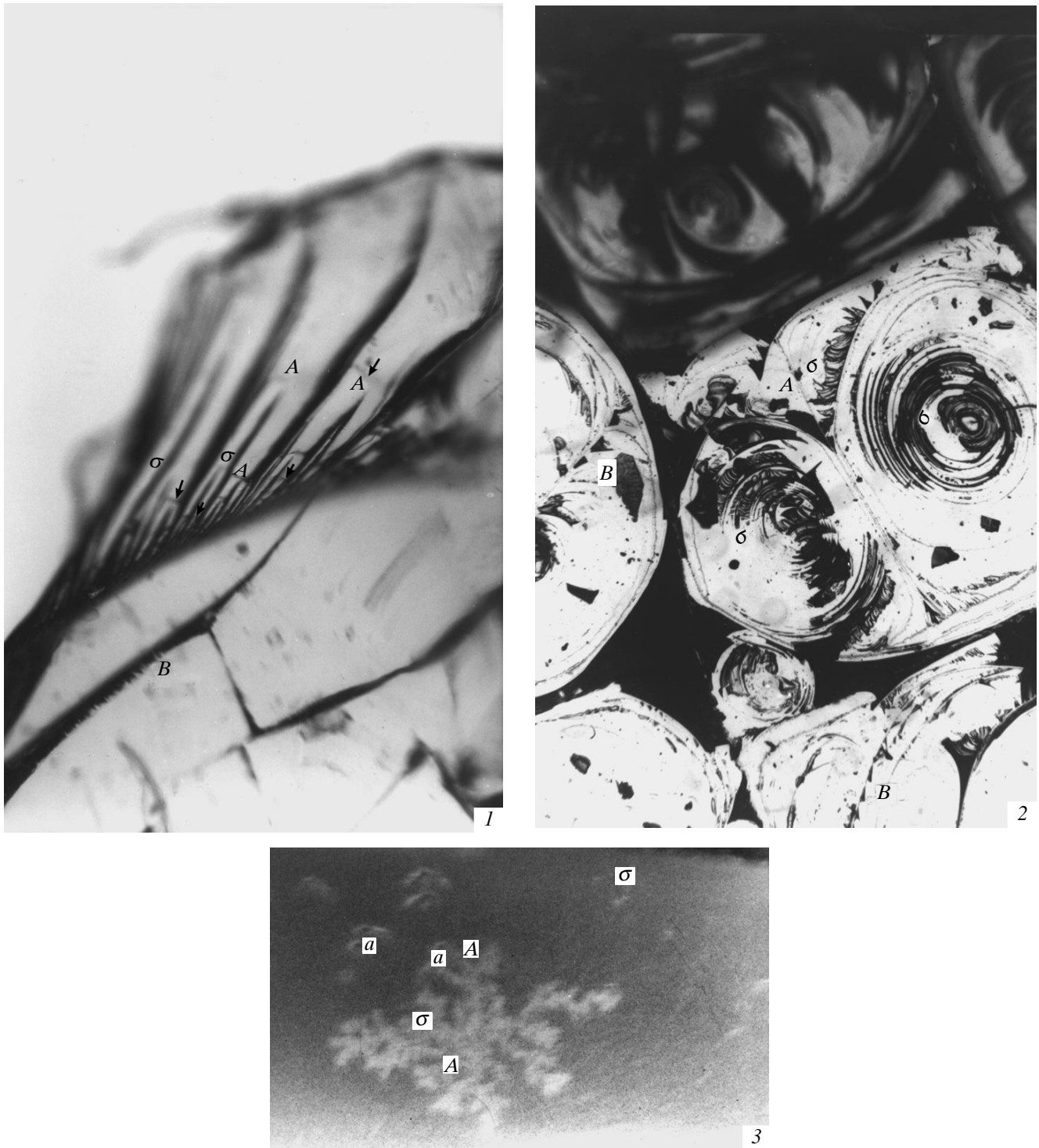
Fig. 3. (*1*, *2*) Macrolevel (magnification 8 × 40), (*3*) microlevel (magnification ×10000.38 mm).

properties of the specimen. At both macro- and microlevels, we usually observed a specific "discontinuous symmetry" (which received the name "porcupine symmetry") with a regularly decreasing "conical gaps" (Fig. 2). Both at the macro- (Fig. 2, (*1–4*)) and the microlevels (Fig. 2, (*5*)), one observes rolling treelike films and dichotomic division (*A*), the zones of defects passing through all the film layers and separating the attraction zones. The alternating zones form structures characterized by helical (σ), chiral (*A*), and porcupine

(*B*) symmetries. There are also defects with dimensions decreasing in a jumpwise manner along the external edge of the film; at a certain angle, the films show transverse striation resembling fingerprints (*B*). All these features are observed both in the *in vitro* and *in vivo* experiments (*B*) (a slice of a native crystalline lens).

Another characteristic feature of these films is their strong adhesion to solid substrates. Figure 3, (*1* and *2*) shows small (10 nm) details and the helical, chiral (*A*), and porcupine (*B*) symmetries, with the latter being interpreted as an energy gap from +2 to 0). In Fig. 3, (*2*) these gaps are indicated by arrows. In this case as well, the adhesion between the protein film and the solid substrate is very strong: it is seen that the larger part of the film surface is cleaved, whereas the layer directly adjacent to the substrate is preserved with all its numerous fine details. Figure 3, (*1*) illustrates the fractal properties of the structures. It is seen that the elements located far from the structure base are of smaller dimensions, and the structure reminds a skeleton of an extremity. This is especially well seen in the zones of the linkage of fractal elements, which are similar to the self-complementing key-to-lock-type linkage well known for fragments of biological objects. (The latter structures were obtained by condensation of protein films onto nylon substrates). Figure 3, (*3*) shows an electron micrograph obtained in a scanning electron microscope, which demonstrates the fractal properties of the protein on a microlevel—self-similarity, the reduction of the volumes of the aggregate complexes during their division depending on the distance from the structure base (the accumulations of "snowflakes" at the side edges of the films). At the nanolevel, one can also see vortices with (a) ~10 nm-large nuclei in the centers.

The supramolecular structure of the films showed considerable birefringence in polarized light. The bright green luminescence spontaneously appearing in the light beam in a confocal laser scanning microscope [7] indicates the high optical activity of the material. In denser materials, one observes sudden (avalanche-like) formation of similar cone-shaped films that are dichotomically divided with the formation of treelike branching structures (Fig. 4). These structures synchronously growing in the helical direction are similar to solitons. The self-similarity of these structures, the changes in the dimensions of some details (from larger to smaller and *vice versa*) accompanied by the formation of cone-shaped patterns with the variations in the volumes of the individual elements depending on their distance from the base indicate that such structures can have fractal properties. We also observed the formation of three-dimensional terracelike, tubular, branching treelike, and cone-shaped chiral structures and also tubes in tubes and helices in helices with the clearly pronounced asymmetry. The pairs of such structures form vortices twisted in the opposite directions with linear defects in between. In each of these vortices, the branches twisted in opposite directions are linked. Figure 4 (*1*) shows cone-shaped (peaked) fractal films of a continuously

decreasing size (*A*), helical and cone-shaped films resembling helical cockles (*B*). Figure 4 (*2*) shows thin treelike films with synchronously and coherently growing dissipative fractal structures (*a*, *b*, *c*). The most numerous and characteristic micro- and macrostructures were spatially divided double-tubular helical blocks forming straight anisotropic rods with a clearly oriented conical top (Fig. 4, (*3*)) similar to well-known diblocks observed in polymers. Such a diblock forming a cone is shown in Fig. 4 (*3*). One can clearly see the thickening linkage regions (*A*), a rolling helical film (σ), and their discrete and dichotomic character (*A, 1*).

It should be emphasized that the structures observed in *in vitro* experiments, their architecture, and behavior were repeated in all the details at both macro- and mesoscopic levels also in the *in vivo* experiments on living biological systems and proteins (Fig. 2, (*4*)) [7]. A clear example here is the well-known property of self-complementarity in the key-to-lock protein linkage usually observed in *in vitro* experiments on living organisms in the condensed phase (Fig. 3, (*1*)). We also managed to observe the dynamics of self-assembly: small dots forming large agglomerates of supramolecular forestlike structures (Fig. 4, (*2*)) and other structures that could also be observed at the macrolevel.

It is seen that these observations confirm our concept of the formation of nanostructures with the properties dependent on their dimensions. The specific features of these structures are observed only at the nanolevel because of the formation of new weak chemical bonds. These new structures and their properties are preserved up to the attainment of the "functioning macrolevel" [9, 13, 22, 23]. This also explains the formation of a new nanoenergetic level (nanomagnetism) [9, 13] determining the functioning of protein "macromachines."

It should also be emphasized that the recent examination of the experimental and theoretical data on the formation of nanostructures in the films of some other materials allowed one to predict the properties of protein nanostructures [22]. These theoretically predicted properties—three-dimensionality, strong adhesion between the protein film and the substrate, shape formation, etc.—completely coincide with the properties of the above structures (Figs. 1–4).

The experimentally observed similarity of mesoscopic (10–100 Å) and macrodimensional structures and the coincidence of theoretically predicted and experimentally observed data give grounds to believe that condensation really takes place during self-assembly of cluster films and reflects the property of protein folding occurring from the nanolevel up to the functioning macrolevel. Therefore, we can state that the experimentally observed self-assembled film structures should be considered as dissipative protein nanostructures formed during protein condensation and folding in the nonequilibrium systems.
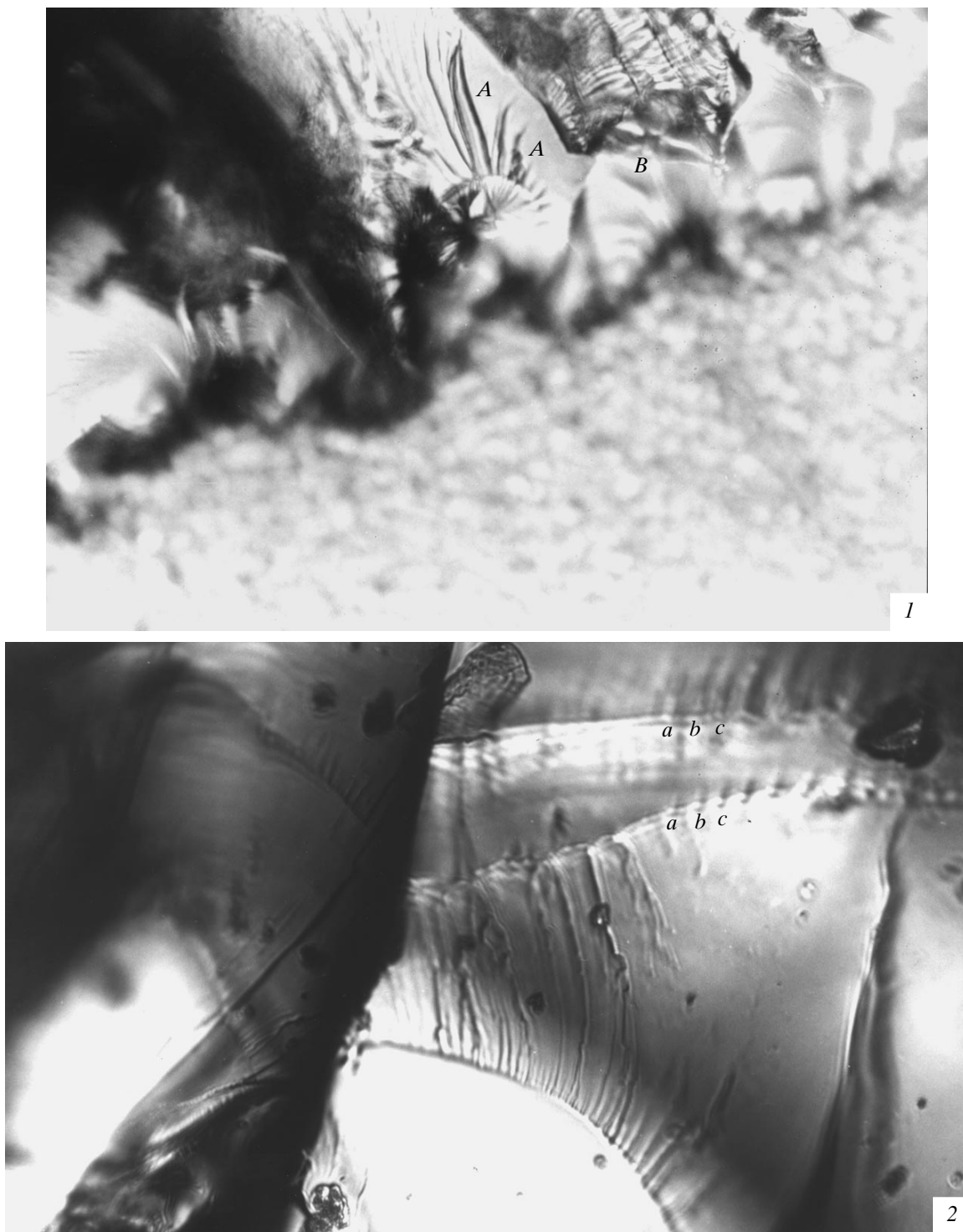
Fig. 4. Micrographs obtained on a microscope with the violet filter. (*1*, *3*) (magnification 8 × 40), (*2*) (magnification 7 × 20).

It was also established that a solid protein film has a mixed structure—mainly amorphous with some crystalline islands. The X-ray diffraction analysis showed that this structure is ordered only at levels of 10.5 and 4.33 Å. In polarized light, protein films show considerable anisotropy (birefringence in crossed Nicol prisms [1, 2]) and bright green fluorescence in laser light (in a confocal laser microscope [7]). The current–voltage characteristics of protein films are similar to those of high-resistance semiconductors [7].
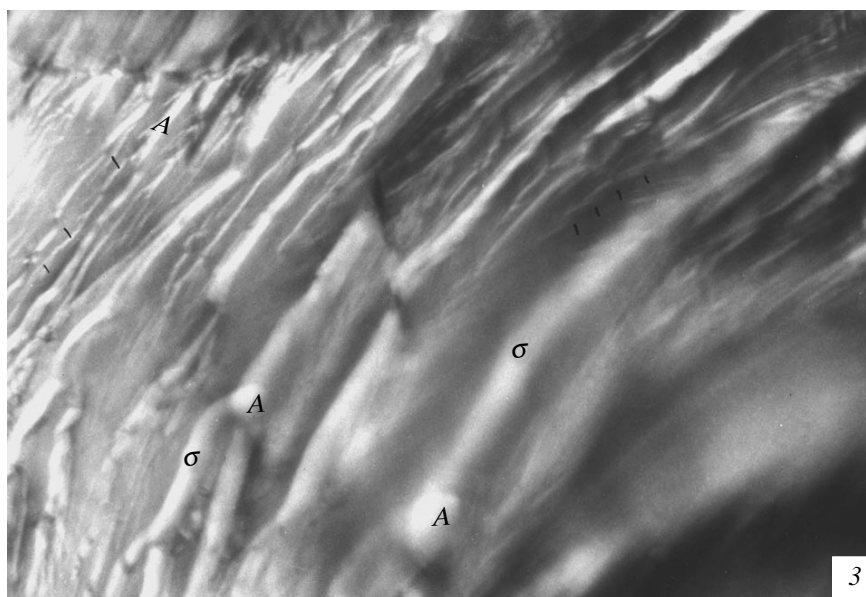
Fig. 4. (Contd.)

DISCUSSION

Our experiments showed that a protein film with the above properties can be formed solely under nonequilibrium conditions. For comparison, we also condensed proteins under more equilibrium conditions—in a closed system with a slow removal of the solvent. Under a microscope (at the macrolevel), the thus formed crystalline protein material looked either like a defect-free monolith or like a material with single defects.

Our experiments showed that protein condensation in a closed system yields quite different results: the protein net thus formed showed birefringence (Figs. 5 and 6 in [4]).

Different properties of solid proteins formed under equilibrium and nonequilibrium conditions can be interpreted within the framework of the classical physicochemical theories of protein and gel formation [24] or the theories of colloidal suspensions in the phase of anisotropic nematic liquid crystals [25] and elastic films [13, 21 26, 27]. These theories consider a protein in the liquid phase as a colloidal suspension, because the dimensions of protein macromolecules are always comparable to the dimensions of particles forming the colloidal suspension. Such a suspension can have a phase of a nematic liquid crystal only under nonequilibrium condensation conditions [25]. The nematic liquid phase of the protein is always characterized by formation of gels and disperse structures possessing various interesting properties [24]. It is shown that there exist only two types of solidification in colloidal systems. Within the framework of the above theories, it was established that, under conditions of the slow removal of the solvent or its preservation in the solution

(i.e., under conditions close to equilibrium), the so-called condensed–crystallized solid phase of the protein is formed (the Izmailova–Rebinder classification [24]). This phase has a rigid gel-like framework with disperse particles being strongly bound to one another and thus forming a continuous protein layer whose X-ray diffraction pattern clearly indicates the formation of the molecular lattice [28]. Thus, in this case, an equilibrium protein crystal structure is formed under more equilibrium conditions.

On the other hand, it was established that the fast removal of the solvent provides the formation of nonequilibrium conditions leading, in turn, to the formation of a heterophase system. In this case, disperse structures of the second type with quite different properties are formed. Moreover, recent studies showed that the colloidal suspension of an anisotropic nematic liquid crystals is, in fact, a metastable colloidal liquid phase whose stabilization is determined by the repulsive forces and defect formation [25, 29]. It is shown that these properties are closely related to the supramolecular process of aggregation of charged particles with the formation of dipoles similar to magnetic ones on the nanometer scale [13]. In a solid protein, these aggregates are characterized by polydisperse structures, high mechanical strength and fragility, and the thermodynamic instability [23, 29, 30] caused by the existence of high internal stresses (absent in the equilibrium structures).

On the whole, the data obtained indicate the existence of some common ("unified") nonlinear and dissipative properties inherent in all the proteins formed under nonequilibrium conditions and dependent, rather, on the nonequilibrium conditions during protein

condensation and aggregation than on the chemical characteristics of the primary protein structures.

It should also be indicated that under nonequilibrium conditions, all the transient phases are, in fact, dissipative systems, such as the liquid phase of a colloidal suspension or a water–protein system [7], the film of an anisotropic nematic liquid crystal [25], an elastic film [27], and a film of unstable fragile solid material (similar to the films of other materials) [23, 29, 30]. Such a behavior of all the protein phases is explained by the presence of pronounced internal stresses dependent on the nonequilibrium conditions of the material formation in all the dissipative systems forming dissipative structures. It is well known that, minimizing the energy, these dissipative structures also play the role of energy sources [15, 17]. This makes it clear why the dissipative structures are necessarily formed at all the stages of protein condensation—they provide the energy for the functioning of "biological machines."

Since we consider here a condensing protein film, it is highly probable that the concrete form of the dissipation of the internal-stress energy of the protein would affect the protein elasticity, in full accordance with the elasticity theory [31]. In other words, there are grounds to believe that protein folding determining the mechanism of its functioning is of a nonlinear nature. It seems that we encounter here a new allotropic protein modification, which is formed as a result of the nonclystalline self-assembly of a cluster film with a morphologically new highly structurized order.

This new protein modification was obtained in this study by varying the kinetics of film formation via the fast removal of the solvent from the protein–water system. This procedure is quite consistent with modern concepts of modifying film structures (in particular, carbon-containing ones) with the aim of obtaining materials with modified properties by varying the kinetic conditions of their formation [32, 33].

The above speculations allow one to realize why real protein folding and the allotropic protein modification were reproduced in all the experiments (100% reproducibility). It seems that both energy accumulation and energy dissipation are provided by a specific mechanism acting under rather simple conditions varying within a rather wide range. Fast water evaporation from an open protein–water system under the conventional conditions (atmospheric pressure and room temperature) of the *in vitro* experiment leads to a complex process of molecular interactions and aggregation via self-assembly of nonequilibrium cluster protein films. It is seen that these conditions are close to the *in vivo* conditions for living systems.

The formation of protein films occurs under simple conditions, whereas the formation of films of nonprotein materials proceeds only under specific conditions, where the chaotic dynamics is observed only at high temperatures and pressures, etc. [8, 29].

The common ("unified") physical and chemical properties of protein molecules and macromolecules irrespective of the chemical nature of various proteins create the unique active media in water-containing systems and determine the "ranges" of the conditions necessary for nonequilibrium self-organization. Thus, the minimum requirements for the external conditions, their natural and simple character, the ability of the material to accumulate stresses "from nothing" and then dissipate the energy during the condensation process determine the place of the new allotropic protein modification in the hierarchy of compounds with "spontaneous energy sources" providing the formation and functioning of living systems.

Thus, the critical conditions for self-assembly of cluster protein films include nonequilibrium conditions, the hierarchy of the length scales, concentration, etc. Of special importance are the rate of the solvent removal and the solvent adhesion to the substrate.

The *in vitro* experiments showed that the dynamics of condensation and self-assembly of protein films acquires the characteristics typical of living objects, namely, the discontinuous helical and chiral symmetry and asymmetry (similar natures of defect formation), the morphological self-similarity of structurized specimens, nucleation, "bronchial growth," and fractal nature of treelike films. On the whole, the process is characterized by a nonlinear chaotic dynamics and manifests coherence, spontaneity, synchronism, etc. It should be emphasized that *in vitro* experiments showed that all the above protein properties can be attained in simple protein–water systems without the use of any enzymatic or energetic ingredients of living organisms. This indicates that the formation of allotropic protein and, thus, also "of simplest biological structures is possible under the abiotic conditions" (E. Braudo). This gives grounds to believe that the protein properties (at least, the morphological framework ones) are determined, to a large extent, by the protein–water system used.

CONCLUSION

Thus, we have established and qualitatively (phenomenologically) estimated the supramolecular packing in protein films self-assembled under nonequilibrium conditions. On the whole, the behavior and the structure of the material are consistent with the general theory of liquid-crystal and elastic cluster films [8, 12, 30], but, at the same time, also have some specific features.

The experiments performed allow us to state that, contrary to equilibrium conditions, nonequilibrium conditions provide the formation of a new liquid-crystal, elastic, or solid phases of the film with highly structurized order at both meso- and macrolevels, a thermodynamically nonequilibrium state with the formation of dissipative systems and structures characterized by a

new nonlinear chaotic dynamics and the fractal geometry, and new chemical and energy properties characteristic of the new allotropic protein modification formed under the nonequilibrium conditions (dissipative nanostructures).

We believe that we have taken only the first steps in the qualitative study of the new allotropic protein modification. However, the simplicity of the method also provides the determination of its quantitative characteristics, very useful for solving the problem of energy control in living systems (both in norm and pathology), which, in turn, can be useful for the development of nanorobotronics, an important field of technology imitating functioning of protein molecules [34].

ACKNOWLEDGMENTS

REFERENCES

1. E. G. Rapis, Pis'ma Zh. Tekh. Fiz. **14** (17), 1560 (1988).
2. E. G. Rapis and G. Yu. Gasanova, Zh. Tekh. Fiz. **61** (4), 62 (1991).
3. E. G. Rapis, A. S. Botin, and A. N. Zaikin, *Proc. Int. Conf. Devoted to the 100th Anniversary of F-groups* (Moscow, 1991).
4. E. G. Rapis, Pis'ma Zh. Tekh. Fiz. **21**, 13 (1995).
5. E. G. Rapis, Symmetry, Culture, and Science **6** (3), 439 (1995).
6. E. G. Rapis, Pis'ma Zh. Tekh. Fiz. **23** (7), 28 (1997).
7. E. G. Rapis, Self-Assembly of Protein Film Clusters (in press).
8. A. Aksay *et al.*, Science **273** (8), 892 (1996).
9. Zh. Zhang and M. Lagally, Science **276** (4), 377 (1997).
10. J. Preskil, Science **272** (5), 966 (1996).
11. E. Pennisi, Science **277** (9), (1997).
12. J. Parkinson, Science **270** (11), (1995).
13. A. Aggeli *et al.*, Nature **386** (6622), 259 (1997).
14. D. Liu *et al.*, Nature **376** (6536), 191 (1995).
15. I. Prigogine, *Self-Organization Phenomena in Physics and Chemistry* (San Francisco, 1980).
16. A. Winfree, *The Geometry of Biological Time* (Springer, Berlin, 1980).
17. A.N. Zaikin and A.M. Zhabotinskiĭ, Nature **225**, 535 (1970).
18. M. Dennin *et al.*, Science **272** (4), 388 (1996).
19. T. Elperin *et al.*, in *Physics of Complex Systems. Book of Abstracts. V Bar Ilan Conf. on Frontiers in Condensed Matter Physics* (Israel, 1997), p. 33.
20. L. Wilen *et al.*, Science **270** (11), 1164 (1995).
21. Ph. Poulin *et al.*, Science **275** (3), 1770 (1997).
22. S. Stupp *et al.*, Science **276**, 384 (1997).
23. Ch. Chen *et al.*, Science **276**, 396 (1997).
24. V. N. Izmaĭlova and P.I. Rebinder, *Shape Formation of Structures in Protein Systems* (Nauka, Moscow, 1974)
25. J. F. Joanuy, Science **275**, 1751 (1997).
26. T. Pekins, Science **276**, June (1997).
27. V. Steinberg, in *Physics of Complex Systems. Book of Abstracts. V Bar Ilan Conf. on Frontiers in Condensed Matter Physics* (Israel, 1997), p. 8.
28. T. Nishizaka *et al.*, Nature **377** (6546), 251 (1995).
29. B. Jerome and J. Commadeur, Nature **386** (6625), 589 (1997).
30. J. Sci *et al.*, Science **271** (2), 937 (1996).
31. O. Steinvock *et al.*, Science **267** (5199), 868 (1995).
32. A. I. Popov, in *Proc. All-Russian Symp. on Amorphous and Microscopic Semiconductors* (Ioffe Physicotechnical Inst. RAS, St. Petersburg, 1998), p. 14
33. V. A. Ligachev, in *Proc. All-Russian Symp. on Amorphous and Microscopic Semiconductors* (Ioffe Physicotechnical Inst. RAS, St. Petersburg, 1998), p. 17.
34. S. Santoli, Biosystems **42**, 77 (1997).

Translated by L. Man

=========================== **BRIEF COMMUNICATIONS** ===========================

Degassing of Water by Means of a Household Filament Lamp

B. G. Emets

Kharkov State University, pl. Svobody 4, Kharkov, 310077 Ukraine
Received October 19, 1998

Abstract—It was shown with the method of nuclear magnetic resonance that half-hour illumination of water by a 30-W filament lamp decreases the amount of dissolved and free air (bubbles) by 12% and 45%, respectively. The air content in water regains its initial value by diffusion, i.e., relatively slowly. The need for taking into account the illumination level in experiments with liquid media is indicated. © 2000 MAIK "Nauka/Interperiodica".

Along with dissolved air, water always contains air in the free state (bubbles) [1]. The air content in these two states is important in many technical and biological applications. For example, the tensile strength of a liquid increases with a decrease in the amount of air bubbles [2], while the lack of dissolved oxygen can be fatal for hydrobionts. It will be shown that the content of both dissolved and free air substantially drops when water is illuminated by a household filament lamp (HFL).

Efficient degassing of water by low-power microwave radiation was reported in [3]. The air was removed by the thermocapillar bubble mechanism. The same mechanism should undoubtedly be responsible for degassing of water illuminated by an HFL.

A 0.6-cm³ sample of distilled water was placed into a glass test tube and illuminated through a lens with a 30-W HFL for 30 min. The diameter of the light sport produced by the lens on the side wall of the test tube was 5 mm, and the incident light power, 60 mW. The experiments were carried out in the laboratory conditions at room temperature (20°C). Thirty-minute illumination raised the sample temperature by 4.5°C. (Further increase in the illumination time practically did not change the sample temperature.)

The method of nuclear magnetic resonance (NMR) was used to separately measure the concentrations of dissolved and free air in water [4]. The rates of spin-lattice proton magnetic relaxation, T_1^{-1}, and spin-spin proton magnetic relaxation, T_2^{-1}, in water are known to be proportional to the concentration of paramagnetic impurities [5]. We obtained the empirical relationships between the relaxation rates and the number N of oxygen molecules dissolved in 1 m³ of water (molecules of paramagnetic oxygen):

$$T_1^{-1}(N) = 0.258 + 7.44 \times 10^{-25} N, \qquad (1)$$

$$T_2^{-1}(N) = 0.308 + 12.27 \times 10^{-25} N. \qquad (2)$$

The values of T_1^{-1} were measured with the standard three-pulse method, while those of T_2^{-1} were obtained by using the Carr–Parcell multipulse method modified by Meiboom and Gill [5]. The working frequency of an NMR relaxometer was 15.9 MHz. Relationships (1) and (2) were derived by measuring the dissolved oxygen concentration, which was varied as follows: A water sample was first kept in a thermostat at a high temperature over a long period of time and then sharply cooled down to room temperature. (The effect of lowering the solubility of gases in water with increasing temperature was employed). Relationships analogous to (1) and (2) with slightly different numerical coefficients were reported in [6]. The discrepancies are explained by the fact that the samples were variously prepared: in [6], oxygen was introduced into predegassed water under pressure, while we kept our samples under normal atmospheric pressure. It was found [4] that relationship (1) can be used to determine the concentration of the dissolved air, while the amount of the free air (in the form of bubbles) can be obtained from (1) and (2). Our measurements showed that the volume of the free air in 1 m³ of the unilluminated water was $(4.7 \pm 1.3) \times 10^{-8}$ m³ at 20°C. (More accurate evaluations performed for the equisized bubbles gave the number of bubbles in one cubic meter of water equal to 2×10^{15} and the average radius of a bubble, approximately to 2×10^{-8} m.) The latter value is close to 5.23×10^{-8} m³, measured by a nephelometer [7] (the confidence interval was not indicated). Our measurements imply that, at 20°C, the concentration of the dissolved oxygen is $(1.7 \pm 0.11) \times 10^{23}$ m⁻³. This means that one cubic meter of the water contains 1.83×10^{-2} m³ of dissolved air, which is in agreement with the literature data [8].

Since 30-min illumination heats up the sample by 4.5°C, as was mentioned above, the illuminated sample was placed for 10 min into a 16-l water thermostat with a temperature of 20°C. This let the temperature of the sample return to its initial (before illumination) value,

20°C. Only after this procedure did we measure the rates of proton magnetic relaxations to evaluate the amount of the dissolved and free air. It was found that, after HFL illumination, the volume of the free air per m^3 of the water became equal to $(2.6 \pm 1.3) \times 10^{-8}$ m^3; i.e., it decreased almost by half. The concentration of the dissolved oxygen was $(1.49 \pm 0.15) \times 10^{23}$ molecules/m^3 [the volume of the dissolved air was $(1.61 \pm 0.16) \times 10^{-2}$ m^3]; i.e., it decreased by 12%. This value is consistent with the literature data [8] on the oxygen solubility in water, which indicate that it drops when a temperature rise equals that observed in our illumination experiments.

The decreased gas content in the water after illumination persisted at least 30 min. The fact is that the initial air concentration in the water is restored by diffusion, i.e., relatively slowly.

Thus, such a seemingly minor influence on water as illumination by a 30-W HFL provides substantial degassing of the liquid. This should be taken into account in treating results of experiments with liquids where light sources were used.

REFERENCES

1. L. R. Gavrilov, *Physical Principles of Ultrasonic Technology*, Ed. by L. D. Rozenberg (Nauka, Moscow, 1970).

2. M. G. Sirotyuk, Akust. Zh. **12,** 87 (1966) [Sov. Phys.–Acoust. **12**, 67 (1966)].

3. B. G. Emets, Pis'ma Zh. Tekh. Fiz. **22,** 22 (1996).

4. B. G. Emets, Pis'ma Zh. Tekh. Fiz. **23,** 42 (1997).

5. A. Abragam, *The Principles of Nuclear Magnetism* (Clarendon, Oxford, 1961; Inostrannaya Literatura, Moscow, 1963).

6. R. Hausser and F. Noak, Z. Naturforsh. **20a,** 1668 (1965).

7. V. K. Makarov, N. V. Chulkova, T. I. Kovaleva, and N. I. Shestakov, *Acoustics and Ultrasonic Technology* (Tekhnika, Kiev, 1988).

8. *Reference Book on Solubility*, Ed. by V. V. Kafarov (Izd. Akad. Nauk SSSR, Moscow, 1961).

Translated by A.P. Lytkin

BRIEF COMMUNICATIONS

Microelectronic Crossed-Field Cold-Cathode Amplifier

D. V. Sokolov and D. I. Trubetskov

"College" State Teaching-Research Center, Chernyshevskiĭ State University, Saratov, 410026 Russia

Received October 19, 1998

Abstract—The results of theoretical investigation of the structure of a microelectronic two-compartment crossed-field SHF amplifier with distributed cold cathode are presented. © *2000 MAIK "Nauka/Interperiodica".*

Technological achievements of the past decades, which led to the development of thin-film cold-cathode vacuum microtubes of various types, might seem to open unlimited possibilities for the creation of new superminiature amplifier and generator tubes for SHF and UHF ranges based on the classical triode schemes [1]. These devices were expected to possess higher stability with respect to radiation and temperature factors as compared to their solid-state counterparts. In addition, it was hoped that the free motion of electrons in the new devices would allow their working frequency to be markedly increased. However, the level of parameters of the cold-cathode triodes actually reached at present is by no means sufficient to compensate for the increased losses in microelectronic devices, which are markedly greater in miniature tubes than in ordinary ones because of the small thickness of conducting films. Moreover, a large input capacitance of the cold-cathode triodes restricts their working frequency to such a great extent that this parameter is frequently even lower as compared to that for the analogous solid-state devices. A natural consequence is the rejection of triode schemes in favor of the classical schemes of SHF electronics based on the long-time interactions of electron beams with retarded waves.

Below we will consider the possibility of realizing these interactions in a microelectronic variant of the triode amplifier with crossed electric and magnetic fields, based on a thin-film cold cathode. A usual scheme of such an amplifier [2, 3], with a distributed cold cathode situated in the interaction space in front of the retarding system, is not quite acceptable for microelectronic devices. In this scheme, the current flowing from a cathode to the interaction space increases along the cathode. Initially, the current is small and the build-up rate of the field generated by the electron beam is also rather low. As a result, the length of the interaction space corresponding to the field saturation is so large that the existing decrement (which, as noted above, is markedly greater in microelectronic retarding systems than in the usual ones) reduces the amplification effect

almost to zero. Note that an increase in the microcathode length is extremely undesirable for technological reasons.

A more effective solution is offered by a two-compartment amplifier tube in which the distributed cold cathode is situated in the first compartment and a flat conducting electrode is mounted on the same place in the second compartment. A signal to be amplified is applied to a segment of the retarding system in the first (input) compartment. The input compartment with a small length serves only as a source of electron bunches. Electrons leaving the cathode move by cycloid trajectories, their average velocities being close to the wave velocity in the retarding system. The electrons falling within the accelerating phase of electric field in the retarded wave move back and are absorbed by the cathode. The electrons within the retarding phase of the field are carried away to reach the retarding system.

Thus, the proposed system produces selection of the emitted electrons and modulation of the electron flux. Electrons in the acceleration phases, which might subsequently take the energy from the field, are excluded from the interaction process. Note that this selection takes place at a comparatively small field amplitude and, hence, is virtually independent of the decrement in the input compartment. The modulated electron flux enters the output section and induces a field in the second retarding system. Since the interactions from the

Table 1. Limiting (minimum) values of the magnetic field strength in crossed-field two-compartment cold-cathode amplifiers

f, GHz	B, T
10	0.18
50	0.9
100	1.8
200	3.6
400	7.2

very beginning involve electrons emitted from the whole cathode, the length of the interaction space reaching the field saturation state is markedly lower as compared to that in a single-compartment device.

Before proceeding to the results of calculations, it is necessary to consider one special feature inherent in the microelectronic variant of crossed-field devices. In the tubes of usual size, the distance h from anode to cathode is markedly greater compared to the cyclotron radius of electrons. For this reason, the cyclotron motion of electrons rather weakly affects their interaction processes. In contrast, the cathode to anode spacing in the microelectronic devices is small and the electrons moving by these trajectories may immediately strike the anode and escape from the interaction process. The limiting (minimum) values of the magnetic field B and anode voltage U_0 corresponding to this phenomenon may be determined from a relationship between the parameters of cycloidal motion and the static field values, assuming that the average electron beam velocity equals to the retarded wave velocity:

$$r = 0.036 f/(By), \quad U_0 = 5.11 \times 10^5/(rn^2).$$

Here $r = R/h$. R is the cycloid radius, h is the distance from cathode to retarding system, f is the frequency (GHz), $y = \beta h$, $\beta = 2fn/c$, n is the retardation factor, and c is the velocity of light. The limiting (minimum) values of the magnetic field strength (for $y = 4$) and anode voltage are listed in Tables 1 and 2, respectively.

According to Table 1, the limiting magnetic field strength falls within the range f really achievable values (at least for the frequencies below 200 GHz). Table 2 indicates that, for the microelectronic variant under consideration not to require the application of very high voltages, the retardation factor must be sufficiently large (significantly exceeding the values for usual SHF systems). Fortunately, the microelectronic thin-film retarding systems are characterized (see, e.g., [4]) by high retardation factors ($n \approx 20$–80) because of the presence of insulating substrates.

The amplifier was calculated in a two-dimensional approximation using a method of large particles conventionally used in SHF electronics. Distributed flux from a cold cathode was modeled by a set of planar ejects with the escape coordinates uniformly distributed along the cathode. Every ejected beam was represented as a set of noninteracting "large" rodlike particles. The retarding system parameters were specified using the data from [4].

The results of calculations are illustrated in Figs. 1 and 2. Figure 1 shows distribution of the gain coefficient G along the interaction space length l for the amplifiers with one and two compartments, which demonstrates efficacy of the preliminary modulation of the electron beam. For a set of parameters selected, this modulation leads to decrease in the field saturation length for the output compartment (by a factor of almost 2.5) and to the resulting increase in the maxi-
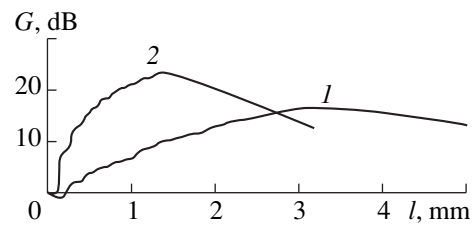
Fig. 1. Gain coefficient distribution along the interaction space in (*1*) single-compartment and (*2*) two-compartment amplifiers calculated for $f = 50$ GHz, $n = 46$, $l_1 = 1$ mm, $h = 0.04$ mm, $U_0 = 600$ V, and $B = 2.3$ T.
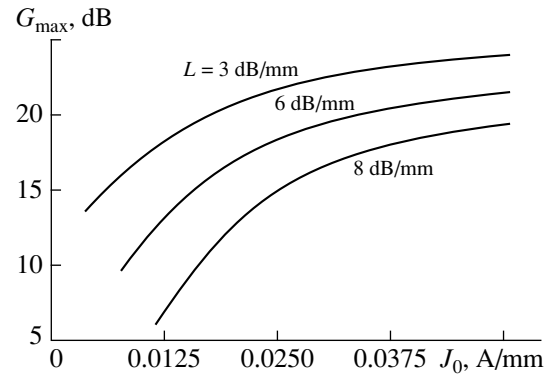
Fig. 2. The plots of maximum gain versus linear current density from the cold cathode for various values of the parameter of losses in the retarding system.

mum amplifier gain (by 7 dB) as compared to the single-compartment device. Figure 2 shows the plots of maximum gain coefficient G_{max} of a two-compartment amplifier versus the linear current density J_0 of the cold- emission microcathode and the parameter of losses L in the retarding system. It is important to note that quite satisfactory G_{max} vales are achieved with not too high (i.e., realistic) parameters of the microcathode and rather large losses in the retarding system.

The results of calculations (obviously having an estimation character) indicate that the electron interactions with crossed fields offer a promising approach to the development of microelectronic amplifiers for the SHF and UHF frequency ranges, with high parameters

Table 2. Limiting (minimum) values of the anode voltage in crossed-field two-compartment cold-cathode amplifiers

n	U_0, kV
10	10.2
20	2.55
30	1.23
50	0.41

achievable at the present-day state of vacuum micro-electronics.

REFERENCES

1. I. Brodie and C. A. Spindt, Adv. Electron. Electron Phys. **83**, 1 (1992).

2. *Crossed-Field Microwave Devices. Principal Elements of Crossed-field Devices*, Ed. by E. Okress (New York, 1961; Inostrannaya Literatura, Moscow, 1961).

3. D. I. Trubetskov, D. V. Sokolov. A. G. Rozhnev, and M. V. Gavrilov, in *Techn. Dig. 6th Int. Conf. Vacuum Microelectronics* (Newport, 1993), p. 70.

4. N. I. Sinitsyn *et al.*, J. Vac. Sci. Technol. B **11** (2), 477 (1993).

Translated by P.P. Pozdeev

============================ **BRIEF COMMUNICATIONS** ============================

Specific Features of the Ionization of Atmospheric Air Resulting from Radioactive Pollution

K. A. Boyarchuk, A. V. Karelin, and A. M. Lomonosov

Institute of General Physics, Russian Academy of Sciences, ul. Vavilova 38, Moscow, 117942 Russia

Received January 28, 1999

Abstract—Specific features of inhomogeneous ionization resulting from radioactive pollution in the lower atmosphere are investigated. A model of track ionization of air is proposed that allows for such elementary processes as kinetics, molecular diffusion, and drift of ions in an electric field. The applicability limits of the model are determined. © *2000 MAIK "Nauka/Interperiodica"*.

Atmospheric air under the action of hard ionizing radiation is a low-temperature nonequilibrium recombining plasma that contains a great number of various charged and neutral components with concentrations unusual for normal conditions [1]. It is commonly accepted that gaseous media are simpler with respect to ion–molecular processes than are liquids and solids, because, due to the low specific mass of gases, the track effects are almost absent when high-energy particles propagate in a gas, so that the primary products of radiolysis (electrons, ions, free radicals, and exited molecules) are uniformly distributed in space [2]. This significantly simplifies the theoretical analysis of ion–molecular kinetics in gases. Therefore, when analyzing the processes of radiochemical conversion of atmospheric air, it is usually supposed that ionization is spatially uniform [3–5]. Nevertheless, if the pressure is rather high (0.1–1 MPa), microscopic regions with a high concentration of ionized particles can form in air at room temperature. These are so-called tracks produced by high-energy corpuscles (protons, α- and β-particles, etc.); the radiolysis products are mainly localized in these tracks [1, 2]. Thus, when investigating the processes of the production of negative and positive ions in air at radioactive pollution, one should distinguish between the volume ionization by X- or γ-rays and the local track ionization by high-energy particles.

The main effect of ionizing radiation on air is due to secondary electrons emerging when atoms and molecules are ionized. The energy of these electrons lies in the range from the thermal energy to the energy of fast primary particles. The track structure depends on the particle species and energy that is spent on the production of secondary electrons emitted from the primary track and forming spur tracks. When the track is produced by an α-particle or a fission fragment, the spurs are located so close to each other that a continuous ionized region is formed around the main track.

The purpose of this paper is to study the influence of a track structure on the kinetic processes occurring in ionized atmospheric air.

Currently, there is no consistent theory describing the processes occurring in the tracks of ionizing particles. Therefore, the results of experimental investigations of the elementary processes in tracks are often interpreted in terms of simplified theoretical models by Onsager and Jaffe. The first model was developed to describe recombination processes in the tracks of high-energy particles in dense gaseous media at a pressure of ~20 MPa [6]. The second model [7], which is aimed at a description of recombination processes in tracks of α-particles in gas at a pressure of ~1 MPa, is more appropriate for a description of processes in the troposphere. The main thesis of the Jaffe theory is that the ions resulting from propagation of a particle are contained in a cylinder surrounding the trajectory of the particle. Ion–molecular processes in such tracks can differ from those in the case of volume ionization caused by X- or γ-rays.

For lower layers of the troposphere, we will suppose that positive and negative ions are uniformly distributed in the cylindrical volume around the particle trajectory. To estimate the diameter of the track, we assume that it is determined by the average length of spurs, i.e., tracks of secondary electrons forming an ionized region around the main track.

The cross section for ionization of air by α-particle is $\sigma \sim 5.1 \times 10^{-16}$ cm^2 [8]. For the mean free path l of the α-particle in air with respect to ionization events (i.e., for the average distance between the spurs), we obtain $l = 1/(\sigma N_L) = 7.3 \times 10^{-5}$ cm, where $N_L = 2.69 \times 10^{19}$ cm^{-3} is the Loschmidt number. The total mean path length Z of an α-particle with an energy of $E_\alpha \sim 5$ MeV in the troposphere is less than 4 cm; thus, the average energy loss ΔE per one ionization event is $\Delta E = (L/l)^{-1} E_\alpha \approx 100$ eV. The same value pertains to the energy of the secondary electrons. Using the Bethe–

Born approximation (its applicability to the electrons is validated by a high velocity of the incident electron compared to the velocities of the atomic electrons), we determine the cross section for ionization of air molecules by the secondary electrons. According to [9], we have

$$\sigma_2 = 4 \times 10^{-14} \xi \frac{\ln(\Delta E/I)}{\Delta EI} \approx 3 \times 10^{-16} \text{ cm}^2, \quad (1)$$

where $l = 15.58$ eV is the ionization potential and $\xi = 6$ is the number of equivalent electrons in the nitrogen molecule.

The minimal radius of the track will be assumed to be equal to the mean free path of secondary electrons, $r = \dfrac{1}{(\sigma_2 N_L)} \approx 1.3 \times 10^{-4}$ cm. Then, the track volume will be $V = \pi r^2 L \approx 2 \times 10^{-7}$ cm^3.

Let us determine the density of electron–ion pairs produced by the incident particle in this volume. The velocity of an α-particle in air can be found from the kinetic energy formula as $v = \sqrt{2E_\alpha/M} \approx 1.5 \times 10^9$ cm/s, where M and E_α are the mass and energy of the α-particle, respectively. Taking into account that the mean time of flight of the particle is $t = L/v \approx 2 \times 10^{-9}$ s, the averaged (over the time of flight) rate of the energy deposition is $w = E_\alpha/Vt \approx 3 \times 10^3$ W/cm^3. The ionization frequency is defined as $\nu = \dfrac{w}{(N_L E_i)} \approx 11$ s^{-1}, where E_i is the energy needed to produce one ion–electron pair in air (~34 eV). Finally, for the initial electron density in the α-particle track in air, we have $N_e = \nu N_L t \approx 10^{12}$ cm^{-3}.

Major elementary ion–molecular processes in the ionized troposphere ($T = 300$ K, $T_e = T$)

Reaction no.	Reaction	Rate α, cm^3 s^{-1}	References
1	$O_2^+ + e \longrightarrow O + O$	2.2×10^{-7}	[4]
2	$N_2^+ + e \longrightarrow N + N$	2.9×10^{-7}	[4]
3	$e + O_2 + O_2 \longrightarrow O_2^- + O_2^*$	2.5×10^{-30} (cm^6 s^{-1})	[4]
4	$O_2^- + O_2^* \longrightarrow O_2 + O_2 + e$	2.0×10^{-10}	[4]
5	$N_2^+ + O_2 \longrightarrow O_2^+ + N_2$	2.0×10^{-11}	[4]
6	$A^+ + B^- \longrightarrow A^* + B$	10^{-6}	[4]
7	$H_2O^+ + H_2O \longrightarrow H_3O^+ + OH$	1.2×10^{-9}	[2]
8	$e + H_3O^+ \longrightarrow H_2 + O + H$	1.3×10^{-6}	[2]

Generally speaking, the dimension and lifetime of the track are determined by both the ambipolar diffusion of ions from the initial track volume and recombination of primary ions in this volume. We note that, due to the small track volume, the probability for minor gaseous atmosphere components (with densities of $\leq 10^8$ cm^{-3}) to participate in ion–molecular processes is almost zero. Hence, we can limit the number of molecules and ions under study. In the model used, until the track decays, we can consider (without significant loss of accuracy) only the main ion–molecular processes involving the ions of major air components (N_2 and O_2), as was done in [3]. The concentrations of these major air components are stable and exceed the concentration of other possible minor gaseous components by several orders of magnitude. The influence of water vapor molecules H_2O on the ion–molecular processes and ion production should also be taken into account.

In the table, we present the list of elementary processes in a track that are accounted for in our model. Note that here we consider a simplified kinetic model of a track. In reality, the electron–ion recombination rates substantially depend on the electron temperature T_e and, since the electron temperature in the track can be fairly high, the recombination rates can be lower than those given in the table. The set of equations describing the ion kinetics in a track with allowance for molecular diffusion and ion drift in an electrical field, being written in the cylindrical coordinates, has the form

$$\frac{\partial n_i}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left[r\left(D_i \frac{\partial n_i}{\partial r} + \mu_i \frac{\partial}{\partial z}(E n_i) \right) \right]$$
$$+ \sum_{i \neq j}^{N} \alpha_{ij} n_i n_j \frac{1}{r}\frac{\partial}{\partial r}(rE) = 4\pi\rho, \quad (2)$$

where n_i, D_i, and μ_i are the densities, diffusion coefficients, and mobilities of electrons, positive and negative ions, and excited molecules ([e], [O_2^+], [N_2^+], [O_2^-], [O_2^*], [H_2O^+], and [H_3O^+]); α_{ij} are the ion–molecular reaction rates; ρ is the charge density; E is the electrical field; and r is the coordinate across the track.

Fig. 1 presents the results of the numerical solution of the above set of equations. It is seen that the most significant recombination processes involve positive and negative oxygen ions, because all of the positive ions rapidly transform into positive oxygen ions as a result of the charge-exchange reaction (5) (see table). Electron attachment to neutral oxygen molecules proceeds at almost the same rate. The rate of formation of the positive H_3O^+ ions is several orders of magnitude lower; thus, this process weakly affects the main processes, which occur during 1 ns. Then, the density of negative ions decreases due to recombination and molecular diffusion (ions leave the reaction region).

For comparison, Fig. 2 presents the time evolution of the radial profile of the density of negative oxygen ions. The lifetime of the track region in which the degree of ionization is quasi-uniform can be estimated from the figure as the time during which the ion density on the track axis decreases by a factor of e. The volume of the quasi-uniform ionization region increases due to diffusive spreading approximately by a factor of 2.5, reaching the value of $V_m \sim 5 \times 10^{-7}$ cm^3 for the time $t_m \sim 125$ ns.

Allowance for the track structure of air ionization is important in solving certain specific problems, e.g., in the analysis of active gaseous media for lasers with nuclear pumping [10] or a radar pulse reflection from an ionized cloud formed due to radioactive pollution [11].

An overlapping of the tracks can serve as a "strong" criterion for the necessity of taking into account the track structure of air ionization. If we denote the quantity characterizing the number of the tracks produced in a unit volume per second (i.e., the volume density of nuclide activity) by $a_{\alpha, \beta}$, then the condition under which the tracks do not overlap can be written as $a_{\alpha, \beta} V_m t_m \ll 1$. Otherwise, if $a_{\alpha, \beta} V_m t_m \gg 1$, the tracks do overlap and the influence of the track structure can be neglected. In other words, if the ionization rate is sufficiently high, the tracks merge and ionization can be regarded as spatially uniform. Thus, it is easily seen that, for an α-active pollution source, the above criterion is violated when the concentration of radioactive substances is such that $a_{\alpha, \beta} \sim 500$ Ci/cm^3, which corresponds to a volume ionization rate of $\sim 10^{18}$ cm^{-3} s^{-1}. Such levels of radioactive pollution can be observed only during major accidents [12]. Therefore, according to the strong criteria, tracks of nonuniform ionization will always be observed in the troposphere under moderate levels of radioactive pollution.

The effect of track ionization must certainly be taken into account in calculating the ion kinetics. The corresponding criterion can be defined as the level of radioactive pollution of the atmosphere above which the relaxation kinetic processes change their character. According to the proposed track ionization model, in the initial stage of ionization, all of the ion–molecular processes with the participation of primary ions involve the basic air components (N$_2$ and O$_2$) and occur in the vicinity of the tracks of high-energy particles. In these processes, which occur while the track exists, the reactions of electron–ion and ion–ion recombination are dominant. Only after 10^{-7} s, does the track blur out and do the ions generated in it mix with the bulk of union-ized air and start to react with neutral molecules of minor gaseous components of the atmosphere, e.g., nitrogen- or carbon-containing compounds. Depending on the concentration balance between the primary ions produced in the tracks and minor gaseous components that are present in the whole volume, either the processes of ion–ion recombination or reactions of ions
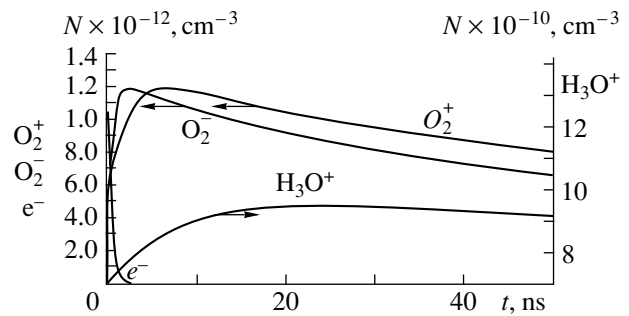

Fig. 1. Time variations of the densities of electrons and major ions in the track caused by molecular diffusion and recombination. The oxonium ion density is scaled up by a factor of 2×10^2.
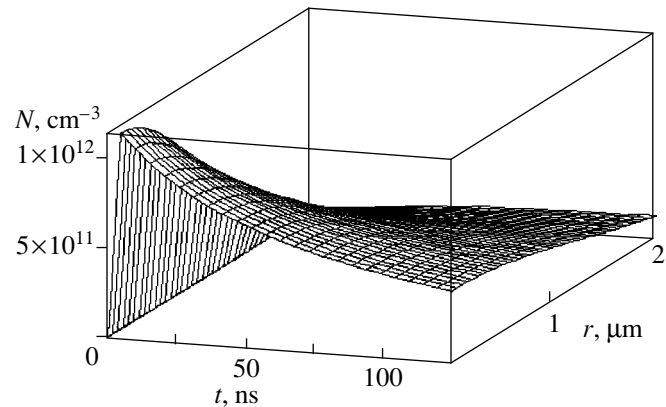

Fig. 2. Density of negative ions of molecular oxygen in the track as a function of time and distance from the track axis.

with neutral molecules can prevail. These two cases are characterized by different production rates and final concentrations of the dominant ions in the atmosphere [1].

Such a "weak" criterion can be expressed as follows:

$$X = N_i^2 \alpha_6 - \alpha N_i N_{NO}, \qquad (3)$$

where N_i is the averaged density of ions in the track, N_{NO} is the density of the most abundant minor gaseous components of the atmosphere (e.g., the typical concentration of nitrides in the troposphere is $\sim 10^{10}$ cm^{-3}), α_6 is the ion–ion recombination rate constant (see table), and $\alpha \sim 10^{-10}$ cm^3 s^{-1} is the rate constant of reactions between ions and neutral molecules [1].

If $X > 0$, then the track structure can be neglected, because the relaxation process proceeds, at first, mainly through the ion–ion recombination channel and no new specific features appear in the relaxation kinetics of media. If $X < 0$, then the track ionization can, in principle, affect the relaxation, because, in this case, the reactions of ions with neutral molecules are the dominant processes of the ion relaxation and they can lead, e.g.,

to the fast production of stable NO_3^- ions that are specific to the troposphere. However, the last assumption requires additional investigation. Probably, this criterion will also be found to be too strong.

In the case of α-active pollution, the critical value of the radioactivity density, according to (3), will be $\sim 10^{-2}$–10^{-3} Ci/cm^3, which corresponds to an ion production rate of $\sim 10^{12}$ cm^{-3} s^{-1}. Similar estimates can easily be obtained for β-particles. For β-active pollution, we obtain $\sim 10^{-4}$ Ci/cm^3 and $\sim 10^9$ cm^{-3} s^{-1}, respectively. Since, in the case of real radioactive pollution, the sources of the α-, β-, γ-, and other kinds of radiation are present simultaneously, it is appropriate to average our estimates and consider an ion production rate of 10^9–10^{12} cm^{-3} s^{-1} as the boundary value for the "weak" criterion. Expressed in units of the power of the absorbed radioactive emission, this corresponds to $\sim 4.2 \times 10^{-3}$ Gy/s.

In this paper, we have analyzed the specific features of nonuniform ionization of air resulting from radioactive pollution of the lower atmosphere. A model for describing the track ionization of the atmosphere is proposed, and the applicability limits of the model are determined.

REFERENCES

1. V. V. Smirnov, *Ionization in the Troposphere* (Gidrometeoizdat, St. Petersburg, 1992).
2. A. K. Pikaev, *Modern Radioactive Chemistry: Radiolysis of Gases and Liquids* (Nauka, Moscow, 1986).
3. K. A. Boyarchuk and Yu. P. Svirko, Pis'ma Zh. Tekh. Fiz. **22** (14), 47 (1996).
4. K. A. Boyarchuk, Izv. Akad. Nauk, Ser. Fiz. Atmos. Okeana **33**, 236 (1997).
5. K. A. Boyarchuk and A. V. Karelin, Bull. Russ. Acad. Sci., Suppl. Ser. Phys., Phys. Vib. **61**, 249 (1997).
6. L. Onsager, Phys. Rev. **54**, 554 (1938).
7. G. Jaffe, Ann. Phys. (Leipzig) **42**, 313 (1913).
8. *Physical Quantities: Reference Book*, Ed. by I. S. Grigor'ev and Ye. Z. Meĭlikhov (Energoatomizdat, Moscow, 1991).
9. L. I. Gudzenko and S. I. Yakovlenko, *Plasma Lasers* (Atomizdat, Moscow, 1978).
10. A. P. Budnik and I. V. Dobrovol'skaya, Kvantovaya Elektron. (Moscow) **24**, 506 (1997).
11. K. A. Boyarchuk, G. A. Lyakhov, and N. V. Suyazov, Zh. Tekh. Fiz. **67** (2), 76 (1997) [Tech. Phys. 42, 190 (1997)].
12. Yu. A. Izraél', *Radioactive Fallouts after Explosions and Accidents* (Progress-Pogoda, St. Petersburg, 1996).

Translated by A. E. Kramida

===== **ATOMS, SPECTRA, AND RADIATION** =====

Observation of a Circular Dichroism of an Ensemble of Triplet Metastable Helium Atoms in a Na–He Gas-Discharge Plasma with Laser-Oriented Sodium Atoms

S. P. Dmitriev, N. A. Dovator, R. A. Zhitnikov, V. A. Kartoshkin, and V. D. Mel'nikov

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

Received July 31, 1998

Abstract—We describe an experiment on the observation of a circular dichroism of an ensemble of triplet metastable helium atoms in a sodium–helium gas-discharge plasma where Na atoms are optically oriented with circularly polarized radiation of a laser tunable around the resonant sodium doublet (with wavelength of 589.0 and 589.6 nm, corresponding to the $3^2S_{1/2} \rightleftharpoons 3^2P_{1/2}$ and $3^2S_{1/2} \rightleftharpoons 3^2P_{3/2}$ transitions). Conditions of observation and the polarization degree of 2^3S_1 atoms are determined. © *2000 MAIK "Nauka/Interperiodica"*.

Ensembles of spin-polarized atoms in gases and gas-discharge plasmas can be prepared not only by means of direct optical orientation, i.e., under the action of resonant circularly polarized radiation, but also with the use of other methods based on the transfer of spin polarization from an ensemble of optically oriented atoms of one element to an atomic ensemble of another element. Such a polarization transfer in gases and gas-discharge plasmas may be associated with several spin-dependent processes accompanying collisions of various paramagnetic atomic particles [1]. Polarization of an ensemble of triplet metastable helium atoms in an alkaline–helium plasma through the optical orientation of alkaline-metal atoms is one of such indirect polarization methods. Spin polarization can be transferred from optically oriented $n^2S_{1/2}$ alkaline-metal atoms to 2^3S_1 metastable He atoms through several simultaneous spin-dependent processes of interatomic and electron–atom collisions. These processes include the spin-dependent Penning ionization of alkaline-metal atoms in the interaction of these atoms with metastable helium atoms and interatomic spin exchange accompanying elastic collisions of these atoms [2–6]. Collisions of metastable helium atoms with electrons polarized in the process of spin exchange between free electrons in a plasma and an ensemble of optically oriented alkaline-metal atoms may play an important role in the polarization of an ensemble of metastable helium atoms [2, 3, 5, 6].

This method was employed to prepare ensembles of spin-polarized metastable He atoms excited in Cs–He [3], Rb–He [2, 6] and K–He [5] plasmas under conditions of pulsed high-frequency discharge, which were later used in physical experiments and devices of quantum electronics. Alkaline-metal atoms were optically oriented in these studies with circularly polarized radiation of spectral lamps containing the relevant alkaline metals. Polarization of helium atoms was detected as the change in the absorption of pump radiation by alkaline-metal atoms under conditions when this polarization decayed due to a magnetic resonance excited in the 2^3S_1 state of He atoms.

In this study, we implemented a spin polarization of triplet metastable helium atoms in a sodium–helium plasma excited by a continuous high-frequency discharge under conditions of optical orientation of $3^2S_{1/2}$ Na atoms with circularly polarized radiation of a tunable laser. We propose a method of direct optical detection of the polarization of metastable He atoms thus induced based on the observation of the circular dichroism of an ensemble of 2^3S_1 helium atoms arising due to the polarization of the magnetic moments of these atoms.

A diagram of the experimental setup is shown in Fig. 1. Sodium atoms were optically oriented with radiation of a continuous-wave tunable Rhodamine 6G dye laser. Radiation of a tunable laser polarizes Na atoms by the emission of one of the *D* lines of the resonant doublet (we were able to use either D_1 or D_2 lines with wavelengths 589.0 and 589.6 nm, respectively). As a pump laser, we employed an LGN-404a argon laser. The cavity of the tunable laser included a Lyot filter and a thin etalon. The power of the pump laser was about ≈ 4 W. The radiation power of the tunable laser at the wavelength of one of the lines of the sodium resonant doublet ($\Lambda_1 = 589.6$ nm and $\Lambda_2 = 589.0$ nm) was about ≈ 70 mW. The bandwidth of laser radiation was ≤ 0.01 nm. The laser beam passed through a quarter-wave plate ($\Lambda/4$). Orientation of this plate determined the sign of the circular polarization (σ^+ or σ^-) of laser radiation. Mechanically rotating the quarter-wave plate ($\Lambda/4$) in the plane perpendicular to the pump laser beam, we were able to periodically change the sign ($\sigma^+ \rightleftharpoons \sigma^-$) of the circular polarization of laser radia-

tion. The rotation frequency was $\Omega \approx 100$ Hz, and the sign of light polarization ($\sigma^+ \rightleftharpoons \sigma^-$) was changed with a frequency equal to 2Ω. The circularly polarized pump beam thus produced entered an absorption chamber. The absorption chamber (a glass cylinder with a length of 5 cm and a diameter of 5 cm) contained metallic sodium and HE gas (at a pressure of 1 torr). A glass-enclosed electrode placed inside the chamber excited a continuous high-frequency (45 MHz) discharge. We were able to vary the concentration of metastable helium atoms by changing the high-frequency discharge current. The absorption chamber was placed in a glass thermostat, which ensured the heating of the chamber up to a temperature of about 150°C, allowing the density of Na vapor corresponding to the concentration of sodium atoms equal to approximately $\approx 10^{11}$ cm^{-3} to be achieved. The thermostat with the chamber was positioned at the center of a multilayer cylindrical ferromagnetic screen. A system of Helmholtz rings inside the screen induced a magnetic field $H_0 \approx 40$ mOe directed along the screen axis. The pump beam employed for the optical orientation of sodium atoms was directed along H_0. The spin polarization of sodium atoms thus induced is transferred to metastable helium atoms through collision processes, giving rise to a circular dichroism of an ensemble of 2^3S_1 He atoms at the wavelength $\lambda = 1083$ nm of the relevant helium line (2^3S_1–$2^3P_{0,1,2}$ transitions). The direction of magnetization of an ensemble of helium atoms under these conditions depends on the direction of spin orientation of Na atoms, i.e., on the sign of the $\sigma^\pm$ polarization of pump radiation of sodium atoms. The spin polarization of 2^3S_1 He atoms induced in collision processes was detected with the use of a light beam of a helium lamp directed along the field H_0 in the direction opposite of the pump beam (the line at $\lambda = 1083$ nm was separated with an interference filter placed in front of the photodetector). Nonpolarized light of the helium lamp passed through the absorption chamber, reaching a circular analyzer ($\lambda/4$, P), which was adjusted in such a way as to transmit σ^+-polarized helium light. Variations in the intensity of detecting light arising in response to changes in the orientation of the spin moments in the ensemble of He atoms [($\sigma^+ \rightleftharpoons \sigma^-$) changes in the sign of the polarization of laser radiation employed for the orientation of Na atoms] were registered with a photodetector, narrow-band amplifier, and a synchronous detector tuned to the frequency 2Ω, equal to twice the modulation frequency. This detection scheme offered important advantages over the detection scheme employed earlier, where the polarization of helium atoms was detected from variations in the absorption of pump radiation by alkaline-metal atoms, as the above-described detection scheme was insensitive to the noise of laser radiation in the detection path and excluded the influence of direct optical orientation or alignment of helium atoms by detecting radiation on the signal being registered. To detect the signals corre-
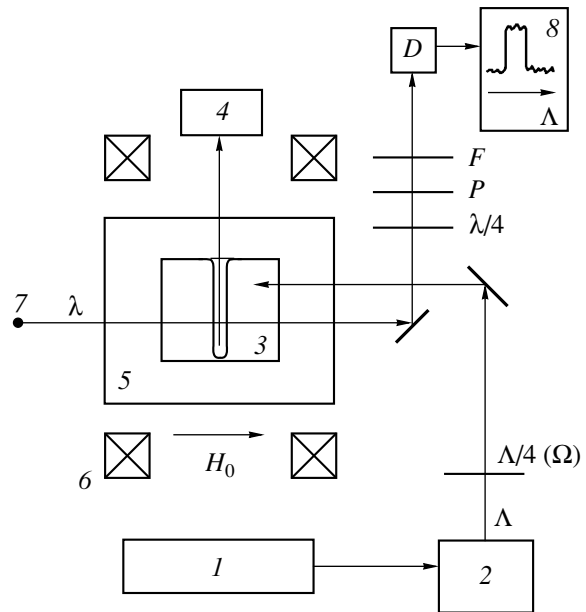

Fig. 1. Diagram of the experimental setup: (*1*) argon pump laser; (*2*) tunable laser (Λ); (*3*) absorption chamber; (*4*) high-frequency oscillator; (*5*) thermostat; (*6*) Helmholtz rings; (*7*) detecting helium lamp; (*8*) detection scheme; *D*, photodetector; *F*, interference filter ($\lambda = 1083$ nm); $\Lambda/4$, rotating quarter-wave plate ($\Lambda = 589$ nm); and $\lambda/4$, *P*, circular polarization analyzer ($\lambda = 1083$ nm).
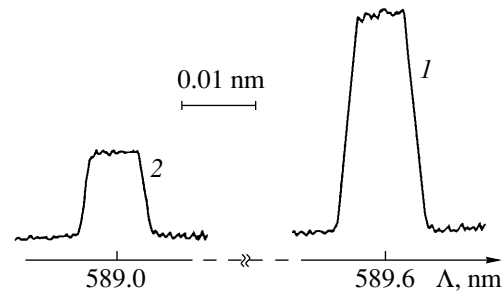
Fig. 2. Dependence of the intensity of detecting light ($\lambda = 1083$ nm) on the wavelength of laser radiation around spectral lines corresponding to the sodium resonant doublet: (*1*) around the D_1 line of the doublet and (*2*) around the D_2 line of the doublet.

sponding to the variation in the intensity of detecting radiation (the line at $\lambda = 1083$ nm) under conditions of optical orientation of sodium atoms with $\sigma^\pm$-polarized laser radiation, we tuned the wavelength of laser radiation around one of the resonant lines in the sodium doublet (Λ) by means of a thin wavelength etalon ($\Lambda_{1,2}$).

Figure 2 displays the variation in the intensity of detecting radiation as a function of Λ with fixed analyzer orientation ($\lambda/4$, *P*) corresponding to σ^+-polarized helium light in the case when the sign of the circular polarization of pump radiation changes ($\sigma^+ \rightleftharpoons \sigma^-$) with the frequency 2Ω. As can be seen from Fig. 2, when the wavelength of laser radiation (Λ) coincides

δI, arb. units
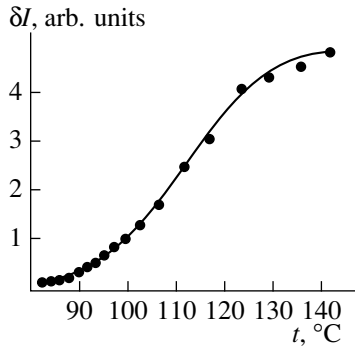
Fig. 3. Temperature dependence of the amplitude of the signal δI in the case when emission of the D_1 line of sodium is employed for pumping.

with one of the lines in the resonant doublet of sodium atoms, the signal corresponding to the absorption of helium-lamp detecting radiation by an ensemble of spin-polarized 2^3S_1 metastable helium atoms changes. The amplitude of these signals in the case when sodium atoms are pumped by the emission of the D_1 line (589.6 nm) was three times higher than in the case when sodium atoms are pumped by the emission of the D_2 line (589.0 nm), which indicates a more efficient polarization of spin momenta in the $3^2S_{1/2}$ state of Na atoms induced by the emission of the D_1 line as compared with polarization induced by the emission of the D_2 line with the same intensity. The ratio of the amplitude of the detected signal δI to the difference ΔI_0 of the intensities of helium radiation in the cases when the discharge is switched on and off was equal to $\delta I/\Delta I_0 = 0.002$ (or 0.2%). This ratio characterizes the stationary absorption of helium radiation by an ensemble of metastable atoms. The relative absorption of the line at $\lambda = 1083$ nm, which characterizes the concentration of 2^2S_1 helium atoms, was equal to $\Delta I_0/I_0 = 0.35$ (where I_0 is the intensity of detecting radiation in the absence of metastable helium atoms). The ratio $(1/2)(\delta I/\Delta I_0)$ can be employed to estimate the degree of spin polarization in an ensemble of 2^3S_1 He atoms due to collision processes in a Na–He plasma under conditions of optical orientation of sodium atoms. We should emphasize that the possibility of determining the ratio $\delta I/\Delta I_0$ is one of the main advantages of the implemented detection scheme over the schemes employed in earlier studies [2–6].

The signals of absorption variation δI were observed within the range of temperatures t from 80 up to 140°C. The temperature dependence of δI is presented in Fig. 3. The increase in the amplitude of signals with temperature growth is due to the increase in the concentration of sodium atoms, which serve as a source of spin polarization of metastable He atoms. The fact that the growth of the amplitude of δI slows down for temperatures above 120°C can be accounted for by

the decrease in the polarization degree of sodium atoms due to the growth in the optical density of metal vapor, lowering the effective pump rate of sodium atoms in the volume of the absorption chamber.

We should note that the above-described experimental scheme provides an opportunity to determine the bandwidth of laser radiation corresponding to the lines of the sodium resonant doublet with an apparatus width equal to the bandwidth of the absorption line of Na atoms. Indeed, taking into account that the bandwidth of the absorption line of sodium atoms under our experimental conditions is much less than the bandwidth of laser radiation, we can determine the bandwidth of laser radiation from the width of the detected signals (Fig. 2). In the case under consideration, this width was about 0.007 nm. This experimental technique is of certain interest because the detected signal is initiated by the absorption of laser radiation by sodium atoms, determining the apparatus width in measurements of the laser-line bandwidth. At the same time, detection is performed with a radiation of a low-noise helium lamp at the line with $\lambda = 1083$ nm, which lies far from the sodium doublet. This circumstance is of special importance when the laser beam is characterized by a high level of intensity noise associated with the laser itself and with optical elements employed in the experimental setup. In the proposed scheme, laser radiation under investigation has no influence on the photocurrent of the detector.

Thus, in this paper, we observed a circular dichroism of an ensemble of triplet metastable helium atoms excited in a Na–He gas-discharge plasma where sodium atoms were optically oriented with circularly polarized laser radiation. We determined the observation conditions and the magnitude of the circular dichroism induced by atomic collisions in a plasma and proposed a method for measuring the bandwidth of laser radiation employed as pump light for optical orientation of sodium atoms.

REFERENCES

1. W. Happer, Rev. Mod. Phys. **44**, 169 (1972).

2. G. M. Keiser, H. G. Robinson, and C. E. Jonson, Phys. Lett. A **51**, 5 (1975).

3. E. V. Blinov, R. A. Zhitnikov, and P. P. Kuleshov, Pis'ma Zh. Tekh. Fiz. **2**, 305 (1976).

4. S. P. Dmitriev, R. A. Zhitnikov, and A. I. Okunevich, Zh. Éksp. Teor. Fiz. **70**, 69 (1976).

5. E. V. Blinov, R. A. Zhitnikov, and P. P. Kuleshov, Zh. Tekh. Fiz. **54**, 287 (1984) [Sov. Phys. Tech. Phys. **29**, 168 (1984)].

6. E. V. Blinov, R. A. Zhitnikov, and P. P. Kuleshov, Zh. Tekh. Fiz. **54**, 2315 (1984) [Sov. Phys. Tech. Phys. **29**, 1362 (1984)].

Translated by A.M. Zheltikov

=========== **GASES AND FLUIDS** ===========

Effect of the Marangoni Convection on the Injection Mechanism of Instability

R. V. Birikh and A. V. Lyushnin

Perm State Pedagogical University, ul. K. Marksa 24, Perm, 614600 Russia
Received May 26, 1998; in final form, January 18, 1999

Abstract—Electroconvective instability of a nonisothermal layer of a weakly conductive liquid with a free boundary whose surface tension depends linearly on temperature is considered for the case where charge injection is performed through this surface. When calculating the unperturbed stationary distribution of the charge and field, we supposed that the injector is separated from the liquid by an air gap of finite thickness. It was, however, assumed when analyzing the stability of the system that the motion in the air gap has no effect on the motion in the liquid phase and the disturbances of the electric field and charge in the air gap decay rapidly because of its high conductivity. © 2000 MAIK "Nauka/Interperiodica".

1. Studies of the problems of electroconvective instability of the motion of fluids (liquids and gases) is related to numerous applications in various fields of engineering and technology. The greatest interest in this problem due to the development of the theory of breakdown and prebreakdown states. The occurrence of motion in the liquid layer sharply changes the charge transfer through it and the breakdown conditions. Mathematical simulation of this problem may lead to a better understanding of the related physical processes.

In recent years, new technologies have been developed for the intensification of heat and mass transfer in weakly conducting liquids using electric fields. On the one hand, an electric field creates a bulk ponderomotive force in the liquid, and, on the other hand, the charges present in the liquid take part in convective motion, forming convective electric currents. For the interaction of the electric and hydrodynamic fields, a space charge should be present in the liquid.

There are two models describing the mechanism of the formation in the liquid of a space charge that participates in the process of charge transfer: the conductivity model and the mobility model. In the first case, the formation of a free charge occurs as a result of dissociation–recombination; in the second case, the charge is formed due to injection through a boundary. Correspondingly, the laws of charge transfer prove to be different. In this work, we consider the injection mechanism of the formation of a free charge in a liquid.

When a liquid with a spatial distribution of charge is placed in an electric field, a ponderomotive force is developed, which is similar to gravity. Upon a nonuniform distribution of the spatial charge, the mechanical equilibrium of the liquid may prove to be unstable, leading to the appearance of a convective flow.

The convective instability of an isothermal layer of a weakly conducting liquid placed between two flat solid electrodes in the case of the injection mechanism of the formation of space charge was studied in some detail by Atten and Moreau [1]. The authors [1] report the threshold values of the electrical analog of the Rayleigh number for the monotone instability for two limiting cases: strong injection and weak injection. The stability of an isothermal liquid layer with a free boundary through which injection was performed was studied in [2, 3]. The injector was located on the free boundary. The critical value of the Rayleigh number and the related wave number were found.

In a nonisothermal liquid with a free boundary, a mechanism of instability arises, which is related to the thermocapillary effect. If there is formed a temperature gradient along the free boundary, then a tangential force that is capable of causing a convective motion in the liquid is developed due to the temperature dependence of surface tension. The linear theory of the thermocapillary instability of the equilibrium of a liquid layer in the presence of a transverse temperature gradient was considered in [4–8]. It was shown that, when heating was performed from the solid-boundary side, only a monotone instability could exist, whereas upon heating from the free-boundary side, an oscillatory instability mode can exist if the boundary is capable of deforming.

This work is devoted to studying these two mechanisms of instability. Charge injection was effected through the free deformable surface of nonisothermal liquid with surface tension linearly depending on the temperature.

2. Let us consider a flat horizontal layer of a weakly conductive viscous incompressible liquid of thickness h located between two plane-parallel electrodes separated by a distance of H. The origin is placed at the bottom solid surface; the x axis is oriented horizontally, and the z-axis is directed upward. Charge injection into

the liquid is performed through the air gap located above the liquid. The coefficient of surface tension of the interface between the two media depends linearly on temperature and can be described by the formula $\sigma = \sigma_0 - \sigma_1 T$. The equation of motion of liquid and of heat and mass transfer may be written as follows:

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}(\nabla \mathbf{v}) = -\frac{1}{\rho}\nabla p + \nu\Delta\mathbf{v} + \frac{q\mathbf{E}}{\rho}, \quad \text{div } \mathbf{v} = 0, \quad (1)$$

$$\frac{\partial T}{\partial t} + \mathbf{v}\nabla T = \chi\Delta T, \tag{2}$$

$$\text{div } \mathbf{E} = \frac{q}{\varepsilon}, \quad \mathbf{E} = -\nabla\phi, \tag{3}$$

$$\frac{\partial q}{\partial t} + \text{div } \mathbf{J} = 0, \quad \mathbf{J} = q(\mathbf{v} + b\mathbf{E}). \tag{4}$$

Here, $\mathbf{v}$ is the vector of liquid velocity; p is the pressure excess above the hydrostatic pressure; T, the temperature; $\mathbf{E}$, the electric field strength in the liquid; $\mathbf{J}$, the flux density of free charges; ϕ, the electric potential in the liquid; q, the density of free charges; ρ, the material density; ν and χ, the kinematic viscosity and thermal diffusivity, respectively; ε, the dielectric constant of the liquid; and b, the charge mobility in the liquid. The macroscopic motion in the air gap will be ignored, since this motion virtually has no effect on the transfer of charged particles [3] and has only a weak effect on the mechanical motion in the liquid. In this approximation, it is sufficient to write equation for the field and current in the air layer as follows:

$$\text{div } \mathbf{E}_a = \frac{q_a}{\varepsilon_a}, \quad \mathbf{E}_a = -\nabla\phi_a, \quad \mathbf{J}_a = q_a b_a \mathbf{E}_a. \tag{5}$$

The index a here specifies quantities that refer to air.

The boundary conditions for the set of equations (1)–(5) will be written as follows. At the injector, a constant value of the electric potential is maintained and the condition for the unipolar injection is fulfilled:

$$z = H: \phi_a = U_H, \quad E_a = 0. \tag{6}$$

For the unperturbed stationary state, the interface is assumed to be isothermal and equipotential; the temperature and the potential in the liquid are measured from the temperature and the potential of the interface, respectively. Thus, at the interface between two media, the following equalities are assumed to be fulfilled for the ground state:

$$z = h: \phi_a = \phi = 0, \quad T = 0,$$
$$\varepsilon_a E_a - \varepsilon E = 0, \quad J_a = J. \tag{7}$$

For the system of two weakly conductive media with an injection law of conductivity in the absence of specific mechanisms of the formation of surface charge, a situation is realized where the surface charge is zero [3]. This is explained by the fact that with the injection mechanism of conductivity, the free charge that creates the compensation field is distributed throughout the entire insulator. A jump in the mobility at the boundary of the two media is accompanied by a jump in the spatial density of the free charge, which compensates the change in the mobility. The normal component of the electric field induction proves to be continuous.

The boundary condition at the collector is written as follows:

$$z = 0: \phi = -U, \quad T = \Theta. \tag{8}$$

The temperature of the lower boundary may be both positive and negative. The set of equations (1)–(5) with the boundary conditions (6)–(8) has an equilibrium solution corresponding to the quiescent liquid ($v = 0$) with a linear temperature distribution

$$T_0(z) = \Theta\left(1 - \frac{z}{h}\right). \tag{9}$$

The distributions of the electric potential and charge density for the two media have the following solutions: for air,

$$\phi_{0a}(z) = \left(\frac{8J_0}{9b_a\varepsilon_a}\right)^{\frac{1}{2}}\left[(H-h)^{\frac{3}{2}} - (H-z)^{\frac{3}{2}}\right],$$
$$q_{0a}(z) = \left(\frac{\varepsilon_a J_0}{2b_a(H-z)}\right)^{\frac{1}{2}}, \quad h \le z \le H; \tag{10}$$

for weakly conductive liquid,

$$\phi_0(z) = -U + \left(\frac{8J_0}{9b\varepsilon}\right)^{\frac{1}{2}}\left[C^{\frac{3}{2}} - (C-z)^{\frac{3}{2}}\right],$$
$$q_0(z) = \left(\frac{\varepsilon J_0}{2b(C-z)}\right)^{\frac{1}{2}}, \quad 0 \le z \le h. \tag{11}$$

The parameter C is written as

$$C = h + \alpha(H-h), \quad \alpha = \frac{\varepsilon_a b}{\varepsilon b_a}. \tag{12}$$

The current density may be found from the condition of the equality of the potentials at the interface between the two media:

$$U = \left(\frac{8J_0}{9b\varepsilon}\right)^{\frac{1}{2}}\left[C^{\frac{3}{2}} - (C-h)^{\frac{3}{2}}\right]. \tag{13}$$

As was shown by calculations, the current J_0 decreases with increasing thickness of the air layer.

3. To study the stability of the equilibrium state (2) and (8) of a liquid layer, the method of small perturbations may be used. The charge and field perturbations in the air layer are rapidly decaying, in view of the relatively high charge mobility ($b/b_a \le 10^{-4}$ for an air–organic liquid system) and may be ignored. Let v, p, T, ϕ, and q be small perturbations of the velocity, pressure,

temperature, potential, and charge density in the liquid, respectively.

Let us discuss the boundary conditions at the free surface. We assume that under the effect of perturbations, the initially flat free surface can be deformed:

$$z = h + \xi(x, t). \tag{14}$$

Assuming the deformation to be small as compared to h, the boundary conditions at $z = h + \xi$ may be replaced by those at $z = h$ using a Taylor expansion.

The normal component of the stress tensor will be written as follows:

$$z = h: -p + \rho g \xi - \sigma \frac{\partial^2 \xi}{\partial x^2} + \varepsilon \left(\frac{d\phi_0}{dz} \right) \frac{\partial \phi}{\partial z}$$
$$+ 2\nu\rho \frac{\partial v_z}{\partial z} = 0. \tag{15}$$

The absence of the electric term in the tangential component of the stress tensor permits us to write this equation in the form

$$z = h: \nu\rho \left(\frac{\partial v_z}{\partial x} + \frac{\partial v_x}{\partial z} \right) = \frac{\partial \sigma}{\partial x}. \tag{16}$$

These conditions should be supplemented by the kinematic relationship

$$z = h: v_z = \frac{\partial \xi}{\partial t}. \tag{17}$$

When formulating boundary conditions for the temperature perturbations, we assume that the thermal conductivity of the liquid is much greater than that of the upper medium and, although there is some heat transfer at the free surface, we may assume that the temperature disturbances in the liquid do not change the heat flux through the air gap and the free boundary is thermally insulated against disturbances (the Biot number is equal to zero). This condition is most favorable for the development of thermocapillary instability [8]. Thus, the boundary condition for temperature disturbances is written as

$$z = h: T' = 0. \tag{18}$$

The requirement that the potential and field disturbances vanish at the interface with the air gap at the flat boundary $z = h$ is written (to the first order perturbation) as

$$\phi(h) + \xi \frac{d\phi_0}{dz} = 0, \quad \phi'(h) + \xi \frac{d^2\phi_0}{dz^2} = 0. \tag{19}$$

The boundary conditions at the solid isothermal equipotential interface for the velocity, temperature perturbations, and potential are as follows:

$$z = 0: v_z = 0, \quad v_z' = 0, \quad T = 0, \quad \phi = 0. \tag{20}$$

As the units of measure for the distance, time, velocity, pressure, temperature, potential, and the densities of charge and current, we take h, h^2/ν, χ/h, $\rho\chi\nu/h^2$, Θ, U, $\varepsilon U/h^2$ and $\varepsilon b U^2/h^3$, respectively. We will investigate the stability of equilibrium distributions with respect to small "normal" perturbations:

$$(v_z, T, \phi, q, \xi) = (v(z), \Theta(z), \varphi(z), q(z), \zeta)$$
$$\times \exp[-\lambda t + ikx]. \tag{21}$$

Here, λ is the complex perturbation decrement, and k is the wave number. We linearize equations (1)–(4) in the perturbations and write the equations and the boundary conditions in the dimensionless form:

$$\lambda(v'' - k^2 v) = v^{IV} - 2k^2 + k^4 v - \text{Ra} k^2 \left(\frac{dq_0}{dz} \phi - \frac{d\phi_0}{dz} q \right), \tag{22}$$

$$-\lambda \text{Pr} \Theta = \Theta'' - k^2 \Theta + v,$$

$$-\lambda P_j q + \frac{dq_0}{dz} v + 2q_0 q - \left(\frac{d\phi_0}{dz} q' + \frac{dq_0}{dz} \varphi' \right) = 0,$$

$$\varphi'' - k^2 \varphi = -q,$$

$$z = 1: v + \text{Pr}\lambda\zeta = 0, \quad \Theta' = 0, \tag{23}$$

$$\varphi + \zeta \frac{d\phi_0}{dz} = 0, \quad \varphi' + \zeta \frac{d^2\phi_0}{dz^2} = 0,$$

$$v'' + k^2 v - k^2 \text{Ma}(\zeta - \Theta) = 0,$$

$$\text{Cr} \left[(3k - \lambda) v' - v''' + \text{Ra} k^2 \left(q_0 \varphi + \frac{d\phi_0}{dz} \varphi' \right) \right]$$
$$= -k^2 \zeta (\text{Bo} + k^2), \tag{24}$$

$$z = 0: v = 0, \quad v' = 0, \quad \Theta = 0, \quad \varphi = 0.$$

The problem contains six dimensionless parameters: the Ra number, which is an electrical analog of the Rayleigh number; the Marangoni number Ma; the Prandtl number Pr; the Bond number Bo; the capillary parameter Cr; and the number P_j, which describes the mobility of charges,

$$\text{Ra} = \frac{\varepsilon U}{b\rho\nu}, \quad \text{Ma} = \frac{\sigma_1 \Theta h}{\rho\nu\chi}, \quad \text{Pr} = \frac{\nu}{\chi},$$
$$\text{Bo} = \frac{g\rho h^2}{\sigma_0}, \quad \text{Cr} = \frac{\rho\nu\chi}{\sigma_0 h}, \quad P_j = \frac{\nu}{bU}. \tag{25}$$

The spectral problem (22)–(24) defines the characteristic numbers λ as functions of the wave number and the parameters of the problem. The numerical solution of the differential equations was performed by the step-by-step integration using the Runge–Kutta–Merson method. To ensure the linear independence of the solutions, the orthogonalization of the partial solutions at each step of solution was used.
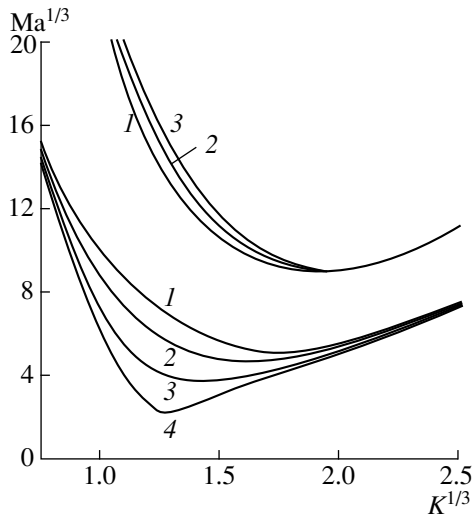
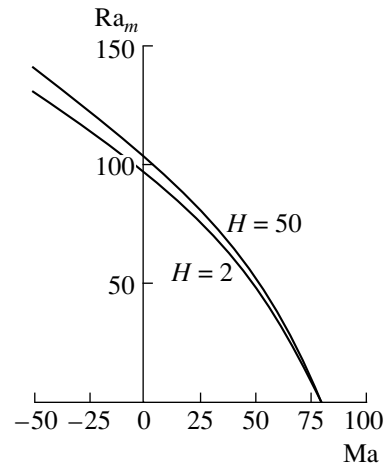**Fig. 1.** Neutral stability curves for the monotone mode.

Fig. 2. Variation of the critical Rayleigh number Ra_m as a function of the Marangoni number for the monotone instability mode. The instability region is above the curves.

4. Let us now study the stability of the state of equations (22)–(24) with respect to monotone perturbations. In the above calculations, a relation was found between three main parameters: the Ma and Ra numbers and the wave number k for the marginal stability at other parameters being fixed. By determining the minimum numbers Ra_m and k_m for a given value of the Ma number, we can determine the influence of the thermocapillary effect on the electrical mechanism of instability.

Figure 1 displays the results of calculations of several neutral curves ($\lambda = 0$) at $H = 2$, $\mathrm{Cr} = 10^{-5}$, and $\mathrm{Bo} = 10^{-1}$ for two lower modes of the injection instability and various values of the Marangoni number. Curves (*1*)–(*4*) correspond to Ma = –50, 0, 50, and 75, respectively. The upper three curves refer to the perturbations with a two-level system of cells. The instability region lies above these curves. The limiting case of Ma = 0 corresponds to the isothermal liquid subjected to injection from the free-boundary side. At Cr = 0 (flat boundary), the results obtained agree with the values reported in [2], $\mathrm{Ra}_m^{(1)} = 99$ and $k_m^{(1)} = 4.0$.

With increasing the positive values of the Marangoni number (heating from the solid-boundary side), the critical value of the electrical analog of the Rayleigh number $\mathrm{Ra}_m^{(1)}$ of the first mode is reduced and becomes zero at Ma = 80, which corresponds to the conventional problem of the development of the Marangoni convection. The threshold value of the wave number is $k_m = 2$. Upon heating from the free-boundary side (negative Marangoni numbers), an increase in the threshold value for the first mode $\mathrm{Ra}_m^{(1)}$ occurs. For the second level of instability, there occurs an increase in the minimum value of Ra_m with increasing positive Marangoni numbers, whereas at negative Ma, the $\mathrm{Ra}_m^{(2)}$ decreases.

Figure 2 displays a map of the stability with respect to the principal mode for two values of the current $J_0 \cong 1.261$ and 1.116, corresponding to the relative thicknesses of the air gap $H = 2$ and 50.

The changes in the shape of the convective cells along the vertical with changing values of the Marangoni number are shown in Fig. 3. The odd numbers at the curves mark the profiles of the vertical component of the velocity for the first instability mode; the even numbers correspond to the second mode. As is seen, the differences in the structure of motion characteristic of the first and second modes vanish with increasing absolute value of the Marangoni number.

5. An oscillatory thermocapillary instability in a layer with a deformable free boundary can only exist, as was shown in [6], when heating is performed from the free-surface side. The results shown in Fig. 1 indicate that with increasing absolute magnitudes of the negative Marangoni numbers, the neutral curve increases for the first instability mode and decreases for the second mode. At a certain negative value of the Marangoni number ($|\mathrm{Ma}| \cong 120$), they contact each other, and the layer becomes unstable with respect to oscillating perturbations. The characteristic neutral curves are given in Fig. 4a. At $\mathrm{Cr} = 10^{-5}$, $\mathrm{Bo} = \mathrm{Pr} = P_j = 10^{-1}$, curves (*1*)–(*4*) correspond to Marangoni numbers Ma = 0, –100, –150, and –250, respectively. The dashed line delineates the marginal stability to oscillatory perturbations. For the above-specified parameters, the oscillatory instability in the absence of an electric field occurs at $\mathrm{Ma}_m \cong -10^4$ and $k_m \cong 0.1$ [6]. The region of the oscillatory instability arises in a certain range of wave numbers at such values of the Marangoni number, at which the values of the Ra number corresponding to the two different instability modes become equal to one another. Note that a similar transition from two levels of a monotone instability to an oscillatory instability
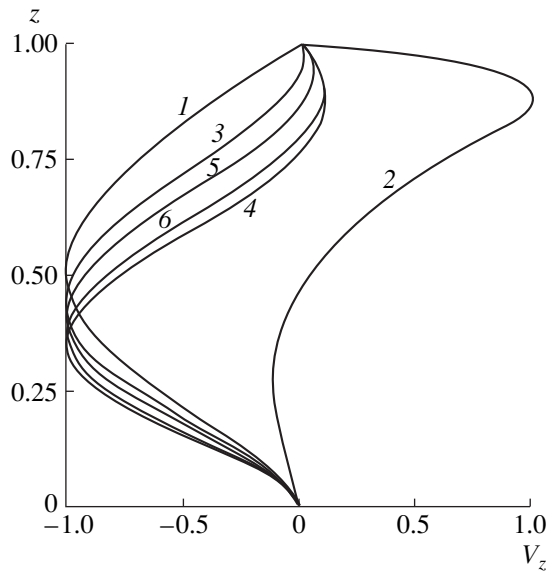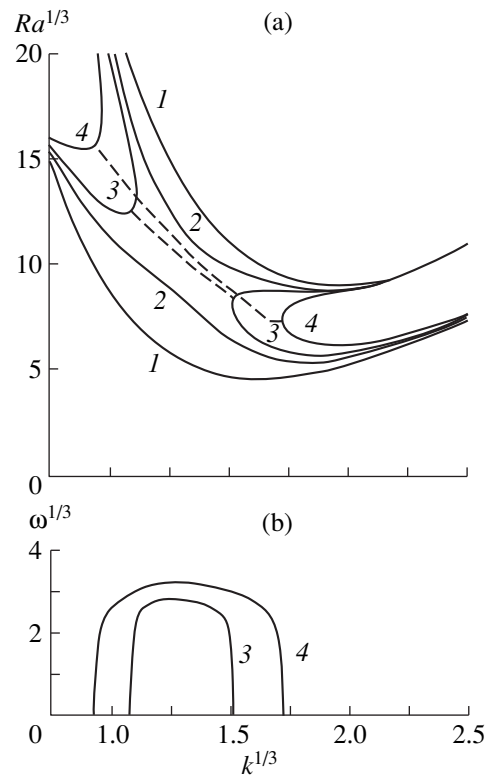
Fig. 3. The profiles of the vertical component of velocity for two lower modes of the monotone instability as a function of the coordinate z for various Marangoni numbers: (*1*, *2*) Ma = 0; (*3*, *4*) Ma = –100; (*5*, *6*) Ma = –150 at $k = 4$.

Fig. 4. (a) Neutral instability curves for two lower levels of the monotone and oscillatory modes and (b) the dispersion relation $\omega(k)$.

was also observed in the case of gravitational convective instability in a two-layered systems of immiscible liquids heated from below [9].

Figure 4b displays the wave frequencies ω as functions of the wave number for the neutral curves (*3*) and (*4*) shown in Fig. 4a. With increasing heating from the free-boundary side, the region of the oscillatory instability expands, mainly toward the shorter wavelengths. The frequency of oscillatory perturbations increases with increasing Marangoni number.

6. The above calculations show that the interaction of the electroconvective mechanism of instability of the injection and thermocapillary types upon heating of the system from the solid-surface side disturbs the stability of the system toward monotonically growing perturbations. At negative values of the Marangoni number (heating from the free-surface side), the stability of the system increases.

With a further increase of heating from the free-surface side, the critical values of the Rayleigh number for the principal monotone mode of stability and for the instability to perturbations with a two-level cell becomes closer to one another. At a certain negative value of the Marangoni number, a single oscillatory mode develops instead of two monotone instability modes. The frequency of the oscillating perturbations increases from zero with increasing Marangoni number. With increasing electrode spacing (width of the air gap) at a fixed potential difference, an increase in the threshold values of the Rayleigh number is observed.

REFERENCES

1. P. Atten and R. Moreau, J. Mech. **11** (3), 471 (1972).

2. J. M. Schneider and P. K. Watson, Phys. Fluids **19** (8), 1948 (1970).

3. P. Atten and D. Koulova-Nenova, 12th Int. Conf. on the Conduction and Breakdown in Dielectric Liquids (IEEE, Rome, 1996), pp. 476–479.

4. M. Lampert and P. Mark, *Current Injection in Solids* (Academic, New York, 1970; Mir, Moscow, 1973).

5. J. A. Pearson, J. Fluid Mech. **4** (5), 489 (1958).

6. M. Takashima, J. Phys. Soc. Jpn. **50** (8), 2745 (1981).

7. E. A. Ryabitskiĭ, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 19 (1992).

8. R. V. Birikh and R. N. Rudakov, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 5, 30 (1996).

9. G. Z. Gershuni and E. M. Zhukhovitskiĭ, Dokl. Akad. Nauk SSSR **265** (2), 302 (1982).

Translated by S. Gorin

=========================== **GASES AND FLUIDS** ===========================

Instability of a Charged Plane Interface between Two Liquids with Respect to the Tangential Discontinuity of a Time-Dependent Velocity Field

A. I. Grigor'ev

Demidov Yaroslavl' State University, Yaroslavl, 150000 Russia
Received November 10, 1998

Abstract—The differential equation that describes the evolution of perturbations of a charged plane boundary between immiscible liquids when the upper liquid moves relative to the lower one with a time-dependent velocity parallel to the boundary is the Hill equation. In this system, the interface can exhibit instabilities of three types at various values of physical parameters: the Kelvin–Helmholtz, Tonks–Frenkel, and parametric instability. When physical parameters have certain values, the interface that is unstable with respect to surface charge and the tangential discontinuity of the velocity field across the interface can be parametrically stabilized. © 2000 MAIK "Nauka/Interperiodica".

INTRODUCTION

Analyses of the Kelvin–Helmholtz and Tonks–Frenkel instabilities are motivated by numerous applications in geophysics, technical physics, and chemical technology [1–4]. One problem of current interest is the instability of a tangential discontinuity between half-spaces occupied by immiscible liquids with different densities, ρ_1 and ρ_2, in the case when the upper liquid moves with a velocity $U = U(t)$ parallel to the interface and the interface carries an electric charge characterized by a constant surface density σ. This problem combines the classical Kelvin–Helmholtz and Tonks–Frenkel instabilities with the parametric instability. It provides an adequate mathematical and physical model of St. Elmo's fire, which appears during storms in the neighborhoods of tall objects covered by water film or droplets and exposed to winds [5].

1. The analysis presented below is conducted in the simplest case of inviscid incompressible liquids. In this system, the evolution of capillary waves can be determined by solving a problem for the harmonic velocity potentials of wave motion in the upper and lower liquids, $\Psi_1(\mathbf{r}, t)$ and $\Psi_2(\mathbf{r}, t)$, in the Cartesian coordinate system with the plane XOY aligned with the unperturbed interface and the axis OZ directed downwards (parallel to the gravitational force) [3, 6, 7]:

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

$$z \longrightarrow -\infty \quad \Psi_1 - xU(t) \longrightarrow \text{const} = 0; \tag{2}$$

$$z \longrightarrow \infty \quad \Psi_2 \longrightarrow \text{const} = 0; \tag{3}$$

$$z = \xi(x; t): \frac{\partial \Psi_1}{\partial z} \approx U\frac{\partial \xi}{\partial x} + \frac{\partial \xi}{\partial t}; \tag{4}$$

$$\frac{\partial \Psi_2}{\partial z} \approx \frac{\partial \xi}{\partial t}; \tag{5}$$

$$\rho_1 \frac{\partial \Psi_1}{\partial t} + \rho_1 g\xi + \frac{1}{2}\rho_1[(\nabla\Psi_1)^2 - U^2(t)]$$
$$= \rho_2 \frac{\partial \Psi_2}{\partial t} + \rho_2 g\xi + P_\sigma - \alpha\frac{\partial^2 \xi}{\partial x^2}, \tag{6}$$

where $\rho(x, t)$ is an interface perturbation associated with thermal capillary wave motion, which can be treated as independent of the coordinate y without loss of generality [3, 6, 7]; $\mathbf{U} = \mathbf{U}(t)$ is the velocity of the upper liquid relative to the lower one, uniform with respect to x and z; the vector $\mathbf{U}$ defines the direction of the axis OX; α is the surface tension of the interface; $P_\sigma = 4\pi\sigma^2 k\xi$ is the electrostatic pressure on the interface [7] associated with the surface perturbation $z = \xi$; and k is the wavenumber.

The velocity potential of the liquid motion induced by the directional motion of the upper liquid defined by equation (7) below has the form $xU(t)$, which is taken into account in writing boundary condition (2). The total potential of the velocity field in the upper liquid has the form

$$\Psi_1(\mathbf{r}, t) = xU(t) + \Psi_1^0(\mathbf{r}, t),$$

where the component $\Psi_1^0(\mathbf{r}, t)$ describes the capillary wave motion in the upper liquid.

To derive the differential equation that describes the evolution of the amplitude of a harmonic of the thermal capillary wave driven by the electric-field pressure and the tangential discontinuity across the interface, we use the fact that the velocity potentials in both liquids van-

ish as the distance from the interface is increased to infinity. According to [7–9], this implies that $\Psi_1^0(\mathbf{r}, t) \sim \exp(kz)$ and $\Psi_1^0(\mathbf{r}, t) \sim \exp(-kz)$. This requirement leads to the relations

$$\frac{\partial \Psi_1^0}{\partial z} = k\Psi_1^0; \quad \frac{\partial \Psi_2}{\partial z} = -k\Psi_2. \tag{7}$$

Next, we note that an interface perturbation induced by capillary wave motion must have the periodic form

$$\xi \sim \exp(ikx) \tag{8}$$

and substitute (7) and (8) into boundary conditions (4) and (5) to obtain

$$\Psi_1^0(\mathbf{r}, t) = \frac{1}{k}\left(U(t)\frac{\partial \xi}{\partial x} + \frac{\partial \xi}{\partial t} \right) = \frac{1}{k}\left(ikU(t)\xi + \frac{\partial \xi}{\partial t} \right);$$

$$\Psi_1(\mathbf{r}, t) = xU(t) + \frac{1}{k}\left(ikU(t)\xi + \frac{\partial \xi}{\partial t} \right); \tag{9}$$

$$\Psi_2(\mathbf{r}, t) = -\frac{1}{k}\frac{\partial \xi}{\partial t}, \tag{10}$$

where i is the imaginary unit.

Now, we substitute (9) and (10) into (6) and use linear approximation with respect to ξ to derive the desired differential equation describing the evolution of a capillary-wave amplitude:

$$\frac{\partial^2 \xi}{\partial t^2} + 2ik\rho U(t)\frac{\partial \xi}{\partial t} + ik\rho\frac{\partial U(t)}{\partial t}\xi + \frac{k}{(\rho_2 + \rho_1)}$$

$$\times [g(\rho_2 - \rho_1) + \alpha k^2 - 4\pi\sigma^2 k - k\rho_1 U^2(t)]\xi = 0; \tag{11}$$

$$\rho = \frac{\rho_1}{(\rho_2 + \rho_1)}.$$

When $U(t) = U_0 = \text{const}$, this reduces to an ordinary differential equation with constant coefficients describing the evolution of a capillary-wave amplitude under the conditions of the Kelvin–Helmholtz and Tonks–Frenkel instabilities. Its solution is

$$\xi = \xi_0 \text{Re}\{\exp[-ik\rho U_0 t + it F^{1/2}(k)]\};$$

$$F(k) \equiv F(k, \alpha, \sigma, U_0, g, \rho_1, \rho_2)$$

$$\equiv \frac{k}{(\rho_2 + \rho_2)}[g(\rho_2 - \rho_1) + \alpha k^2 \tag{12}$$

$$- 4\pi\sigma^2 k\xi - k\rho_1 U_0^2 + kU_0^2\rho\rho_1].$$

The criteria for both instabilities are determined by setting the function $F(k)$ to zero. They can be readily obtained by means of a standard analysis (see a detailed description in [6, 7]).

2. To make the analysis that follows more specific, consider a particular form of the time dependence of the upper-liquid velocity:

$$\mathbf{U}(t) = \mathbf{U}_0 + \mathbf{U}_* \cos \omega_0 t \quad (\mathbf{U}_0 \parallel \mathbf{U}_*). \tag{13}$$

Substituting (13) into (11), we obtain a differential equation with variable coefficients that characterizes the evolution of a capillary-wave amplitude. To find its solutions, we perform a change of the dependent variable [9]:

$$\xi(t) = \zeta(t)\exp[i\Phi(t)];$$

$$\Phi(t) = -k\rho\left(U_0 - \frac{U_*}{\omega_0}\sin \omega_0 t \right). \tag{14}$$

As a result, (11) is rewritten as

$$\frac{\partial^2 \zeta}{\partial t^2} + \zeta F(k) - \zeta D(k)\cos \omega_0 t + \zeta L(k)\cos^2 \omega_0 t = 0;$$

$$D(k) \equiv D(k, U_0, U_*, \rho_1, \rho_2) \equiv 2k^2\frac{\rho_1\rho_2}{(\rho_2 + \rho_1)^2}U_0 U_*; \tag{15}$$

$$L(k) \equiv L(k, U_*, \rho_1, \rho_2) \equiv k^2\frac{\rho_1\rho_2}{(\rho_2 + \rho_1)^2}U_*^2.$$

When U_0 and U_* are arbitrary nonzero quantities, the resulting equation with time-dependent coefficients is the Hill equation. It has either parametrically stable or parametrically unstable (exponentially increasing with time) solutions, depending on the relation between the coefficients F, D, and L and the frequency ω_0. In the inviscid approximation considered here, the parametric instability develops when the amplitude U_* of the variable velocity component is arbitrarily small. In a real liquid, the parametric instability will manifest itself starting from a certain threshold value of U_* depending on the capillary-wave wavelength and the liquid viscosity [9–13].

When $U_0 = 0$, the upper liquid executes an oscillatory motion parallel to the interface. Under this condition, $D(k) = 0$ in equation (5). Therefore, one obtains the Mathieu equation

$$\frac{\partial^2 \zeta}{\partial t^2} + \omega^2(1 + h\cos 2\omega_0 t)\zeta = 0; \tag{16}$$

$$\omega^2 \equiv F(k) + 0.5L(k); \quad h \equiv 0.5L(k)\omega^{-2}.$$

The domains where it has stable and unstable solutions have been analyzed and described in sufficient detail (e.g., see [10]). A detailed analysis of the solution to equation (15) for $D(k) = 0$ was presented in [9]. In particular, these analyses have demonstrated that, when $h \ll 1$, the instability domains of solutions to (16) are narrow zones in the neighborhoods of the points defined by the equation $(2\omega/\omega_0) = n$, where n is an inte-

ger. The widths of the instability zones, being determined by the parameter h, rapidly decrease with increasing n. In the second-order approximation with respect to h, the first three zones of instability are determined by the relations (see [11])

$$1 - \frac{1}{2}h + \frac{7}{32}h^2 \ll \left(\frac{2\omega}{\omega_0}\right)^2 \ll 1 + \frac{1}{2}h + \frac{7}{32}h^2;$$

$$4 + \frac{3}{2}h^2 - h^2 \ll \left(\frac{2\omega}{\omega_0}\right)^2 \ll 4 + \frac{3}{2}h^2 + h^2;$$

$$9 + \frac{81}{64}h^2 - \frac{3^6}{2^9}h^3 \ll \left(\frac{2\omega}{\omega_0}\right)^2 \ll 9 + \frac{81}{64}h^2 + \frac{3^6}{2^9}h^3.$$

Here, ω should be interpreted as a capillary-wave frequency. When the lower liquid is treated as a viscous one characterized by a kinematic viscosity ν, in which case the capillary-wave decrement is given by the well-known expression $\gamma \equiv 2\nu k^2$ [6], there exists a threshold value of U_* (as indicated above), or an equivalent threshold value of h for equation (16), which corresponds to the critical condition for the parametric instability. The threshold values increase with the instability-zone number n. For the first three zones, the critical values of h are determined by the following relations derived in [11]:

for the first zone,

$$h > 4\gamma\omega_0^{-1};$$

for the second zone,

$$h > 2(2\gamma\omega_0^{-1})^{1/2};$$

for the third zone,

$$h > \frac{8}{3}\left(\frac{2\gamma}{3\omega_0}\right)^{1/3}.$$

Since h is treated here as a small parameter, these relations are applicable only when the decrement γ is small.

When $U_0 \gg U_*$, equation (15) can be linearized with respect to the small parameter (U_*/U_0) and, once again, reduced to a Mathieu equation similar to (16), but with a parametric-excitation frequency lower by a factor of 2. In this case, the results obtained above for solutions to (16) remain valid up to the corrections due to the lower value of the parametric-excitation frequency.

When $U_* = 0$ (i.e., when the relative velocity of the upper liquid is constant), the coefficients D and L vanish, (15) becomes an ordinary differential equation, and its solution, which can be used to determine the critical conditions for the Kelvin–Helmholtz and Tonks–Frenkel instabilities corresponding to $F(k) \le 0$ ($\omega^2 \le 0$) is identical to (12) by virtue of (14).

In the general case, when neither U_0 nor U_* vanishes, solutions to the Hill equation (15) are also unstable in the greater part of the domain $\omega^2 \le 0$ (where capillary waves are essentially unstable in the absence of parametric excitation, because the Kelvin–Helmholtz and Tonks–Frenkel instabilities develop when $\omega^2 \le 0$). Nonetheless, there exists a domain of stable solutions to equation (15) [14] in the vicinity of the origin in the coordinate plane (U_*, ω^2) when $\omega^2 \le 0$. The capillary waves that are unstable in the Kelvin–Helmholtz and Tonks–Frenkel sense are parametrically stabilized in this domain. This observation suggests that the unstable behavior can be controlled and underlines the importance of the problem analyzed here.

CONCLUSION

When a tangential discontinuity of velocity across a charged interface between two liquids has a time-dependent component, the domain of instabilities that can develop in the system may expand, and the critical conditions for the instabilities can be controlled through parametric damping of unstable capillary waves.

REFERENCES

1. V. P. Vlasov, S. K. Zhdanov, and B. A. Trubnikov, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza, No. 3, 10 (1991).
2. A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 3 (1994).
3. O. A. Grigor'ev and S. O. Shiryaeva, Zh. Tekh. Fiz. **66** (2), 23 (1996).
4. P. M. Lushnikov, Izv. Ross. Akad. Nauk, Fiz. Atmos. Okeana **34,** 413 (1998).
5. A. I. Grigor'ev, I. D. Grigor'eva, and S. O. Shiryaeva, J. Sci. Expl. **5,** 163 (1991).
6. L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, 2nd ed. (Nauka, Moscow, 1986; Pergamon, Oxford, 1987).
7. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, 2nd ed. (Nauka, Moscow, 1982; Elsevier, New York, 1985).
8. V. G. Levich, *Physicochemical Hydrodynamics* (Fizmatgiz, Moscow, 1959; Prentice-Hall, Englewood Cliffs, 1962).
9. D. V. Lyubimov, M. V. Khenner, and M. M. Shots, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 3, 25 (1998).
10. *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*, Ed. by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).
11. N. N. Bogolyubov and Yu. A. Mitropol'skiĭ, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Nauka, Moscow, 1974).
12. A. I. Grigor'ev, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza, No. 1, 50 (1989).
13. S. O. Shiryaeva and A. I. Grigor'ev, Zh. Tekh. Fiz. **62** (11), 49 (1992).
14. D. R. Merkin, *Introduction to the Stability Theory* (Nauka, Moscow, 1987; Springer-Verlag, New York, 1997).

===== GAS DISCHARGES, PLASMA =====

Application of Wide-Aperture Resonators with Small Diffraction Loss to Diagnostics of a Tokamak Plasma by the Laser Photoionization Technique

Yu. V. Petrov, G. T. Razdobarin, and S. Yu. Tolstyakov

Ioffe Institute for Physics and Technology, Russian Academy of Sciences, St. Petersburg, 194021 Russia

Received September 28, 1998

Abstract—To increase the sensitivity of the photoionization diagnostics, efficient laser systems for plasma probing are developed based on stable resonator configurations with small diffraction loss. The systems meet the requirements for multipass intracavity probing and make it possible to increase the resonator length to conform to a large size of the tokamak chamber. Demonstration experiments were carried out in the FT-1 tokamak. The results obtained confirm the possibility of reliable measurements of the hydrogen neutral density in the range 10^8–10^9 cm^{-3} in tokamak plasmas. © *2000 MAIK "Nauka/Interperiodica".*

1. DESIGN PRINCIPLES OF LASER DIAGNOSTIC SYSTEMS

In previous papers [1, 2], we presented experiments with the use of a new diagnostic technique based on laser photoionization of excited hydrogen atoms in a tokamak plasma. For further development of these diagnostics, it is necessary to improve the accuracy of measuring a weak light signal produced by photoionization in the presence of the intrinsic plasma radiation, which is attained by increasing the probing energy. It is worthwhile to increase the energy even under the saturation conditions, when the recorded photoionization signal becomes independent of the probing power. In this case, the sensitivity of the diagnostics can be increased by increasing the generation duration, the number of generated pulses, and the number of emitting atoms in the probing plasma volume due to an increase in the diameter of the laser beam. To suit these requirements, laser systems based on resonators with small diffraction loss and tailored for diagnostic experiments in tokamaks were developed.

The development of these systems is based on two principles. The first principle is a multipass plasma probing with the use of a pair of mirrors situated on each side of the discharge chamber of a tokamak. In this case, the probing energy increases with the number of passes. The second principle is to introduce a plasma under study into the closed laser cavity. This allows us to obtain a low generation threshold because of small end loss and, consequently, to increase the generation energy. Both principles can be realized separately; however, the most effect can be attained if these principles are used simultaneously. It is suggested that there are at least two coupled resonator sections, one of which is intended to contain the active medium, and another one is intended to contain the diagnosed plasma. Such a separation of the resonator to the sections coupled with the help of a matching lens ensures the generation regimes with smallest loss. The higher order modes provide the worst confinement of a beam. As a result, the diffraction loss per pass has a sharp mode boundary, which depends on the Frenel number N. For the confocal configuration, the loss near the boundary mode j_0 can be presented as a function of the Fresnel number $e^{-(4\pi N - j_0)}$ [3]. For $N \gg 1$, the loss almost vanishes as soon as the length of the resonator section reduces by half and, accordingly, the Fresnel number increases by half. Thus, the increase in the number of sections, provided that the total length of the composite resonator remains the same, leads to the proportional increase in the Fresnel number for each section and is beneficial for energy confinement, especially, in the case of an inhomogeneous active medium.

2. ARRANGEMENT AND TEST RESULTS OF DIAGNOSTIC LASER SYSTEMS

2.1. System with a Semiconfocal Resonator

The basic requirements of the systems with multipass intracavity plasma probing are that they should allow increasing the resonator length to conform to a large size of the tokamak discharge chamber and they should have a low generation threshold. These two conflicting requirements can be satisfied if most stable resonator configurations close to the confocal ones [4] are used. Two versions of the system in which the tokamak plasma is situated inside a long semiconfocal resonator are shown in Fig. 1. Here, the role of a spherical mirror is played by a catadioptric system consisting of a flat
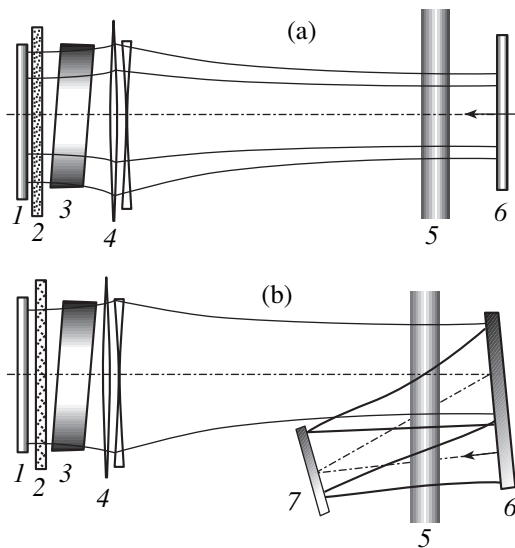
Fig. 1. Optical scheme of the semiconfocal resonator (a) in the absence and (b) in the presence of the multipass probing system: (*1*) flat mirror, (*2*) phototropic shutter, (*3*) active element, (*4*) objective, (*5*) plasma volume, and (*6, 7*) flat mirrors.
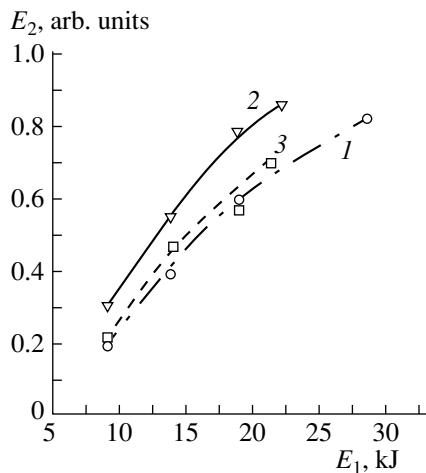
Fig. 2. Total probing energy E_2 for (*1*) 4, (*2*) 6, and (*3*) 8 passes of the beam through the plasma volume versus the pumping energy E_1.

mirror *1* and a matching lens *4*. An active element from $\varnothing 45 \times 300$ neodimium phosphate glass and a phototropic shutter with an initial transmission of nearly 90% are positioned between the lens and the mirror. By varying the focal length of the matching lens, it is possible to vary the system configuration over a wide range from the concentric to the flat one. The maximum generation occurs when the confocality conditions are accurately fulfilled, and the focus of the catadioptric system is made coincident with a flat mirror *6*. The scheme of the semiconfocal resonator with a total length of ~10 m, which is shown in Fig. 1b, meets the requirements for the multipass plasma probing. This is

achieved by zigzag ray trajectories inside the resonator. In the figure, six (three direct and three reverse) passes of the ray through the plasma volume are shown; however, the system was easily adjusted for four and eight passes. In the confocal scheme, we obtained the multipulse generation regime, in which the number of pulses in the train during the pumping was above 10 and the energy of each pulse was up to 6.5 J. In this case, the individual pulses were amplitude modulated with a period of ~70 ns in accordance with the time of phase-conjugation of the wave front in the resonator. The number of the amplitude-modulation periods reached 30 during the pulse of ~2 μs full duration. Such a great number of the laser beam passes through the resonator is evidence of a low level of divergence loss. The low loss level is also evidenced by a low generation threshold. Compared to a resonator of flat configuration used in the first experiments [1, 2], the threshold pumping energy reduced from 15 to 2.7 kJ. However, a deterioration was observed in energy confinement if the resonator length was longer than 10 m, which corresponded to more than 6 passes of the probing beam through the plasma. This effect is more typical for the regimes with a high pumping power, when thermal distortion leads to violation of the confocality condition. It is seen from Fig. 2 that, for the number of passes between 4 and 6, the gain in the total pumping energy is about 30%. As the number of passes is further increased (to eight), even a somewhat decrease is observed in the total probing energy. Thus, the increase in selective loss due to thermal aberrations did not allow us to efficiently use the scheme when the number of the probing-beam passes was increased above six. Nevertheless, in spite of the limitation of the number of passes, the presented scheme of the semiconfocal resonator has an evident advantage over the flat resonator configuration with respect to the values of the threshold pumping energy and the total energy of plasma probing.

2.2. System with a Composite Confocal Resonator

The schemes of resonators shown in Figs. 3 and 4 use the principle of sectioned composite resonator. Figure 3 illustrates the scheme of coupling of two resonator sections. If the focus of a lens *2* coincides with a flat mirror *3*, then the position of waists is uniquely determined by the focal lens planes [5]. Minimum loss in the section *I* containing a plasma is attained if we form the confocal configuration ($R_0 = 2L_0$) by making the focuses of the spherical mirror *1* and the lens *2* coincident.

Writing the confocality parameter [3] of a Gaussian beam in terms of the curvature radii of wave fronts at the spherical mirror (section *I*) and the lens (section *II*), we obtain the coupling equation for the section lengths and the corresponding radii of curvature

$$b = 2\sqrt{L_1(R_1 - L_1)} = 2\sqrt{L_0(R_0 - L_0)} = 2L_0. \quad (1)$$

In the special case of equal lengths $L_0 = L_1 = f$, equation (1) has a unique solution $R_1 = 2L_1$, and the configuration becomes symmetric if the field distribution in all three resonator sections is exactly the same. In the case $L_0/L_1 \gg 1$ or $L_0/L_1 \ll 1$, the solutions for section II correspond to either the flat or concentric resonator configurations. The ratio of the beam diameters at both mirrors and in the waists can be easily calculated. From the known law of propagation of a Gaussian beam [3] and the lens formula, we can derive the ratio of the beam diameters for any pair of the resonator components. Thus, if the beam cross size is limited by an aperture diaphragm d at the flat mirror 1, the beam diameters at the other components are equal to $D = d\sqrt{1 + (L_0/f)^2}$, $d_0 = dL_0/f$, and $D_0 = d\sqrt{2}\,L_0/f$, respectively. The preference should be given to the configuration in which the beam diameter best matches the limiting diaphragms. Tokamak windows situated inside the section I and the working diameter of the active element situated in section III play the role of such diaphragms.

For testing, we prepared a scheme, presented in Fig. 4, with a coupling lens of focus $f = 2$ m and a spherical mirror with a curvature radius $R = 3.3$ m. An active element 3 and a phototropic shutter 2 were positioned near a flat mirror 1. The total length of the composite resonator, with allowance for the refraction of the active medium, was 5.76 m. In our case, the beam was limited by the size of the active element $d = 45$ mm, and the beam diameters at the lens, in the waist, and at the spherical mirror, according to the above expressions, were equal to $D = 58$ mm, $d_0 = 37$ mm, and $D_0 = 53$ mm.

In the optical scheme shown in Fig. 4b, there is an additional component, which represents a flat mirror 7 used for multipass plasma probing. The flat mirror is placed in the focal plane of a spherical mirror 6, thereby forming an additional section representing a semiconfocal resonator. The multipass system is adjusted for four additional passes of the probing beam through the plasma volume. Unlike in the configuration presented in Fig. 1b, the implementation of the multipass probing in this sectioned system does not involve a decrease in the Fresnel number and diffraction loss. The loss increases primarily due to the additional optical components and can be reduced to minimum by employing highly reflecting coatings and antireflection lens optics. This is confirmed by the results of measurements of the number of generated pulses for various values of the pumping power (Fig. 5). As is seen, the presence of the multipass system has little effect on the number of generated pulses and, consequently, does not lead to a noticeable increase in energy loss. As a result, the multipass probing gives the evident gain in the plasma-probing energy. The saturation observed in the probing energy at high pumping energies is a consequence of the formation of a thermal lens that leads to the violation of the confocality condition.
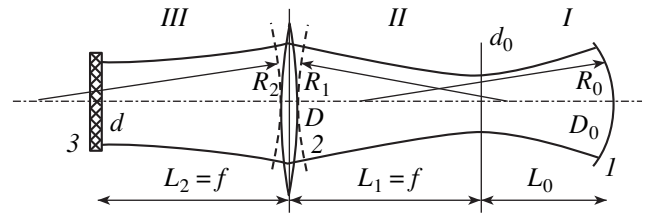
Fig. 3. Composite resonator: (*1*) spherical mirror, (*2*) lens, and (*3*) flat mirror; R_0 is the curvature radius of the spherical mirror, R_1 and R_2 are the curvature radii of the beam wave fronts on the right and left from the lens, respectively; L_0, L_1, and L_2 are the lengths of the resonator sections I, II, and III, respectively; D and D_0 are the beam diameters at the lens and the spherical mirror, respectively; d and d_0 are the beam-waist diameters, and f is the focal length of the lens.
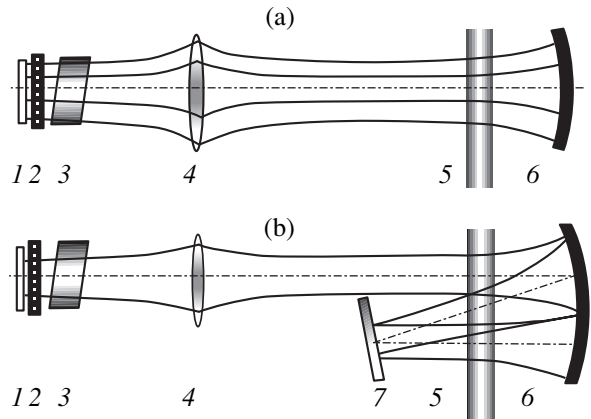

Fig. 4. Optical scheme of the composite confocal resonator (a) in the absence and (b) in the presence of the multipass probing system: (*1*) flat mirror, (*2*) phototropic shutter, (*3*) active element, (*4*) lens, (*5*) plasma volume, (*6*) spherical mirror, and (*7*) flat mirror.

The scheme demonstrated the low level of selective loss, which resulted in the low value of the threshold pumping energy ~1.9 kJ. This made it possible to realize the multipulse regime with a great number of generated pulses at sufficiently low values of the pumping energy. For a given resonator length, it is possible to vary the diameters of the beam at the components of this system over a wide range without decreasing the generation energy and, thereby, to achieve the optimum filling of the active element and the tokamak windows by radiation.

3. DEMONSTRATION EXPERIMENTS IN A TOKAMAK

Demonstration experiments were carried out in the FT-1 tokamak ($R = 62.5$ cm, $a = 15$ cm) under the standard discharge conditions with the use of the scheme of recording the emission of the H_β line, presented in the previous papers [1, 2]. The scheme of intracavity multipass plasma probing with the use of the semiconfocal resonator with a total length of ~10 m is shown in
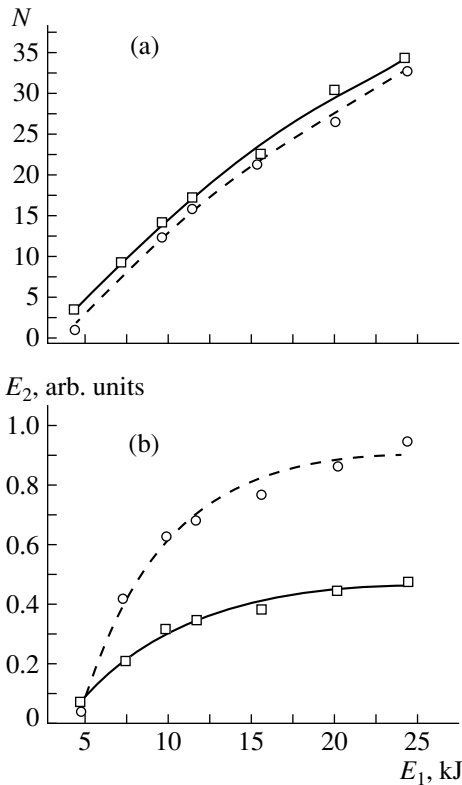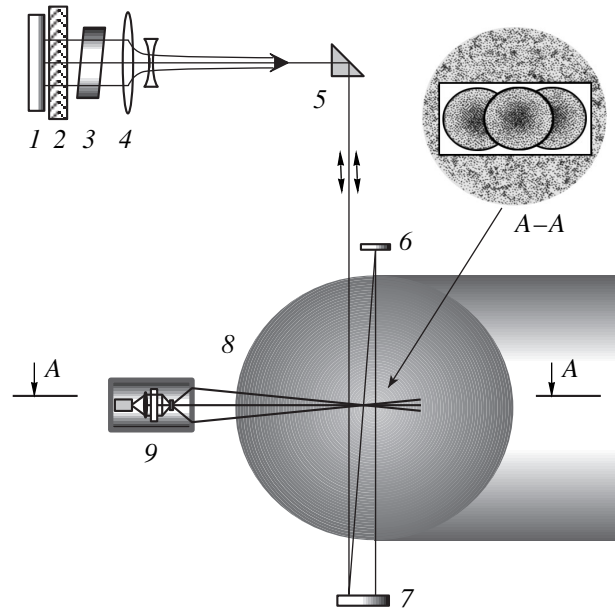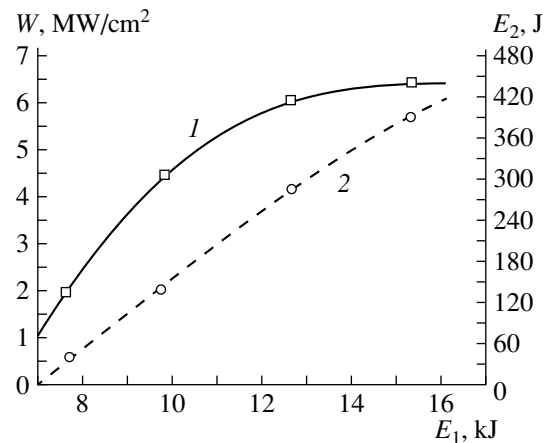
Fig. 5. (a) Number of generated pulses and (b) total probing energy for the composite confocal resonator in the presence (dashed curve) and absence (solid curve) of the multipass system.

Fig. 6. Schematic of the experiment in the FT-1 tokamak with the use of a semiconfocal resonator: (*1*) flat mirror, (*2*) phototropic shutter, (*3*) active element, (*4*) objective, (*5*) deflecting prism; (*6, 7*) flat mirrors, (*8*) plasma, (*9*) light-collecting system; *A–A* is the projection of the beam cross section onto the center of the plasma column for 6 passes of plasma probing.

Fig. 6. Here, it is also shown how the tokamak windows with a 20×40 mm cross section are filled by radiation for 6 beam passes through a plasma. In spite of an additional nonselective loss arising because the tokamak chamber is placed inside the resonator, we obtained the high parameters of probing radiation. At a moderate pumping energy of 15 kJ, the number of pulses in the train was $N \geq 10$, and the pulse energy in a single was 5–6 J. In Fig. 7, the peak value of the power density of probing radiation and the total probing energy in the multipass system are shown as functions of the pumping energy. The observed saturation of the power density corresponds to the increase in the number of generated pulses. The power-density level attains 6.5 MW/cm^2, which markedly exceeds a saturation value of ~2 MW/cm^2 under the discharge conditions of the FT-1 tokamak [2]. The obtained generation parameters significantly improve the diagnostic potentialities of the method in comparison with the previous experiments [1, 2]. In the previous measurements, the energy of individual pulses was 3–5 J for a maximum pumping energy of ~40 kJ. In addition, the duration of probing pulses increased by more than twofold and reached ~1 μs at a half-maximum.

An increase in the duration of probing can substantially increase the sensitivity of the diagnostics. As a

Fig. 7. (*1*) Power density of probing radiation and (*2*) probing energy as functions of the pumping energy.

result, it became possible to record signals for each pulse of the train of laser pulses in the presence of the plasma noise. This is seen from Fig. 8, in which the oscillograms of the photoionization signals corresponding to the sequence of laser pulses generated during the pumping time are presented. These results of measurements correspond to the plasma region in which the density of unexcited hydrogen atoms are below 10^{10} cm^{-3}. The averaging of signals over a series of pulses in the train allows us to carry out reliable mea-
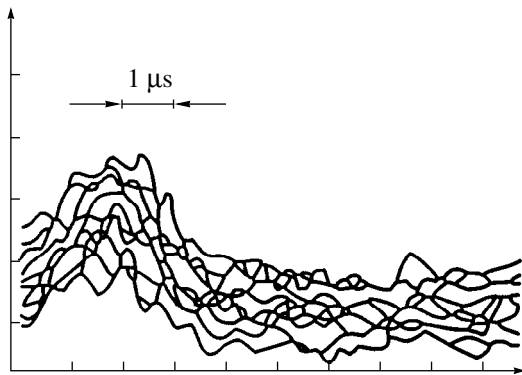
Fig. 8. Oscillograms of the photoionization signal.

ters. Since the confocal resonators have a small diffraction loss, less stringent requirements are imposed on the inhomogeneity of the active element. As a result, the scheme becomes more stable against thermal optical distortions and misalignment. In the experiments, a marked gain was obtained in the threshold pumping energy and the duration and energy of probing pulses. This makes it possible to carry out reliable measurements in the density range 10^8–10^9 cm^{-3} typical of the central region of the plasma column.

surements in the density range 10^8–10^9 cm^{-3} typical of the central region of the plasma column.

4. CONCLUSION

A new diagnostic system is developed for measuring the neutral component of tokamak plasmas by the laser photoionization technique. The testing experiments carried out in the FT-1 tokamak demonstrated the advantage of the laser systems based on the confocal resonator configurations over the flat resonator configuration with respect to the most important parame-

REFERENCES

1. V. K. Gusev, A. V. Dech, D. V. Kuprienko *et al.*, Pis'ma Zh. Tekh. Fiz. **21**, 32 (1995).

2. V. I. Gladuschak, V. K. Gusev, M. Yu. Kantor *et al.*, Nuclear Fusion **35**, 1385 (1995).

3. H. Kogelnik and T. Li, Appl. Opt. **5**, 1550 (1966).

4. Yu. A. Anan'ev, *Laser Resonators and Beam Divergence Problem* (Adam Hilder, New York, 1992).

5. A. L. Mikaélyan, M. L. Ter-Mikaélyan, and Yu. G. Turkov, *Solid-State Optical Oscillators* (Sov. Radio, Moscow, 1967).

Translated by N. F. Larionova

=========== **SOLIDS** ===========

Mesoscopic- and Macroscopic-Level Plastic Deformation and Fracture of Ion-Modified $Ni_{50}Ti_{40}Zr_{10}$ Alloy with the Shape-Memory Effect

L. L. Meisner, V. P. Sivokha, Yu. P. Sharkeev, S. N. Ku'lkov, and B. N. Gritsenko

Tomsk Institute for Physics of Strength and Materials Science, Siberian Division,
Russian Academy of Sciences, Tomsk, 634021 Russia

Received October 1, 1998

Abstract—The influence of the ion-modified surface layer on the ductility, shape-memory effect, and mesoscopic structure of the surface of fracture was studied for $Ni_{50}Ti_{40}Zr_{10}$ alloy. It was found that ion implantation increases the microhardness of the surface layer about 1 μm thick but makes the bulk of the alloy more ductile. The mesoscopic structures of the fracture surfaces of the initial and implanted specimens suggest the formation of a specific layer immediately under the irradiated surface. Its thickness is a multiple of the grain size of the $B2$ phase. The mesorelief of this layer and its ductile properties differ from those of the nonirradiated specimens. After such a surface treatment, the shape memory effect parameters do not degrade, and the temperature cycling resistance is even improved. © 2000 MAIK "Nauka/Interperiodica".

INTRODUCTION

It is well known that deformation in alloys with the shape memory effect depends largely on the ductility of the initial phase. In a low-ductility material, the formation of martensite due to stress or temperature variations, as well as the reorientation or retwinning of martensite grains, may give rise to grain-boundary cracking, lower the resistance to temperature cycling, and cause premature fracture [1]. The initial ductile parameters of a material can be changed by modifying the microstructure and chemical composition of the surface.

High-dose implantation of ions into the surface is an efficient technique to strengthen metallic materials. The modified phase and structure of the near-surface layers change such mechanical properties of the irradiated materials as microhardness, wear resistance, and coefficient of friction. Such a treatment, impeding the formation and propagation of microcracks, significantly affects the development of plastic deformation and failure of the materials on all characteristic scales (i.e., on micro-, meso-, and macroscopic levels) [2].

The deformation of alloys with the shape memory effect (SME alloys) under loading has been extensively studied on the microscopic level [3–5]. The formation of a thin modified near-surface layer in these alloys would be expected to produce a significant effect both on the overall ductility (macroscopic level) and on their ductile properties on smaller, micro- and mesoscopic, levels. In the literature, there are no data as to how thin modified near-surface layers influence the evolution of plastic deformation and fracture of SME materials on different scales. This paper is aimed at studying the effects of the modified surface layers on the ductility, SME, and postfracture evolution of the $Ni_{50}Ti_{40}Zr_{10}$ strained structure. It is also important to reveal the characteristic features of the strained structure on different scale levels.

## MATERIALS AND EXPERIMENTAL TECHNIQUE

The choice of $Ni_{50}Ti_{40}Zr_{10}$ alloy was dictated by the following: (1) in this alloy, martensite transformation responsible for the shape-memory effect proceeds according to the simple scheme $B2 \rightleftarrows B19'$, (2) the characteristic temperatures of the martensite transformation and shape-memory effect lie within the temperature ranges convenient for study, and (3) the alloy is promising for applications. Earlier [4, 5], we studied in detail "high-temperature" SME $Ni_{50}Ti_{50-x}Zr_x$ alloys. It was found that the fundamental inelastic properties of this system are the most pronounced for the composition $Ni_{50}Ti_{40}Zr_{10}$ and that the alloy offers high workability.

As an implant, we chose molybdenum, which forms a continuous series of solid solutions with β-Ti titanium [6] and is known as one of the best plasticizers of titanium-based β alloys [7].

The preparation of ingots and specimens are described in [4]. In our work, the specimens were needles (average size $1 \times 1 \times 20$ mm) and flat platelets ($1 \times 10 \times 20$ mm). Their surfaces were electrolytically prepolished in a hot 89 ml H_2O + 2 ml HF + 2 ml HNO_3 solution.

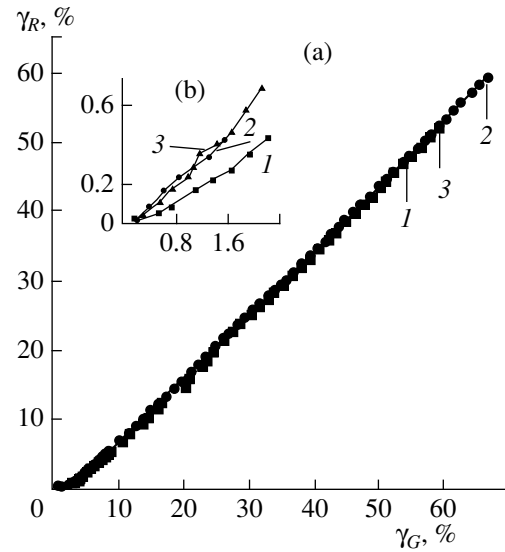The ion processing of the surface was performed with a Diana-2 vacuum arc pulsed ion source. Molyb-
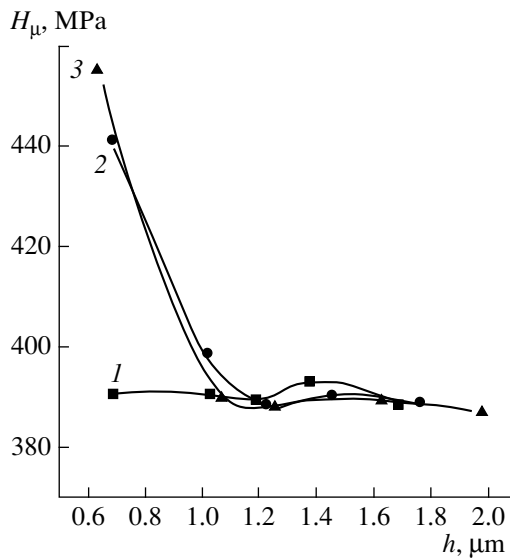
Fig. 1. Alloy microhardness versus indentation depth: (*1*) starting specimen, (*2*) first-dose, and (*3*) second-dose implantation.

Fig. 2. Residual plastic strain versus applied load for loading–unloading cycles under torsion: (a) general view and (b) initial portion of the curves (small loads). For (*1*)–(*3*), see Fig. 1.

denum ions were implanted at an accelerating voltage of 50 kV with a pulse repetition rate of 50 pulses per second. The first and second rated implantation doses were 10^{17} and 5×10^{17} ion cm^{-2}, respectively. During the implantation, the specimen temperature did not exceed 100–150°C. Two side faces of the needles and one side of the platelet were treated simultaneously. The needles were clamped into a package so as to provide complete processing of their side faces.

The mechanical properties were studied with a Reverse Torsion Pendulum setup in the torsion mode. This technique makes it possible to study the mechanical properties of surface layers, which are the first to deform during torsion. Three groups of needlelike specimens (three per group) were tested: with the initial (unmodified) surface, with the side surfaces irradiated by the first dose, and with the side surfaces irradiated by the second doses. The test temperature was maintained above the point of direct strain martensite transformation (M_d) and was equal to 170°C. This allowed us to maintain the single-phase $B2$ state in the specimens and eliminate the effect of martensite-transformation-related inelastic components on deformation.

In the experiment, a needle was placed into the setup and loaded (twisted through a certain fixed angle. Then, the specimen was unloaded, and the angle set after partial stress relief was measured. The general, γ_G, and residual, γ_R, strains were calculated from the relationship

$$\gamma = \frac{d\Delta n}{2lR},$$

where d and l are the mean values of the specimen diameter and length, respectively; $R \equiv$ const is a constant accounting for the setup geometry; and Δn is the difference between the maximum twist and that set in after unloading (in the units of the setup scale).

We repeated the loading and unloading procedures, increasing the twist to specimen failure.

The microhardness of the starting and irradiated alloys was measured on the platelets by a PMT-3M1 instrument. The indentation load was varied within the range 0.02–0.3 N.

To determine the grain size in the starting material, the platelets were ground, electrolytically polished, and then ion-etched in a VUP-5 general-purpose vacuum unit. The surface microstructure and the grain sizes were studied with a Neofit-32 optical television complex. The surfaces of fracture were examined by an SEM-200 scanning electron microscope.

EXPERIMENTAL RESULTS

In the initial state, the alloy quenched from 800°C had the following phase composition: ~90 vol % $B2$ + ~10 vol % ($B19'$ + secondary phases). The Vickers microhardness of the alloy was about 40 kg/mm^2, and the mean grain size of the main $B2$ phase was 30–40 μm.

The implantation of molybdenum ions significantly improved the surface microhardness of the $Ni_{50}Ti_{40}Zr_{10}$ alloy as compared to that for the initial state (Fig. 1). At small indentation loads, an increase in the microhardness is related to the contribution from the near-surface, irradiated layers. At higher loads (≥0.1 N), the contri-
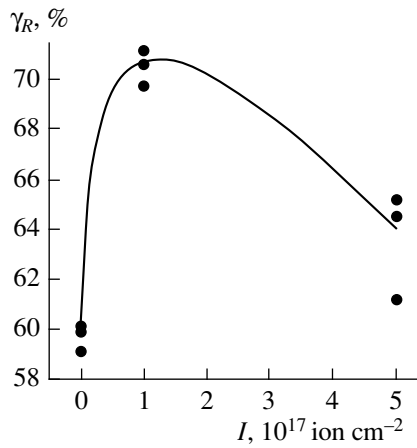
Fig. 3. Maximum plastic deformation $\gamma_R^{\max}$ to fracture in the $\mathrm{Ni_{50}Ti_{40}Zr_{10}}$ alloy versus ion-beam intensity.

250 μm
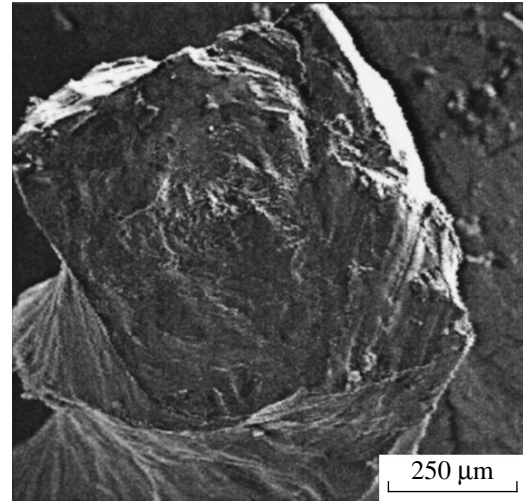
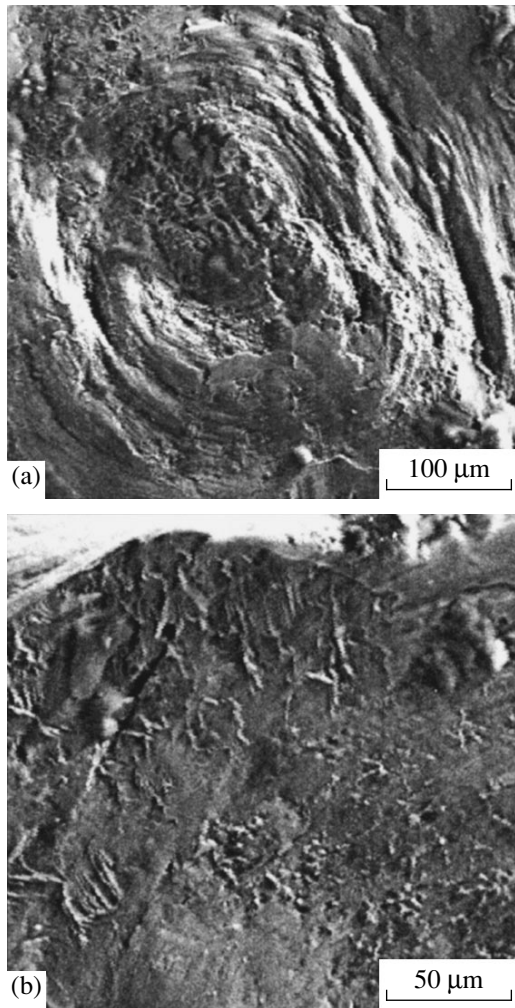**Fig. 4.** Surface of fracture of the starting alloy.

bution from the lower layers is dominant. The microstructure of the latter must differ from that in the initial state [2]. Owing to the "long-range interaction effect," a specific microcrystalline state with a modified dislocation structure is formed in these layers. In such a zone, the ductility, strength, microhardness, etc., change. According to the $H_\mu(h)$ plots presented in Fig. 1, the depth at which the modified microcrystalline layer forms in $\mathrm{Ni_{50}Ti_{40}Zr_{10}}$ is about 1 μm and does not depend on the implantation dose. At loads exceeding 0.2 N, the microhardness is determined mainly by the still lower layers with the unchanged microscopic and mesoscopic structures rather than by the doped layer and the layer with the modified microstructure. In this situation, the $H_\mu(h)$ plots for the irradiated and nonirradiated specimens coincide (Fig. 1).

The results of the torsion tests as the strain (γ_R)–stress (γ_G) curves are presented in Fig. 2. Note that the experimental values of γ_R and γ_G coincide to within the experimental error in all the tests both for the starting and for the irradiated specimens. It turned out that the $\gamma_R(\gamma_G)$ curves for the starting and irradiated specimens run in a similar way, except for the extreme regions (the beginning and the end of the tests for each of the groups)—see Fig. 2a. In the irradiated specimens, the residual strains appear at the lower applied load and the rate of their accumulation in the low-load region (at γ_G ranging from zero to about 5%) is higher than in the starting specimens (Fig. 2b). For all types of the specimens, the accumulation of residual strains in this load range is a nonlinear process. When γ_G exceeds 5%, all the $\gamma_R(\gamma_G)$ curves for the starting and irradiated specimens coincide and the residual strains are accumulated linearly. An important feature of the $\gamma_R(\gamma_G)$ curves is the significantly different maximum values, $\gamma_R^{\max}$, of the residual strains accumulated before fracture (these values for the three kinds of surface processing are indicated by arrows in Fig. 2a). The values of $\gamma_R^{\max}$ for the irradiated specimens turned out to be higher than those for the nonirradiated ones by 10% on average. The larger values were found for the specimens irradiated by the first dose (Fig. 3). Thus, the $\gamma_R(\gamma_G)$ curves suggest that the accumulation of the residual strain after the implantation proceeds in the same manner. At the same time, the strain ranges where $\gamma_R \geq 0$ expand toward lower values of γ_G at small loads and toward higher values of γ_G at large loads. These results imply that the ductility of the irradiated specimens becomes appreciably higher as compared to the starting specimens.

The surface of fracture of the starting specimen suggests that they underwent plastic intergranular fracture. Figure 4 clearly shows pitting in the center of the specimen, which is typical of such fracture, and the evidence of strong local deformation of the material along the interfaces and grain boundaries. In the near-surface regions, the image is more smeared owing to the high ductility of the material.

In the central parts of the fractograms taken from the irradiated (Figs. 5a and 6a) and starting (Fig. 4) specimens, the surface relief is almost identical. As in the former case, the irradiated specimens, exhibiting high ductility, underwent plastic intercrystalline fracture with the formation of the typical relief, once irreversible deformation had exceeded 50%. An important feature of all of the irradiated specimens is the net-shaped relief due to brittle fracture of the $B2$ phase grains because of cleavage along the crystallographic planes (Figs. 5b and 6b). This net-shaped relief appears along the specimen perimeter immediately under the modified surface. In the specimens irradiated by the first dose, the width of the brittle fracture zone is nearly equal to the grain size of the $B2$ phase (about 40 μm). In the case of second-dose implantation, this zone is wider, covering three or four layers of grains. Thus, the

(a)

100 μm

(b)

50 μm

Fig. 5. Surface of fracture of the irradiated alloy: (a) central part of the section and (b) region under the side surface modified by the first dose.

(a)

200 μm

(b)

100 μm
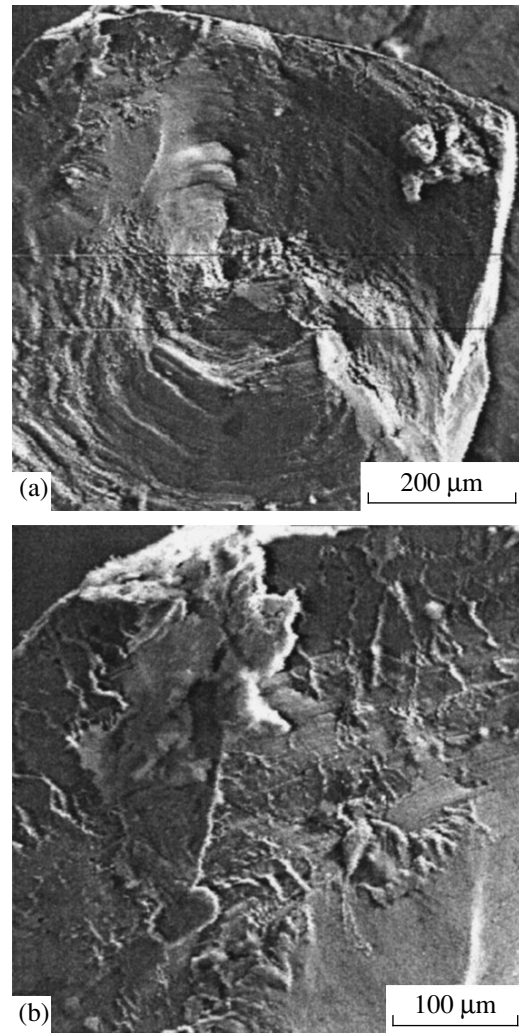
Fig. 6. Surface of fracture of the irradiated alloy: (a) central part of the section and (b) region under the side surface modified by the second dose.

extent and properties of the region formed under the modified microstructural layer cannot be treated on the microlevel. On the other hand, this region does not embrace yet the whole specimen and therefore can be referred to as a mesostructural deformation layer in ion-doped $Ni_{50}Ti_{40}Zr_{10}$ alloy (following the concepts put forward by Panin [8]). The mechanical properties of such a mesostructural layer (microhardness, ductility, and fracture surface morphology) are qualitatively different from those of the bulk material. The depth of this layer depends on the irradiation dose.

The SME parameters in the starting and irradiated specimens were similar. The SME temperature range is also the same. The amount of inelastic strain also remains unchanged and attains the maximum possible value for the given composition (about 10–11%) in all the specimens under study. However, the modified specimens fail due to temperature cycling at stresses

exceeding the applied load by 10–15%. The general strain accumulated at these loads is compensated for by the plastic component.

CONCLUSION

Ion modification significantly improves the general ductility of the material. At the early stages of deformation in the modified specimens, the surface layers are involved in the process of plastic deformation at lower loads. The ductility of the whole material increases because of the modified strained structure, which forms in these layers and prevents cracking. The existence of the mesostructural strained layer with specific ductility that lies immediately under the modified surface follows from the fractograms taken of the alloy. The structure and properties of such a layer, being one grain thick after the first dose and three or four grains thick after the second dose, vary. This suggests that a grain of

the initial $B2$ phase can take part as a whole in an elementary act of plastic deformation and be the main strained element of $Ni_{50}Ti_{40}Zr_{10}$ alloy on the mesoscopic structural level. The enhanced ductility of the initial $B2$ phase after $Ni_{50}Ti_{40}Zr_{10}$ surface modification improves the resistance to fracture due to high-load temperature cycling without affecting the SME parameters.

ACKNOWLEDGMENTS

REFERENCES

1. M. I. Gol'dshteĭn, V. S. Litvinov, and B. M. Bronfin, *Physical Metallurgy of High-Strength Alloys* (Metallurgiya, Moscow, 1986).

2. Yu. P. Sharkeev, N. V. Girsova, *et al.*, Fiz. Khim. Obrab. Mater., No. 4, 14 (1996).

3. E. F. Dudarev, *Microplastic Deformation and Yield Strength of Polycrystals* (Tomsk, 1988).

4. L. L. Meĭsner and V. P. Sivokha, Fiz. Met. Metalloved. **81** (5), 160 (1996).

5. L. L. Meisner and V. P. Sivokha, J. Phys. IV (Paris), Colloque C8 **5** (C8), 765 (1995).

6. R. P. Elliott, *Constitution of Binary Alloys, First Supplement* (McGraw-Hill, New York, 1967; Metallurgiya, Moscow, 1970).

7. E. W. Collings, *The Physical Metallurgy of Titanium Alloys* (American Society for Metals, Metals Park, Ohio, 1984; Metallurgiya, Moscow, 1988).

8. V. E. Panin, Izv. Vyssh. Uchebn. Zaved, Fiz. **41** (1), 7 (1998).

Translated by K.I. Kugel

=========== **OPTICS, QUANTUM ELECTRONICS** ===========

Energy Exchange between Four Copropagating Light Waves under Conditions of Diagonal Bipolar Nonlinearity

S. A. Podoshvedov

South Ural State University, Chelyabinsk, 454080 Russia
Received April 15, 1998

Abstract—Stationary energy exchange between four copropagating light beams in a medium with diagonal bipolar nonlinearity is studied. This type of nonlinearity is common in nematic liquid crystals. Solutions of underlying Hamilton's equations are thoroughly analyzed with the phase-plane approach for different wave-vector mismatches. The existence of an unstable eigenmode is demonstrated geometrically. This mode results from the considerable difference between the temperature dependences of the ordinary and extraordinary refractive indexes. Exact analytic solutions related to various regions of the phase plane are presented. Optical switching is demonstrated to be feasible under the stated conditions. It is found that the energy exchange essentially depends on the input phase mismatch. Intensities required for observing the predicted effects are evaluated. © 2000 MAIK "Nauka/Interperiodica".

INTRODUCTION

Optical switching, bistability, multistability, modulation instability, etc., have become burning issues. Produced by the interaction between intense laser radiation and matter, nonlinear optical phenomena are important in terms of both fundamental science and engineering. In all probability, they will provide practical means of controlling light with light, thus contributing to the progress in optical communications and computing [1, 2].

Quite a few studies, both theoretical and experimental, in the field of optical switching have been reported. Without intending to give a comprehensive overview of them, we cite a number of relevant interactions in various media where the effect occurs: wave interactions in tunnel-coupled waveguides [3, 4], the nonlinear propagation of differently polarized waves through a single-mode birefringent fiber or crystal [5], wave interactions at two or three different frequencies under second-order nonlinearity [2, 6, 7], and two-, three-, or four-wave mixing under third-order nonlinearity [8–14]. In media whose third-order susceptibility is governed by the nonlinear electron response to an electromagnetic wave, the last-named interaction offers a switching time (the transition time of nonlinear response) of an order of 10^{-13} s. On the other hand, the light-beam intensity necessary for optical switching to be observable is so large that it may cause optical damage of the sample. In conventional nonlinear optical media, the input power can be lowered by increasing the interaction length to centimeters or even hundreds of centimeters. This makes the practical implementation of this effect very expensive.

Recently, interest in nonlinear phenomena in liquid crystals has gained momentum [15–20]. It has been found that the mesophases may give dramatically high values of the third-order effective susceptibility $\chi^{(3)}$. Nematic liquid crystals are remarkable in this respect. Compared to isotropic liquids, the susceptibility of nematics may be larger by six to ten orders of magnitude. Such a strong third-order nonlinearity may stem from orientation [21] or photoconformation [22] effects. Thermal effects [23, 24] are also important, since the refractive indexes of nematics are very sensitive to temperature variations.

Due to very high third-order susceptibilities, many well-known nonlinear phenomena, such as self-focusing, stimulated light scattering, and four-wave mixing (FWM), can be obtained in spite of a very small interaction length, available in optical experiments with liquid crystals. As is known, it is strongly dependent on the orientation inhomogeneity of the samples and is significantly limited by severe light scattering in the mesophase; in practice, the interaction length is shorter than 100 μm. Nevertheless, under certain conditions, the nonlinear phenomena are feasible at a very low energy density of the incident radiation, such as that provided by milliwatt cw gas lasers. This fact both simplifies the design of the experimental setup and facilitates creating optical devices with extremely low switching powers.

This paper presents a detailed study of stationary energy exchange between four light beams under diagonal bipolar nonlinearity [25]. It was demonstrated [26] that such a nonlinearity may be associated with a thermal effect, namely, the dissipation of light energy in nematic liquid crystals. We will give both the phase

portraits and the exact analytic expressions for solutions to a system of FWM equations at different wave-vector mismatches.

BASIC EQUATIONS

Liquid crystals are known to have a remarkably strong dependence of the refractive indexes on temperature: $\partial n_\parallel / \partial T \approx -(4\text{–}6) \times 10^{-3}$ deg^{-1}, whereas $\partial n_\perp / \partial T$ is positive but more than 4 or 5 times smaller in magnitude. Therefore, light-dissipation nonlinearities become significant once the intrinsic absorption of the liquid crystal has attained a high level [21, 25, 26]. In a homogeneous nematic, a temperature variation causes local variations in the refractive indexes $n_\parallel$ and $n_\perp$, but the orientation of the director remains uniform.

The permittivity tensor ε_{ij}^0 of a liquid crystal is usually a uniaxial tensor of the form

$$\varepsilon_{ij}^0 = \varepsilon_\perp^0 + \varepsilon_a^0 n_i n_j, \tag{1}$$

where $\varepsilon_a = \varepsilon_\parallel - \varepsilon_\perp$ is the permittivity anisotropy, n_i are the Cartesian components of the unit vector aligned with the optical axis, $\varepsilon_\perp = n_o^2$, and $\varepsilon_\parallel = n_e^2$.

For the principal components of ε_{ij}^0, the local variations due to light absorption are $\delta\varepsilon_{\perp, \parallel} = \partial\varepsilon_{\perp, \parallel}/\partial T \delta T$, where δT is the variation of the sample temperature.

The temperature field evolves according to the diffusion equation

$$\frac{d\delta T}{dt} - \chi\Delta\delta T = \frac{\sigma c n |E|^2}{8\pi\rho C_p} - \Gamma\delta T, \tag{2}$$

where Δ is Laplacian, σ is the absorption coefficient (cm^{-1}), ρC_p is the heat capacity per unit volume (erg/cm^2 deg), $\chi \sim 10^{-3}$ cm^2/s is the thermal diffusivity, Γ^{-1} is the temperature transition time when external heat sources are applied to the sample, $\Gamma \sim \chi(a^{-2} + (\pi^2/L^2))$, a is the beam diameter; L is the cell length, and $|E|^2$ is the intensity of the interacting light waves. The
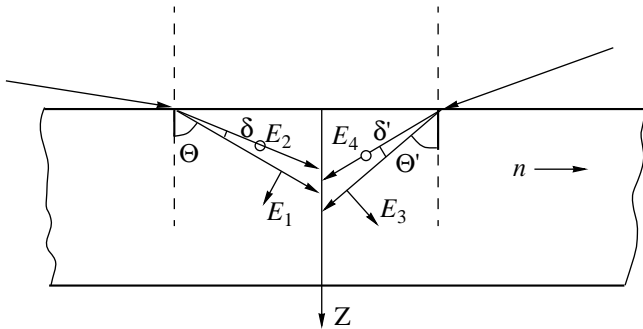
Fig. 1. FWM geometry. The angles δ and δ' are due to birefringence.

cell is assumed to easily dissipate heat through the walls, which is allowed for by the term $\Gamma\delta T$ if heating is uniform.

Figure 1 shows the energy exchange geometry. Two elliptically polarized waves are incident on a uniaxial nematic liquid crystal, where each of them splits into linearly polarized waves of the ordinary and the extraordinary type. We thus have four frequency-degenerate light waves propagating in the z direction. The optical axis of the crystal is oriented in the y direction. The extraordinary waves propagate at the angles Θ and Θ', whereas the directions of the ordinary waves are rotated through the angles δ and δ', respectively. The total wave amplitude inside the crystal is

$$\begin{aligned}\mathbf{E} &= \mathbf{e}_x(E_2(z)\exp(i\mathbf{k}_2\mathbf{r}) + E_4(z)\exp(i\mathbf{k}_4\mathbf{r})) \\ &+ \mathbf{e}_1 E_1(z)\exp(i\mathbf{k}_1\mathbf{r}) + \mathbf{e}_3 E_3(z)\exp(i\mathbf{k}_3\mathbf{r}),\end{aligned} \tag{3}$$

where $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ are the Cartesian unit vectors; $\mathbf{e}_1$ and $\mathbf{e}_3$ are the polarization unit vectors of the extraordinary waves; E_2 and E_4 are the ordinary-wave amplitudes; and E_1 and E_3 are the extraordinary-wave amplitudes.

Over the length $\lambda = (2\pi)/k$, the amplitudes E_1, E_2, E_3, and E_4 vary far more slowly than $\exp(i\mathbf{k}\mathbf{r})$.

Light dissipation results in the modulation of the refractive indexes. In what follows, we neglect the contribution from the z-components of the extraordinary waves, assuming Θ and Θ' to be small. In the context of this study, only the diagonal components of the permittivity tensor are variable: $\delta\varepsilon_{xy} = \delta\varepsilon_{yx} = 0$, $\delta\varepsilon_{xx} = \delta\varepsilon_\perp$, and $\delta\varepsilon_{yy} = \delta\varepsilon_\parallel$ [25]. We also assume that the tensor is much less sensitive to director rotation than to light dissipation. Substituting (3) into (2), we obtain an approximate solution as follows:

$$\delta T = \frac{c}{8\pi\rho C_p}\Gamma^{-1}Q(\mathbf{r}, t), \tag{4}$$

where

$$\begin{aligned}Q(\mathbf{r}, t) &= (\sigma_o n_o(|E_2|^2 + |E_4|^2) + \sigma_e n_e(|E_1|^2 + |E_3|^2)) \\ &\times (1 - \exp(-\Gamma t)) + \frac{\sigma_o n_o \Gamma}{\Gamma_o}E_2 E_4^* \\ &\times \exp(i(\mathbf{k}_2 - \mathbf{k}_4)\mathbf{r})(1 - \exp(-\Gamma_o t)) + \text{c.c.} \\ &+ \frac{\sigma_e n_e \Gamma}{\Gamma_e}E_1 E_3^*(\mathbf{e}_1\mathbf{e}_3)\exp(i(\mathbf{k}_1 - \mathbf{k}_3)\mathbf{r}) \\ &\times (1 - \exp(-\Gamma_e t)) + \text{c.c.}\end{aligned} \tag{5}$$

Here, $\Gamma_o = \chi(\mathbf{k}_2 - \mathbf{k}_4)^2$ and $\Gamma_e = \chi(\mathbf{k}_1 - \mathbf{k}_3)^2$. This approximate solution of (2) is valid under the assumption that Γ/Γ_o, $\Gamma/\Gamma_e \ll 1$ (the spatial period is much less than the length of the medium or the beam radius) and that the light amplitudes are constant.

Assume $|\mathbf{k}_2 - \mathbf{k}_4|$ and $|\mathbf{k}_1 - \mathbf{k}_3| \ll |\mathbf{k}_1|$, $|\mathbf{k}_2|$, $|\mathbf{k}_3|$, and $|\mathbf{k}_4|$. Then, time-independent FWM equations have the form

$$\frac{dE_1}{dz} = -\frac{i\pi C_e}{\lambda \cos^2\Theta}\Big((\Theta_0 + A_e b_e(\mathbf{e}_1\mathbf{e}_3)|E_3|^2)E_1 \cos\Theta$$
$$+ A_o b_o \cos\Theta' E_2 E_3 E_4^* \exp(i\Delta\mathbf{k}_z z)\Big), \tag{6}$$

$$\frac{dE_2}{dz} = \frac{i\pi C_o}{\lambda \cos(\Theta+\delta)}\Big((\Theta_0 + A_o b_o|E_4|^4)E_2$$
$$+ A_e b_e(\mathbf{e}_1\mathbf{e}_3)E_1 E_3^* E_4 \exp(-i\Delta\mathbf{k}_z z)\Big), \tag{7}$$

$$\frac{dE_3}{dz} = -\frac{i\pi C_e}{\lambda \cos^2\Theta'}\Big((\Theta_0 + A_e b_e(\mathbf{e}_1\mathbf{e}_3)|E_1|^2)E_3 \cos\Theta'$$
$$+ A_o b_o \cos\Theta E_1 E_2^* E_4 \exp(-i\Delta\mathbf{k}_z z)\Big), \tag{8}$$

$$\frac{dE_4}{dz} = \frac{i\pi C_o}{\lambda \cos(\Theta'+\delta')}\Big((\Theta_0 + A_o b_o|E_2|^2)E_4$$
$$+ A_e b_e(\mathbf{e}_1\mathbf{e}_3)E_1^* E_2 E_3 \exp(i\Delta\mathbf{k}_z z)\Big), \tag{9}$$

where $\Theta_0 = A_o(|E_2|^2 + |E_4|^2) + A_e(|E_1|^2 + |E_3|^2)$; $\Delta\mathbf{k}_z = (\mathbf{k}_2 + \mathbf{k}_3 - \mathbf{k}_1 - \mathbf{k}_4)_z$ is the projection of the wave-vector mismatch onto the propagation direction (z-axis); λ is the wavelength of the interacting waves;

$$|C_o| = \frac{\sigma c}{4\pi\rho C_p\Gamma}(dn_o/dT); \quad |C_e| = \frac{\sigma c}{4\pi\rho C_p\Gamma}(dn_e/dt);$$

$\sigma = (\sigma_o + \sigma_e)/2$ is the mean absorption coefficient; $A_o = \sigma_o n_o/\sigma$, $A_e = \sigma_e n_e/\sigma$, $b_o = \Gamma/\Gamma_o$, and $b_e = \Gamma/\Gamma_e$; and C_o and C_e are coupling constants. The coefficients A_o, A_e, b_o, and b_e are real. The difference between A_o and A_e stems from the absorption dichroism of the ordinary and extraordinary waves. The coefficients b_o and b_e determine the efficiency of recording stationary thermal gratings as against the case of uniform heating (b_o, $b_e < 1$). The bipolar property implies that $C_o > 0$ and $C_e < 0$; hence $C_o = |C_o|$ and $C_e = |C_e|$.

FWM in bipolar-response media was examined in [25], where the conditions for efficient energy exchange between the waves were analyzed in terms of recording of static thermal gratings under zero pump depletion. Our report presents a thorough theoretical study of the energy exchange in view of input conditions, specifically, the input phase mismatch and the distribution of the input total power over the light waves. For example, under certain conditions, a seem-

ingly insignificant variation in this distribution dramatically amplifies two of the four waves at the output, thus providing optical switching. Also, the energy exchange between the waves appears to be highly sensitive to the input phase mismatch.

PHASE-PLANE APPROACH

System (6)–(9) admits of the following integrals of motion:

$$a|E_1|^2 + |E_2|^2 + b|E_3|^2 + c|E_4|^2 = P, \tag{10}$$

$$a|E_1|^2 - |E_2|^2 = D_1, \tag{11}$$

$$b|E_3|^2 - c|E_4|^2 = D_2, \tag{12}$$

where

$$a = \frac{A_e b_e C_o n_e \cos^2\Theta(\mathbf{e}_1\mathbf{e}_3)}{A_o b_o C_e n_o \cos(\Theta+\delta)\cos\Theta'},$$

$$b = \frac{A_e b_e C_o n_e \cos^2\Theta'(\mathbf{e}_1\mathbf{e}_3)}{A_o b_o C_e n_o \cos(\Theta+\delta)\cos\Theta},$$

$$c = \cos(\Theta'+\delta')/\cos(\Theta+\delta).$$

Formula (10) can be viewed as the conservation of total power with allowance for the "dispersion" of the medium. Formulas (11) and (12) indicate that energy quantum exchange is possible only between differently polarized waves, such as the ordinary and extraordinary waves, occurring in the crystal. Let us change over to the variables q_i ($i = 1–4$) defined by $E_1 = q_1\sqrt{P/a}$, $E_2 = q_2\sqrt{P}$, $E_3 = q_3\sqrt{P/b}$, and $E_4 = q_4\sqrt{P/c}$. Then, integrals (10)–(12) take the form $|q_1|^2 + |q_2|^2 + |q_3|^2 + |q_4|^2 = 1$, $|q_1|^2 - |q_2|^2 = d_1$, and $|q_3|^2 - |q_4|^2 = d_2$, where $d_1 = D_1/P$ and $d_2 = D_2/P$. Using the new variables and the integrals of motion, we replace (6)–(9) by as few as two equations for the real-valued functions $\eta(s) = |q_2|^2$ and $\psi(s) = ks + \varphi_2 + \varphi_3 - \varphi_1 - \varphi_4$:

$$\frac{d\eta}{ds} = \sin\psi\sqrt{\eta(\eta + d_1)(1 - d_1 + d_2 - 2\eta)(1 - d_1 - d_2 - 2\eta)}, \tag{13}$$

$$\frac{d\psi}{ds} = \cos\psi$$
$$\times\left(\frac{\sqrt{(d_1 + \eta)(1 - d_1 + d_2 - 2\eta)(1 - d_1 - d_2 - 2\eta)}}{2\sqrt{\eta}}\right.$$
$$\left. - \frac{\sqrt{(d_1 + \eta)(1 - d_1 - d_2 - 2\eta)}}{\sqrt{1 - d_1 + d_2 - 2\eta}}\right.$$

$$+ \frac{\sqrt{\eta(1 - d_1 + d_2 - 2\eta)(1 - d_1 - d_2 - 2\eta)}}{2\sqrt{d_1 + \eta}} \quad (14)$$

$$- \frac{\sqrt{\eta(d_1 + \eta)(1 - d_1 + d_2 - 2\eta)}}{\sqrt{1 - d_1 - d_2 - 2\eta}} \Bigg)$$

$$- \eta(r_1 + r_2 + r_3 + r_4) + r_1 \frac{1 - d_1 - d_2}{2}$$

$$+ r_2 \frac{1 - d_1 + d_2}{2} - r_4 d_1 + k,$$

where $s = z\rho$, $k = \Delta\mathbf{k}_z/\rho$, and $\rho = (\pi C_e A_o b_o P/\lambda\cos\Theta) \times \sqrt{\cos(\Theta + \delta)\cos\Theta/\cos(\Theta' + \delta')\cos\Theta'}$. The coefficients r_i ($i = 1$–4) depend on nothing more than parameters of the medium and interaction geometry.

In the context of this study, $\cos\Theta = \cos\Theta'$, $\cos(\Theta + \delta) = \cos(\Theta' + \delta')$, and $(\mathbf{e}_1\mathbf{e}_3) \approx 1$ in (13) and (14). Therefore, we simply have $r_1 = r_3 = 1/r_2 = 1/r_4 = C_o n_e \cos\Theta/C_e n_o \cos(\Theta + \delta)$. Generally, r_i ($i = 1$–4) do not strictly meet these relationships, because the above conditions for $\cos\Theta$, $\cos\Theta'$, $\cos(\Theta + \delta)$, $\cos(\Theta' + \delta')$, and $(\mathbf{e}_1\mathbf{e}_3) \neq 1$ are not quite correct. However, these discrepancies are small and were neglected. As for $|q_1|^2$, $|q_3|^2$, and $|q|^2$, they can be expressed in terms of η through the integrals of motion d_1 and d_2: $|q_1|^2 = \eta + d_1$, $|q_3|^2 = (1 - d_1 + d_2 - 2\eta)/2$, and $|q_4|^2 = (1 - d_1 - d_2 - 2\eta)/2$.

Note that (13) and (14) are Hamilton's equations ($d\eta/ds = -\partial H/\partial\psi$ and $d\psi/ds = \partial H/\partial\eta$) for the canonical conjugates η and ψ, the Hamiltonian being

$$H = \sqrt{\eta(d_1 + \eta)(1 - d_1 + d_2 - 2\eta)(1 - d_1 - d_2 - 2\eta)}$$

$$\times \cos\psi - \frac{\eta^2}{2}(r_1 + r_2 + r_3 + r_4) \quad (15)$$

$$+ \eta\left(r_1\frac{1 - d_1 - d_2}{2} + r_2\frac{1 - d_1 + d_2}{2} - r_4 d_1 + k\right).$$

The value of H, defined by the input conditions, remains constant as the waves propagate in the medium. This enables us to obtain both analytic and phase-plane solutions of (13) and (14) at given d_1 and d_2. In the right-hand side of (15), the first term corresponds to the parametric term in (6)–(9), which is responsible for energy exchange, and the others (including η_2 and η), to the nonparametric terms, which represent self- and cross-phase modulation. In this study, the constants involved in the Hamiltonian of the two-dimensional oscillator are as follows: $d_1 = 0.1$, $d_2 = 0.1$ (in general, d_1 and d_2 range between –1 and 1), $A = r_1 + r_2 + r_3 + r_4 = 8.5$, and $B = r_1(1 - d_1 - d_2)/2 + r_2(1 - d_1 + d_2)/2 - r_4 d_1 = 1.7$. Drawing trajectories on the phase plane tangibly facilitates understanding exact analytic solutions of the set (13)–(14).

The shape of phase-plane trajectories is essentially dictated by the number of Hamiltonian extreme points ($\partial H/\partial\psi_{\eta_e, \psi_e} = \partial H/\partial\eta_{\eta_e, \psi_e} = 0$), if any, and their stability. They are known as the eigenmodes of (13)–(14) [6–12, 19, 20]. Figure 2 shows the dependence of the eigenmodes on the dimensionless wave mismatch k. Notice that there exist two eigenmodes, with $\eta \neq \eta_{max}$ and the input phase mismatch $\psi_e = 0$ or $\psi_e = \pi$. Their stability is governed by the Hamiltonian parameter A [see (15)]. If A is larger than a certain critical value A_{cr} ($A_{cr} = 4.33$ for the given d_1 and d_2), the diagram of the eigenmode with $\psi_e = \pi$ includes an instability segment (the dashed line in Fig. 2). If $A < A_{cr}$, the eigenmode with $\psi_e = \pi$ is stable for any k. If $A > A_{cr}$, then, at certain k, this eigenmode splits into three (two stable and one unstable) eigenmodes. In the context of this study, $A > A_{cr}$. Note that the presence of an unstable eigenmode under the stated FWM conditions stems from the significant difference between the "strengths" of the gratings recorded by the ordinary and extraordinary waves (C_e is much larger than C_o in magnitude). Other media with diagonal bipolar nonlinearity may be free from an unstable FWM eigenmode if $A < A_{cr}$ (the coupling coefficients in all the four equations for this process are nearly equal). As indicated by Fig. 2, the unstable mode with $\psi_e = \pi$ is eliminated if k is sufficiently large. However, the unstable case seems to be of immediate interest.

In what follows, we consider the cases $k = 0$ and $k = 0.4$. Figure 3 shows the phase portraits of the respective solutions to (13)–(14). Stable eigenmodes
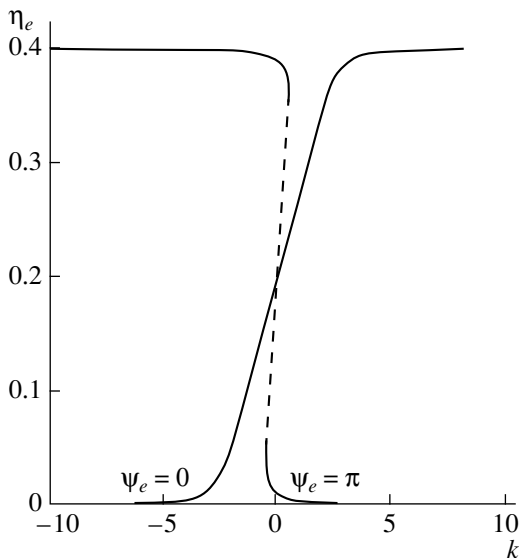
Fig. 2. Bifurcation diagram showing the effect of the wave mismatch on the FWM eigenmodes.
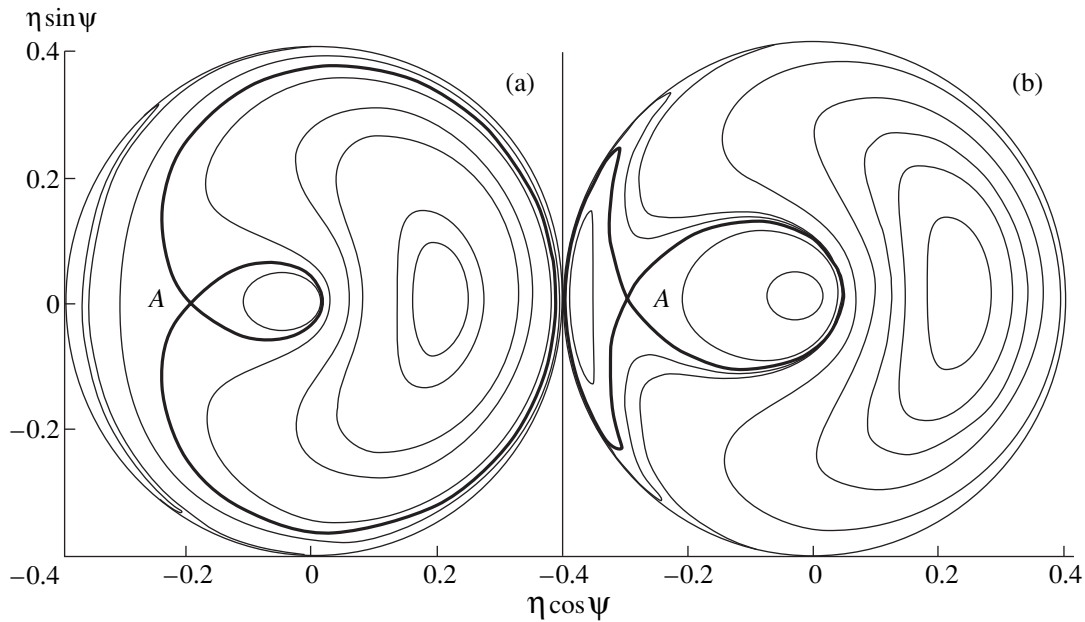
Fig. 3. Phase portraits at $k = $ (a) 0 and (b) 0.4. Point A is the unstable eigenmode.

are represented by stable center points, and unstable eigenmodes, by unstable saddle points, such as point A. The saddle point gives rise to a two-loop separatrix, whose constituent trajectories are drawn with heavy lines.

The convenience of the phase-plane approach to (13)–(14) is that it provides a visual tool to control the input power distribution and phase mismatch so that the energy exchange be maximum. Ideally, i.e., in the case of maximum energy exchange, there should be a phase trajectory going through the points $\eta = 0$ and $\eta = \eta_{max} = 0.4$. However, such a trajectory is nonexistent (see Fig. 3). On the other hand, one can easily find the phase-plane regions where the energy exchange is maximum for given d_1 and d_2. From the phase portraits for a two-dimensional oscillator with Hamiltonian (15), it follows that, if the input conditions situate the initial phase point close to stable eigenmodes, the energy exchange is insignificant. Conversely, if the initial point approaches the two-loop separatrix, the energy exchange may grow considerably.

Look at Fig. 3 to compare the shapes of the separatrix trajectories at the two values of the wave mismatch. Analytically, the two trajectories can be exactly described in terms of hyperbolic functions. Both of them asymptotically tend to point A, which is the unstable eigenmode for both k's. In Fig. 3a, the separatrix consists of a larger and a smaller loop. In Fig. 3b, the loops are nearly equal in size. In both cases, the larger loop avoids the region where $\cos\psi$ is positive. Switching from $k = 0$ to $k = 0.4$ restructures the entire phase portrait, as indicated by the shape of the two-loop separatrix.

The phase planes in Fig. 3 are divided into three regions with differing closed trajectories. Each closed trajectory represents a periodic solution, which can be expanded in elliptic functions. This property also features the trajectories (on both of the phase portraits) that leave the circle at the point with $\eta = 0.4$ and $\psi = 3\psi/2$ and return to the circle at the point with $\eta = 0.4$ and $\psi = \pi/2$. In the coordinates $\eta\cos\psi$ and $\eta\sin\psi$, we cannot display the trajectories in every detail. Note that these can be called simple separatrices, since they separate slightly different types of periodic solutions. Specifically, they serve as boundaries for the trajectories with ψ ranging from 0 to 2π and for those with ψ lying in a narrower interval.

In Fig. 3, a representative point moves toward larger magnitudes of η, when being in the upper half-plane ($\sin\psi > 0$), and toward smaller η's when $\sin\psi < 0$. For example, a point on the larger loop (Fig. 3a) moves clockwise, whereas that on the smaller loop counterclockwise.

ANALYTIC SOLUTIONS

Consider the analytic solutions corresponding to Fig. 3. It is useful to note that equations (13) and (14) describe the motion of a material point in a one-dimensional potential field. System (13)–(14) with Hamiltonian (15) yields the integral representation

$$\sqrt{\frac{A^2}{4} - 4s} = \int_{\eta_0}^{\eta} \frac{dx}{\sqrt{f(x)}}, \qquad (16)$$

where $\eta_0 = \eta(s = 0)$ and $f(x)$ is the potential field. Gen-
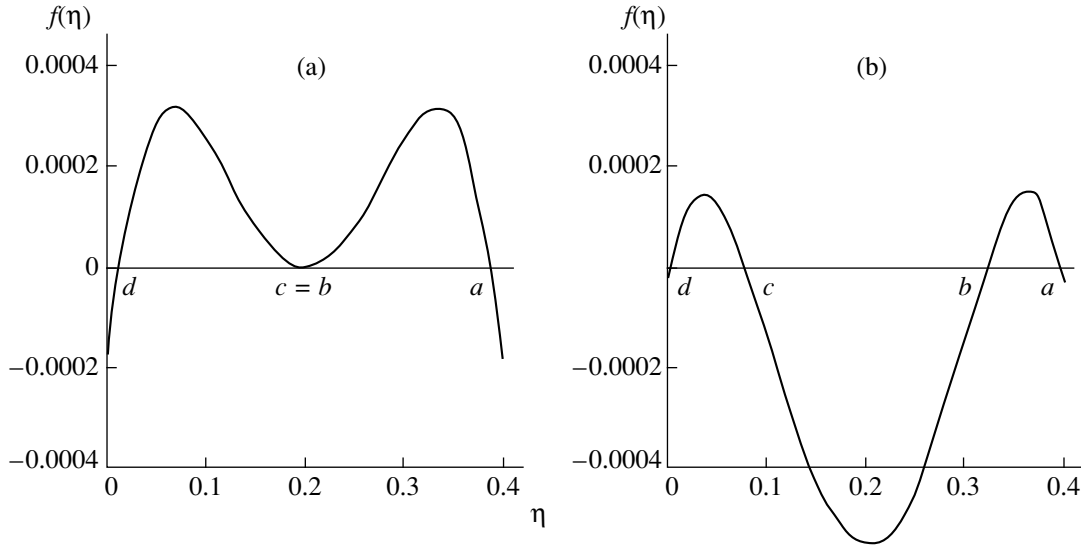
Fig. 4. One-dimensional potential $f(\eta)$ at $k = 0$ for (a) $H = H_{\mathrm{uns}}$ and (b) $H < H_{\mathrm{uns}}$.

erally, $f(x) = -x^4 + \alpha_1 x^3 + \alpha_2 x^2 + \alpha_1 x + \alpha_0$, where

$$\alpha_3 = \frac{4(2d_1 - 1) + A(B + k)}{\dfrac{A^2}{4} - 4}, \tag{17}$$

$$\alpha_2 = \frac{1 - 6d_1 + 5d_1^2 - d_2^2 - B^2 - k^2 - HA - 2Bk}{\dfrac{A^2}{4} - 4}, \tag{18}$$

$$\alpha_1 = \frac{d_1 - 2d_1^2 + d_1^3 - d_1 d_2^2 + 2H(B + k)}{\dfrac{A^2}{4} - 4}, \tag{19}$$

and

$$\alpha_0 = -\frac{H^2}{\dfrac{A^2}{4} - 4}. \tag{20}$$

Thus, solving (13)–(14) implies finding and analyzing the roots of $f(x) = 0$. Since this problem takes much effort in the general case, it is expedient to use a graphic approach. For the unstable eigenmode, the Hamiltonian takes on the value $H \approx 0.05$ at $k = 0$ or $H \approx 0.14952$ at $k = 0.4$. The corresponding plots of $f(x)$ are presented in Figs. 4a and 5a, respectively. Hereafter, the roots of $f(x) = 0$ are denoted by a, b, c, and d, with $a > b > c > d$. As follows from Fig. 4, $c = b$ and $f(x)$ can be expanded as $f(x) \approx (0.38855 - x)(0.2 - x)^2(x - 0.01145)$ at $k = 0$ and $f(x) \approx (0.3978 - x)(0.29975 - x)^2(x - 0.0445)$ at $k = 0.4$. It is seen that the material point is confined to one of two regions [where $f(x) \geq 0$] bounded by the points d and $c = b$ and the points $c = b$ and a. The motion within the first region is depicted by the smaller

loop in Fig. 3a and a trajectory to the right of point A in Fig. 3b. The motion within the second region is depicted by the larger loop in Fig. 3a and a trajectory to the left of point A in Fig. 3b. These solutions can be represented as follows (for $a + d \neq 2b$). For the trajectory to the left of point A in Fig. 3b,

$$\eta = b + \frac{2(b - d)(a - b)}{(a - d)\cosh U_1 + 2b - a - d}, \tag{21}$$

and for that to the right of point A in Fig. 3b,

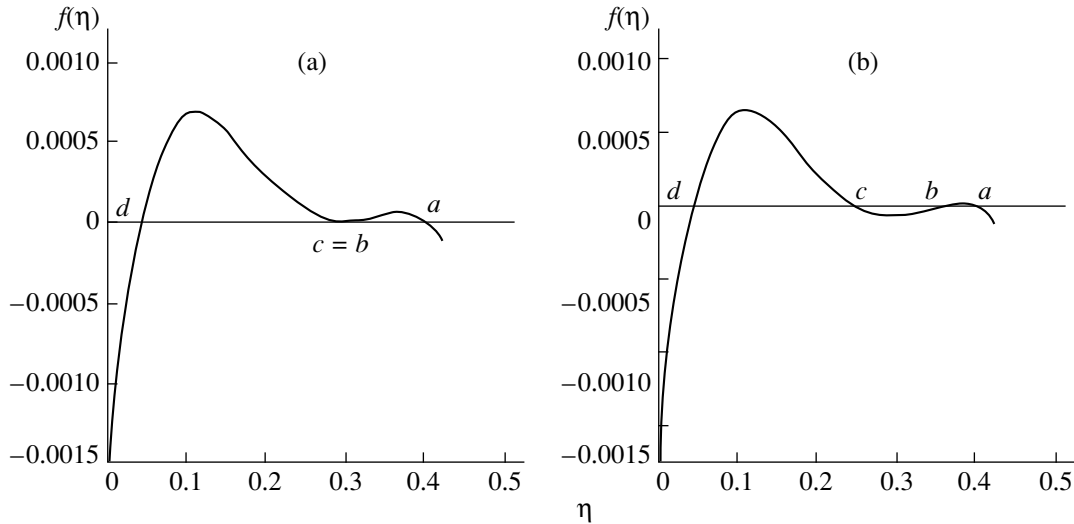$$\eta = b - \frac{2(a - b)(b - d)}{(a - d)\cosh U_2 + a + d - 2b}. \tag{22}$$

In (21) and (22),

$$U_1 = \pm\mathrm{arccosh}((\eta_0(a + d - 2b)$$
$$+ b(a + d) - 2ad)/((a - d)(\eta_0 - b)))$$
$$- \sqrt{(a - b)(b - d)}\sqrt{A^2/4 - 4} \times s,$$

and

$$U_2 = \pm\mathrm{arccosh}((\eta_0(a + d - 2b)$$
$$+ b(a + d) - 2ad)/((a - d)(b - \eta_0)))$$
$$- \sqrt{(a - b)(b - d)}\sqrt{A^2/4 - 4} \times s.$$

Note that the arch values are positive. For U_1, the plus and the minus signs correspond, respectively, to the cases where the trajectory starts from the upper half-plane ($\sin\psi > 0$) or the lower half-plane ($\sin\psi < 0$). For U_2, the reverse holds.

Fig. 5. One-dimensional oscillator potential for a material point involved in stationary energy exchange at $k = 0.4$ for (a) $H = H_{uns}$ and (b) $H < H_{uns}$.

If $a + d = 2b$, (21) and (22) can be recast into a very simple form. This is the case when $k = 0$ with the given d_1, d_2, A, and B (see the preceding section). We have the following numerical expressions for the phase trajectories forming the two-loop separatrix:

$$\eta = 0.2 \pm \frac{0.18855}{\cosh U}, \qquad (23)$$

$$\cos \psi = \frac{1.2591 - \cosh^2 U}{\sqrt{(\cosh^2 U - 0.8888)(\cosh^2 U - 0.395)}},$$

$$\sin \psi = \pm \frac{1.111 \sinh U}{\sqrt{(\cosh^2 U - 0.8888)(\cosh^2 U - 0.395)}}, \qquad (24)$$

where the plus and the minus correspond to the larger and the smaller loop, respectively, and $U = \pm \text{arccosh}(0.18855/(|0.2 - \eta_0|)) - 0.7070625s$. In the expression for U, the plus sign applies if the representative point initially lies on the smaller loop and in the lower half-plane or on the larger loop and in the upper half-plane, and the minus sign applies in the converse case.

Now, consider the case $H < H_{uns}$ (H_{uns} is the Hamiltonian for the unstable eigenmode). Figures 4b and 5b show the corresponding plots of $f(x)$ at $k = 0$ or $k = 0.4$, respectively. The former is drawn for $H = 0.02$, and the latter, for $H = 0.145$. Having evaluated the roots of $f(x) = 0$, we can factorize $f(x)$. For example, $f(x) \approx (0.39715 - x)(0.32175 - x)(0.0782499 - x)(x - 0.00285)$ at $k = 0$. The plots demonstrate that a material point η may move within one of two regions of the potential field $f(\eta)$. These regions are bound by the points d and c and the points a and b, respectively. At $k = 0$, the trajectories corresponding to the first region are inside the smaller loop in Fig. 3a, and those corresponding to the second region are

outside the larger loop. Consider the motions between d and c. By switching from x to $y = \arcsin \sqrt{((a-c)(x-d))/((c-d)(a-x))}$, the integral in (16) is brought into a tabulated form [27]. We thus obtain the solution

$$\eta = \frac{a(c-d)\text{sn}^2(U; r) + d(a-c)}{(c-d)\text{sn}^2(U; r) + a - c}, \qquad (25)$$

where $\text{sn}(U; r)$ is the elliptic sine, $r = \sqrt{((a-b)(c-d))/((a-c)(b-d))}$ is the modulus of the elliptic function, $U = \sqrt{(a-c)(b-d)} \sqrt{A^2/4 - 4} \, s/2 \pm F(\beta_0; r)$, $F(\beta_0; r)$ is the elliptic integral of the first kind, and $\beta_0 = \arcsin \sqrt{((a-c)(\eta_0 - d))/((c-d)(a-\eta_0))}$.

For $H = 0.02$ ($f(x)$ in Fig. 4b),

$$\eta \approx \frac{0.0299 \, \text{sn}^2(U; r) + 0.0091}{0.0754 \, \text{sn}^2(U; r) + 0.3189}, \qquad (26)$$

where $U = 0.5979s + F(\beta_0; r)$ if the initial point is in the upper half-plane and $U = 0.5979s - F(\beta_0; r)$ if the point is in the lower half-plane.

Having computed η, we can find the remaining variables: $|q_1|^2, |q_3|^2, |q_4|^2$, and ψ.

Consider the motions between a and b. (The region where the material point η can move is determined by the input conditions, the value of the Hamiltonian being conserved during the motion.) By switching from x to $y = \arcsin \sqrt{((b-d)(a-x))/((a-b)(x-d))}$, the
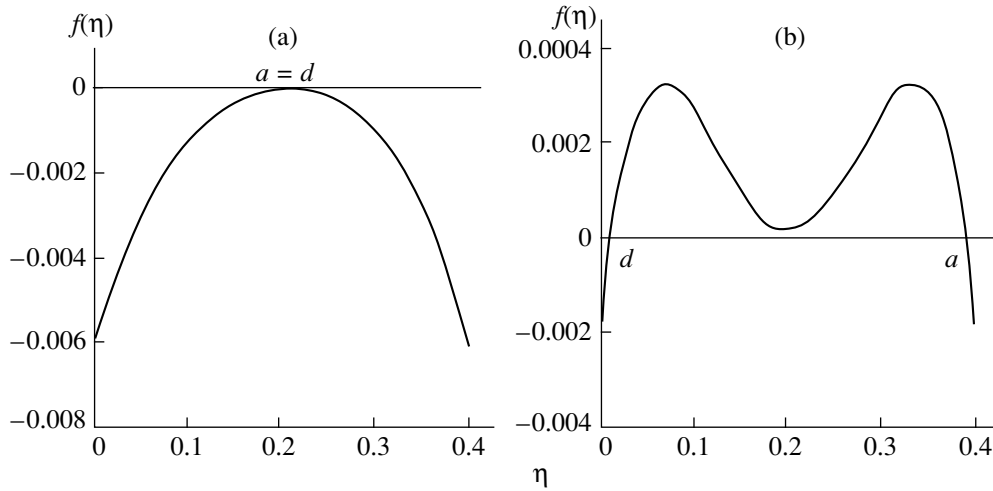
Fig. 6. One-dimensional potential $f(\eta)$ at $k = 0$ for (a) $H = H_{st}$ and (b) $H_{uns} < H < H_{st}$. The phase-plane trajectories for Fig. 6b are inside the larger loop (Fig. 3a).
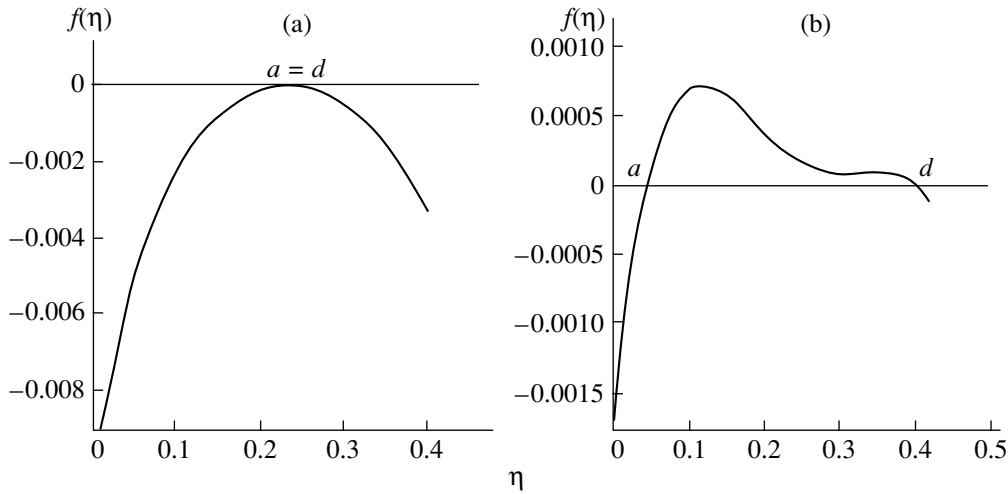
Fig. 7. One-dimensional potential $f(\eta)$ at $k = 0.4$ for (a) $H = H_{st}$ and (b) $H_{uns} < H < H_{st}$. The phase-plane trajectories for Fig. 7b encircle the separatrix (Fig. 3b).

integral in (16) is brought into a tabulated form. We thus obtain the solution

$$\eta = \frac{d(a-b)\mathrm{sn}^2(U; r) + a(b-d)}{(a-b)\mathrm{sn}^2(U; r) + b - d}, \qquad (27)$$

where $U = \pm F(\mu_0; r) - \sqrt{(a-c)(b-d)} \sqrt{A^2/4 - A} \times s/2$, $r = \sqrt{((a-b)(c-d))/((a-c)(b-d))}$ is the modulus of the Jacobian elliptic function, and $\mu_0 = \arcsin\sqrt{((b-d)(b-\eta_0))/((a-b)(\eta_0-d))}$. The sign in the expression for U is dictated by the phase-plane region from which the material point starts (Fig. 3).

Solutions (25) and (27) go over into solutions (22) and (21), respectively, as $H \longrightarrow H_{uns-0}$. Such behavior could be predicted from Fig. 3.

Quite a different situation is observed in the case $H_{uns} < H < H_{st}$, where H_{st} is the value of the Hamiltonian at the stable eigenmode with $\psi_e = 0$. The plots of $f(x)$ for $H = H_{st}$ and $H_{uns} < H < H_{st}$ at $k = 0$ and $k = 0.4$ are shown in Figs. 6 and 7. Notice that only two roots of $f(x)$ are real. If $H = H_{st}$, the solution is the simplest: $\eta(s) = \eta_e$. At $k = 0$, the corresponding factorization is $f(x) \approx (0.2 - x)^2(-x^2 + 0.4x - 0.149511)$. However, as follows from Figs. 6 and 7, of special interest is the case $H - H_{uns} \approx 0$, when the energy exchange is maximum. Figures 6b and 7b are obtained, respectively, for $H = 0.051$, $k = 0$ and $H = 0.155$, $k = 0.4$. At $k = 0$, the H value corresponds to a trajectory inside the larger loop. At $k = 0.4$, we have a trajectory encircling the two-loop separatrix (Fig. 3b). After the roots of $f(x)$ are evaluated, we can factorize $f(x)$. For example, $f(x) \approx (x - 0.01175)(0.38825)(x^2 - $

0.4x + 0.040544) at $k = 0$ and $H = 0.051$. By switching from x to $y = 2\operatorname{arccot}\sqrt{(q(a-x))/(p(x-b))}$, the inte- gral in (16) is brought into a tabulated form. We thus arrive at the solutions

$$\eta = \frac{(p+q)(qa+pd)\operatorname{sn}^2(U;r) + 2pq(a+d)\operatorname{cn}^2(U;r) \pm 2pq(a-d)\operatorname{cn}(U;r)}{(p+q)^2\operatorname{sn}^2(U;r) + 4pq\operatorname{cn}^2(U;r)}, \tag{28}$$

where $U = \sqrt{pq}\,\sqrt{A^2/4 - 4s} \pm F(\alpha_0; r)$, $p^2 = (b'-a)^2 + b''^2$; $q^2 = (b'-d)^2 + b''$; $b = b' + ib''$ (complex root), $\alpha_0 = 2\operatorname{arccot}\sqrt{(q(d-\eta_0))/(p(a-\eta_0))}$, and $r = 0.5\sqrt{((p+q)^2 + (a-d)^2)/pq}$.

Formula (28) gives two analytic solutions of system (13)–(14). The sign in the expression for U before $F(\alpha_0; r)$ is chosen according to the location of the initial point on the phase plane. To be specific, take the solution defined by (28) with the plus sign preceding $\operatorname{cn}(U; r)$. If the initial point is in the upper half-plane, then $U = \sqrt{pq}\sqrt{A^2/4 - 4s} - F(\alpha_0; r)$. Otherwise, $U = \sqrt{pq}\sqrt{A^2/4 - 4} + F(\alpha_0; r)$. The converse rule applies to the solution with the minus sign before $\operatorname{cn}(U; r)$. If $H \longrightarrow H_{\text{uns}+0}$, then solutions (28) go over into (21) and (22); i.e., we have the phase trajectories constituting the two-loop separatrix.

In the next section, the phase-plane and the analytic approach will be used to consider optical switching. We believe that the predicted effects will find practical application.

OPTICAL SWITCHING

Consider the FWM at $k = 0$. Figure 8 gives the power of the ordinary wave E_2 plotted against the distance z traveled by the light waves inside the crystal. The values of E_2 and z are normalized to the total input power P_{in} and to the length L of the medium, respectively. Here, $P_{\text{in}} = |E_{10}|^2 + |E_{20}|^2 + |E_{30}|^2 + |E_{40}|^2$, where E_{10}, E_{20}, E_{30}, and E_{40} are the input strengths of the waves. Unlike P in (10), the total power P_{in} varies with z. Curve 1 is obtained for the input phase mismatch $\psi_0 = \psi(z = 0) = 0$, whereas curve 2 corresponds to $\psi_0 = \pi$, other parameters being equal; both of the curves are computed at $\rho L = 50$. The input total power is distributed as follows: $|E_{10}|^2/P_{\text{in}} = 0.1728$, $|E_{20}|^2/P_{\text{in}} = 2.023 \times 10^{-3}$, $|E_{30}|^2/P_{\text{in}} = 0.757$, and $|E_{40}|^2/P_{\text{in}} = 0.06815$. Notice that $|E_2|^2/P_{\text{in}}$ has a sharp peak if $\psi_0 = 0$, in contrast to the case $\psi_0 = \pi$. The gain in E_2 is as high as 33.4 times when the length of the medium and the total input power are selected appropriately. Under the same conditions, the extraordinary wave E_1 is also amplified, by a factor of about 4.4, whereas the waves E_3 and E_4, which originate from the second incident wave, lose their energy. Such a large difference between the gains

for E_2 and E_1 primarily stems from the difference in dispersion properties of ordinary and extraordinary waves, which affects C_o and C_e. The high sensitivity of the energy exchange to the input phase mismatch can be understood with the help of the phase-plane and the analytic approach, using Fig. 3a or exact solutions (25) and (27), respectively. The matter is that curves 1 and 2 in Fig. 8 correspond to solutions of (13)–(14) from qualitatively differing regions in the phase plane. The considerable dependence of the energy exchange on the input phase mismatch can physically be explained by the fact that nonlinearity (4)–(5) behaves so that each grating is read out by the waves of both polarizations. The process where a grating recorded by extraordinary waves is read out by an ordinary wave gives the parametric term in (6)–(9) that represents energy exchange. The process where a grating recorded by ordinary waves is read out by an ordinary wave gives the term representing cross-phase modulation. The wave E_2 is the sum of two waves: the transmitted wave and the wave E_4, which is reflected from the grating recorded by the extraordinary waves E_1 and E_3. If the two waves are added in phase ($\psi_0 = 0$), the wave E_2 is materially amplified if z/L is not very large (see curve 1 in Fig. 8). If they are added in antiphase, the energy exchange is virtually zero (see curve 2 in Fig. 8).

We now discuss the FWM conditions when optical switching is pronounced. Figure 9 shows another power diagram for the ordinary wave E_2, the coordinates being the same as in Fig. 8. The input phase mismatch is zero for each curve ($\psi_0 = 0$). The input total
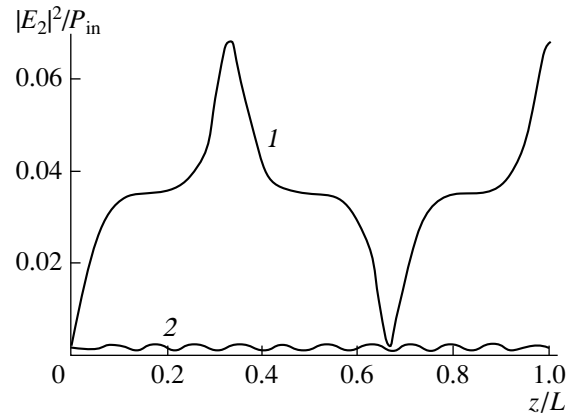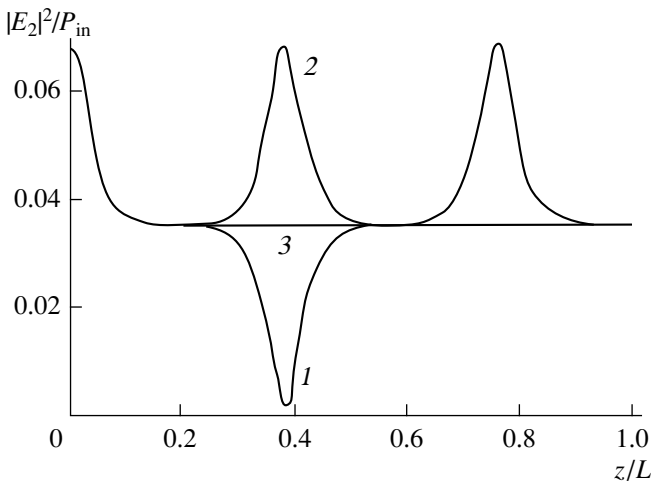

Fig. 8. Normalized power of the wave E_2 vs. normalized distance.

$|E_2|^2/P_{\text{in}}$

Fig. 9. Normalized power of the wave E_2 vs. normalized distance as applied to optical switching.

power is distributed as follows: $|E_{10}|^2/P_{\text{in}} = 0.7579$, $|E_{20}|^2/P_{\text{in}} = 0.0682$, $|E_{30}|^2/P_{\text{in}} = 0.174$, and $|E_{40}|^2/P_{\text{in}} = 0.002$. Curve *1* corresponds to the case where the input conditions place the initial phase point inside the larger loop, whereas curve *2* is plotted for the initial point outside the separatrix (Fig. 3a). The difference in $\eta_0 = |E_2(z=0)|^2/P_{\text{in}}$ between curves *1* and *2* is no more than 10^{-6} (Fig. 9). The period of curve *1* is seen to be nearly twice as large as that of curve *2*. Given the crystal length, the input total power can be adjusted so that the switch is in the ON or the OFF state at the output, depending on the input conditions. The ON state is characterized by $(|E_2(z=L)|^2 = |E_2(z=0)|^2)$, and the OFF state, by $(|E_2(z=L)|^2 \approx 0)$. The ON to OFF power ratio is 33.4 for the ordinary waves and 4.3 for extraordinary ones. Curve *3* is plotted for the input conditions that place the initial point just on the two-phase separatrix, with $|E_2(z)|^2/P_{\text{in}}$ asymptotically approaching the unstable-mode level.

Finally, we shall give some computed data. Let $a = 0.1$ cm, $L = 100$ μm (length of the medium), $\Lambda = 0.5$ μm, $\rho C_p \approx 1.5 \times 10^7$ erg/cm^3, $n_e = 1.71$, $n_o = 1.51$, $A_o \sim A_e = 1$, and $b_o \sim b_e = 0.5$. Also, take a moderate value of absorption: $\sigma = 5$ cm^{-1}. Then we have $\Gamma = 10^3$ s^{-1}. Two-wave amplification is observable at an input intensity of $\sim 7.7 \times 10^4$ W/cm^2, and liquid-crystal FWM switching is possible with intensities larger than $\sim 8.8 \times 10^4$ W/cm^2. On the other hand, the switching time (the transition time of static gratings) in this case is $\sim 10^{-3}$ s. Although this value cannot compete against that for bulk crystals, it is worth bearing in mind that the minimum switching intensity required by liquid crystals is lower by seven orders of magnitude.

CONCLUSION

The stationary energy exchange between four light beams propagating under diagonal bipolar nonlinearity was studied by means of Hamilton's equations. The phase-plane and the exact analytic representations of their solutions were obtained. An unstable eigenmode with the phase mismatch $\psi_e = \pi$ was disclosed. This mode results from the considerable difference between the coefficients in the equations describing the variation of the ordinary and extraordinary amplitudes along the propagation direction. The effect of input conditions on FWM dynamics was analyzed in various cases. It was found that the energy exchange can be both high ($\psi_0 = 0$) and near-zero ($\psi_0 = \pi$). It was demonstrated that optical switching can be performed by means of liquid crystals with thermal nonlinearity. The power required to observe the predicted effects was evaluated.

To the best of our knowledge, no experimental evidence for the effects has been reported. We hope that the theory and recommendations presented here will be of practical value. For example, they may find application in optical switches using nematic liquid crystals.

REFERENCES

1. A. A. Maĭer, Usp. Fiz. Nauk **165**, 1037 (1995).
2. A. A. Maĭer, Usp. Fiz. Nauk **166**, 1171 (1996).
3. A. A. Maĭer, Kvantovaya Élektron. (Moscow) **9**, 2296 (1982).
4. A. A. Maĭer, Kvantovaya Élektron. (Moscow) **11**, 157 (1984).
5. H. G. Winful, Appl. Phys. Lett. **47**, 213 (1985).
6. G. Cappellini, S. Trillo, S. Wabnitz, *et al.*, Opt. Lett. **16**, 637 (1991).
7. S. Trillo and S. Wabnitz, Opt. Lett. **17**, 1572 (1992).
8. S. A. Podoshvedov and F. V. Podgornov, Opt. Spektrosk. **81**, 450 (1996).
9. S. A. Podoshvedov, Opt. Commun. **142**, 79 (1997).
10. S. A. Podoshvedov, Opt. Spektrosk. **83**, 955 (1997).
11. S. A. Podoshvedov, Pis'ma Zh. Tekh. Fiz. **23** (6), 61 (1997).
12. S. A. Podoshvedov, Pis'ma Zh. Tekh. Fiz. **23** (7), 91 (1997).
13. L. A. Bol'shov, D. V. Vlasov, and R. A. Garaev, Kvantovaya Élektron. (Moscow) **9**, 83 (1982).
14. V. D. Solov'ev and A. I. Khizhnyak, Opt. Spektrosk. **53**, 723 (1982).
15. Th. Peschel, T. Dannberg, U. Langbein, *et al.*, J. Opt. Soc. Am. B: Opt. Phys. **5**, 29 (1988).
16. I. C. Khoo and J. Y. Hou, J. Opt. Soc. Am. B: Opt. Phys. **2**, 761 (1985).
17. I. C. Khoo and P. Zhou, J. Opt. Soc. Am. B: Opt. Phys. **6**, 884 (1989).
18. I. C. Khoo, R. R. Michael, R. J. Mansfield, *et al.*, J. Opt. Soc. Am. B: Opt. Phys. **8**, 1464 (1991).

19. S. A. Podoshvedov, in *Topical Meeting on Diffractive Optics'97*, Digest Series, **12** (1997), p. 126.

20. S. A. Podoshvedov, in *Topical International Meeting on Optics of Liquid Crystals'97* (1997), p. 88.

21. B. Ya. Zel'dovich and N. V. Tabiryan, Usp. Fiz. Nauk **147,** 633 (1985).

22. S. G. Odulov, Yu. A. Reznikov, and A. I. Khizhnyak, Zh. Éksp. Teor. Fiz. **85,** 1475 (1982).

23. V. Volterra and E. Wiener-Avhear, Opt. Commun. **12,** 194 (1974).

24. O. V. Garbyan, V. I. Zhdanov, N. I. Zheludev, *et al.*, Kristallografiya **26,** 789 (1981).

25. T. B. Galstyan, B. Ya. Zel'dovich, Yam Chun Ku, *et al.*, Zh. Éksp. Teor. Fiz. **100,** 737 (1991).

26. N. V. Tabiryan and B. Ya. Zel'dovich, in *Nonlinear Optical Materials, Proceedings of SPIE*, Ed. by G. Roosen (SPIE, Bellingham, Wash., 1989), Vol. 1017, pp. 193–198.

27. E. A. Janke, F. Emde, and F. Lösch, *Tafeln Höherer Funktionen* (Tables of Higher Functions), 6te Auflage, neubearbeitet von F. Lösch (Teubner, Stuttgart, 1960; Nauka, Moscow, 1968).

Translated by A.A. Sharshakov

============================= **OPTICS, QUANTUM ELECTRONICS** =============================

Optical Aberrations in a Bent-Crystal Spectrometer

V. A. Bryzgunov

Kurchatov Institute of Atomic Energy, All-Russia Scientific
Research Center, Moscow, 123182 Russia

Received February 17, 1999

Abstract—Aberrations in a bent-crystal X-ray spectrometer were studied by the methods of analytical geometry. Trigonometric equations describing aberrations in a real spectrometer were derived and solved by approximate techniques. The cases of crystal analyzers possessing spherical, toroidal, and cylindrical surfaces were considered. The investigation was aimed at evaluation of the resolving power of a spectrometer with an extended radiation source (tokamak plasma). © *2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

X-ray spectrometers used for hot plasma monitoring are typically based on the Johann scheme [1]. Measurements of the Doppler broadening and shift of the spectral line emitted by a small admixture of heavy ions allow the ion temperature and rotation rate in the hydrogen plasma of tokamaks to be determined [2, 3]. Tokamaks featuring a deuterium–tritium thermonuclear reaction exhibit a manifold increase in intensity of the neutron and gamma emissions. In connection with this, radical measures have to be taken for reducing the background emission level for the X-ray spectroscopic measurements to be performed on large thermonuclear setups such as the International Thermonuclear Experimental Reactor (ITER) currently under construction. Using spherical or toroidal bent crystal analyzers instead of the conventional cylindrical crystals markedly reduces divergence of the X-ray beam at the spectrometer input. This, in turn, allows a decrease in cross-section of the input channel in the reactor protection system by which the spurious radiation may enter (together with the X-ray beam) the spectrometer. Placing a selective polycrystalline graphite mirror [4] in the path of the weakly diverging beam provides additional increase in the signal-to-noise ratio [5, 6].

The X-ray spectrometers for the laser plasma investigations usually employ spherical bent crystal analyzers collecting radiation from a small emitting volume, which are capable of measuring the spectrum using a few short-time (100 fs) emitted pulses [7].

This analyzer system based on a crystal bent in two dimensions provides extremely favorable conditions for carrying out experiments on a modern level requiring both a large relative aperture and a high spectral resolution. Using modern technologies for manufacturing a crystal analyzer, it is possible to retain perfect structure of the initial material [8]. However, high-aperture spectrometers may still feature considerable geometric distortions of the spectral lines.

Development of the Johann spectrometer scheme, as well as other focusing systems suggested many years ago [1], always involved the need in increasing resolving power of the instruments and, hence, was accompanied by solving the task of reducing aberrations in the focusing spectrometers [9]. Aberrations in the classical spectrometer with cylindrical crystal analyzer have been thoroughly analyzed (see, e.g., [10]). At the same time, the problem of aberrations in a crystal analyzer with two-dimensional (2D) bending is still insufficiently studied [11, 12].

APERTURE ABERRATIONS

The Johann scheme does not provide for the ideal X-ray beam focusing. The appearance of aperture aberrations in a spectrometer based on a spherically bent crystal analyzer with the curvature radius R is illustrated in Fig. 1. The crystal analyzer *1* expands the spectrum over a cylindrical surface on which a one-dimensional radiation detector *2* is placed. The base of this cylinder is the focal circle *3* (the Rowland circle) containing the pole P of the crystal. Consider an orthogonal coordinate system xyz with the origin at a point 0 on the circle ($0P = R$). The X-ray $\mathbf{v}$ reflected from an arbitrary point C of the crystal analyzer is registered by the detector at a point D. The orthogonal coordinates of point C are determined by the radius vector R and the polar angles ε and ω. The position of point D with respect to height ($Z_D = DE$) is determined by the reflection angle Φ.

We must bear in mind that the beam is reflected at the Bragg angle $\vartheta = \arcsin\lambda/2d$ (λ is the radiation wavelength, and d is the interplanar distance in the crystal). A monochromatic beam PE lying in the coor-

dinate plane *xy* (the median plane of spectrometer) determines point E on the focal circle, which corresponds to the radiation with the wavelength λ. The reflection angle for this beam is $\Phi = \Phi_\lambda = \pi/2 - (\vartheta + \alpha)$. Aberrations in this spectrometer are created by beams for which $\Delta\Phi = \Phi - \Phi_\lambda \neq 0$. The beam shift by an angle $\Delta\Phi$ corresponds to an arc of the focal circle $\Delta s = R\Delta\Phi$. In the given spectrometer scheme, this shift is equivalent to a change in the radiation wavelength by $\Delta\lambda = -\lambda\Delta\Phi/\tan\vartheta$ and, hence, in the Bragg angle by $\Delta\vartheta = -\Delta\Phi$.

The reflected beam **v** forms an angle $\pi/2 - \vartheta$ with respect to the normal **N** to the crystal surface. Therefore, we obtain the scalar vector product

$$\mathbf{v} \cdot \mathbf{N} = |\mathbf{v}||\mathbf{N}|\sin\vartheta. \qquad (1)$$

At the same time, the normal **N** to the spherical crystal surface is collinear with (or antiparallel to) the radius vector **R**, so that $\mathbf{N} = -\mathbf{R}$. If the reflecting surface makes an angle α with the crystal surface (oblique reflection), the normal to the reflecting plane does not coincide with **R**. In this case, an expression for the unit normal vector is

$$n = (\sin\alpha\sin\omega - \cos\alpha\sin\varepsilon\cos\omega)\mathbf{i}$$
$$- (\sin\alpha\cos\omega + \cos\alpha\sin\varepsilon\sin\omega)\mathbf{j} + \cos\alpha\cos\varepsilon\mathbf{k}, \qquad (2)$$

where **i**, **j**, and **k** are the unit vectors of the coordinate axes *x*, *y*, and *z*, respectively. For the spherically bent crystal, equation (1) yields

$$\sin^2\vartheta(\sin^2\Phi - 2\sin\varepsilon\sin\Phi\sin(\Phi + \omega)$$
$$- 2\cos\varepsilon Z_D/R + Z_D^2/R^2 + 1)$$
$$= [\cos\alpha(1 - \cos\varepsilon Z_D/R - \sin\varepsilon\sin\Phi\sin(\Phi + \omega))$$
$$- \sin\alpha\sin\Phi\cos(\Phi + \omega)]^2. \qquad (3)$$

Upon solving equation (3) by numerical methods, we may readily find the values of Φ for all reflected beams. Then, by integrating over the entire crystal surface, we will determine the intensity of the reflected monochromatic radiation as a function of $\Delta\vartheta = \Phi_\lambda - \Phi$, that is, the instrumental function. It should be noted that real crystals exhibit a mosaic character, which results in that the reflection takes place within a certain angular interval near the Bragg angle. This factor accounts for additional broadening of the spectral line. The total instrumental function is obtained as a mathematical convolution of the two functions.

In a real spectrometer, the relative size of the crystal analyzer with respect to the Rowland circle radius is usually small. Taking this fact into account and neglecting high orders of small parameters in the expansions,
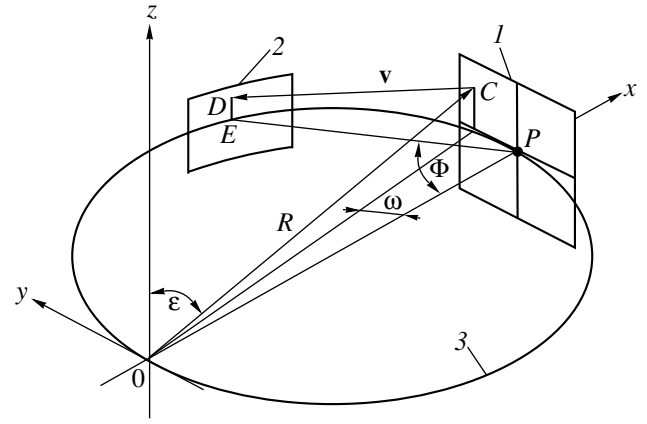
Fig. 1. Schematic diagram of the X-ray beam trajectory in a spectrometer with bent crystal analyzer: (*1*) crystal analyzer; (*2*) detector; (*3*) Rowland circle.

we obtain an approximate solution to equation (3) in the following form:

$$\Delta\Phi \approx \frac{Y_C^2}{2R^2\tan(\vartheta + \alpha)} - \frac{Z_D^2\tan\vartheta}{2R^2\sin^2(\vartheta + \alpha)}$$
$$- \frac{(2Z_C Z_D - Z_C^2\cos^2(\vartheta + \alpha))\sin\alpha\cos(\vartheta + \alpha)}{2R^2\sin^2(\vartheta + \alpha)\cos\vartheta}. \qquad (4)$$

Upon introducing small changes into the scheme of Fig. 1, one may readily calculate aberrations for a spectrometer with cylindrical bent crystal analyzer. Here, the equation describing the aberrations according to (1) and the approximate solution are as follows:

$$\sin^2\vartheta[\sin^2\Phi - 2\sin\Phi\sin(\Phi + \omega)$$
$$+ (Z_D/R - Z_C/R)^2 + 1]$$
$$= [\cos\alpha - \sin\Phi\sin(\Phi + \omega + \alpha)]^2, \qquad (5)$$

$$\Delta\Phi \approx \frac{Y_C^2}{2R^2\tan(\vartheta + \alpha)} - \frac{(Z_D - Z_C)^2\tan\vartheta}{2R^2\sin^2(\vartheta + \alpha)}. \qquad (6)$$

Note that, for $\alpha = 0$, equation (6) coincides with the expression derived in [10].

Similar considerations are used to calculate aberrations for a spectrometer with toroidal bent crystal analyzer. Here, the reflecting surface is formed by a circle of radius R_s rotated about the *z*-axis, with the center of this circle occurring at a distance of $R - R_s$ from the *z*-axis. The meridional radius R determines (as in Fig. 1) the diameter of the focal circle. In the particular case of $\alpha = 0$, equation (1) for a toroidal crystal analyzer acquires the following form:

$$\sin^2\vartheta[\sin^2\Phi - 2\sin\Phi\sin(\Phi + \omega) - 2\{(1 - \sin\varepsilon)$$

$$\times (1 - \sin\Phi\sin(\Phi + \omega)) + \cos\varepsilon Z_D/R\}R_s/R$$

$$+ 2(1 - \sin\varepsilon)R_s^2/R^2 + Z_D^2/R^2 + 1]$$

$$= [\sin\varepsilon(1 - \sin\Phi\sin(\Phi + \omega))$$

$$- \cos\varepsilon Z_D/R + (1 - \sin\varepsilon)R_s/R]^2. \tag{7}$$

The corresponding beam shift in a spectrometer with toroidal crystal is

$$\Delta\Phi \approx \frac{Y_C^2}{2R^2\tan\vartheta} - \frac{Z_D^2}{R^2\sin 2\vartheta} + (1/R - 1/R_s)$$

$$\times \left[\frac{Z_C^2\tan\vartheta}{2R_s} + \frac{2Z_C Z_D - Z_C^2}{R\sin 2\vartheta}\right]. \tag{8}$$

CALCULATION RESULTS AND DISCUSSION

Using modern computational facilities, it is not difficult to solve the exact trigonometric equations (3), (5), and (7) by numerical methods. However, fast estimation and qualitative analysis of the aperture aberrations are conveniently performed using approximate solutions in the form (4), (6), and (8), respectively.

Figure 2 shows a schematic diagram of aberrations in a spectrometer employed for measuring the ion temperature in a thermonuclear reactor using the Doppler shift of a resonance line from the helium-like krypton [6]. The wavelength $\lambda = 0.9454$ Å corresponds to a Bragg angle of $\vartheta = 35.6°$. For a reflecting surface area of 6×8 cm^2 bent with a curvature radius of $R = 6$ m, the level of aberrations is comparable with the natural
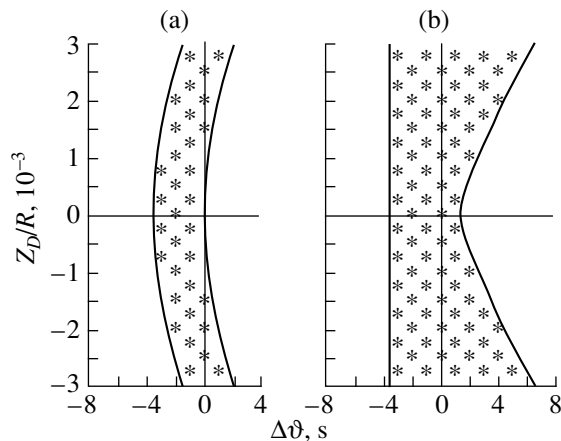
Fig. 2. Schematic diagram of aperture aberrations for a quasimonochromatic spectral line ($\lambda = 0.9454$ Å) in a spectrometer with (a) spherical and (b) cylindrical bent crystal analyzer.

broadening of the spectral line, the line halfwidth corresponding to the Bragg angle variation within $\Delta\vartheta = \pm5.6''$. In this scheme, the role of the crystal analyzer type is clearly manifested.

In a spectrometer with spherical bent crystal analyzer, the monochromatic beam produces an arc-shaped trace on the detector. For $\alpha = 0$, the width of this arc is determined by dimensions of the crystal. The narrower the reflecting surface, the smaller the trace width. In order to retain the relative aperture of the spectrometer on a high level, it is necessary to increase the height of the crystal, which does not influence the level of aberrations. The aberrations can be also reduced by using an oblique-cut crystal analyzer with $\alpha \approx \pi/2 - \vartheta$; this method is used for the measurements at large Bragg angles. In addition, the arc-shaped distortions in a spectrometer can be compensated in a scheme employing a two-coordinate detector.

Aberrations of a different type appear in a spectrometer with cylindrical bent crystal analyzer. These distortions markedly exceed the aberrations observed for the spherical bent crystal, the difference being especially pronounced at large Bragg angles. The large aberrations are manifested at a considerable divergence of beams necessary for the spectrometer filling with radiation. In practice, the tokamak plasma diagnostics is performed via a horizontal beryllium slit window, which restricts the divergence of the X-ray beam (thus decreasing the level of aberrations) at the expense of at least a twofold loss in the relative aperture. Note that the sphere and cylinder can be considered as particular cases of the toroidal surface, and the toroidal bent crystal analyzer possesses intermediate characteristics between the two.

It should be recalled that our analysis referred to the case of X-ray radiation emitted from an extended source. The beam path in the spectrometer can be reversed to find the contour of a virtual diaphragm through which the radiation fills the spectrometer. The X-ray beams from a point source placed on the focal circle are reflected from a small part of the crystal analyzer. For example, the reflection zone on a cylindrical surface has a X-cross shape (see also [13]). The aperture aberrations observed for a local source depend on the working surface area of the crystal analyzer.

REFERENCES

1. H. H. Johann, Z. Phys. **69** (3–4), 185 (1931).

2. M. Bitter, S. von Goeler, R. Horton, *et al.*, Phys. Rev. Lett. **42**, 304 (1979).

3. R. Bartiromo, F. Bombarda, R. Gianella, *et al.*, Rev. Sci. Instrum. **60**, 237 (1989).

4. A. A. Rusakov, D. M. Kheĭker, and G. G. Sazonov, Kristalografiya **35**, 1057 (1990).

5. V. A. Bryzgunov, A. B. Gil'varg, and A. N. Svechkopal, Fiz. Plazmy **20**, 596 (1994).

6. V. A. Bryzgunov and A. B. Gil'varg, *Proc. 23rd EPS Conf. on Controlled Fusion and Plasma Physics*, Ed. by D. Gresillon, A. Sitenko, and A. Zagorodny (Kiev, 1996) [JEPS **20**, 1116 (1996)].

7. B. K. F. Young, A. L. Osterheld, D. F. Price, *et al.*, Rev. Sci. Instrum. **69**, 4049 (1998).

8. A. B. Gil'varg, Byull. Izobret. No. 31, 185 (1977).

9. T. Johansson, Z. Phys. **82**, 507 (1933).

10. Y. Cauchois and C. Bonnelle, in *Atomic Inner-Shell Processes*, Ed. by Grasemann (Academic Press, New York, 1975), Vol. 2, p. 83.

11. J. Eggs and K. Ulmer, Z. Angew. Phys. **20**, 118 (1966).

12. M. Bitter, B. Fraenkel, K. W. Hill, *et al.*, Rev. Sci. Instrum. **66**, 530 (1995).

13. S. A. Ditsman, Izv. Akad. Nauk SSSR, Ser. Fiz. **24**, 376 (1960).

Translated by P.P. Pozdeev

$=$ **ACOUSTICS, ACOUSTOELECTRONICS** $=$

Variation of the Ultrasonic Velocity in Al under Plastic Deformation

L. B. Zuev, B. S. Semukhin, and K. I. Bushmeleva

Institute of Strength Physics and Materials Science, Siberian Division,
Russian Academy of Sciences, Tomsk, 634021 Russia
Received October 1, 1998

Abstract—Variation of the velocity of ultrasound propagation in polycrystalline aluminum under plastic deformation is studied. The dependences of the velocity of ultrasound on the strain and the actual stress are found to consist of three distinct stages. The study of the complex shapes of these dependences allows one to reveal additional stages in the parabolic stress–strain curve of the plastic flow, these features being impossible to observe by conventional methods. The behavior of the ultrasonic velocity observed in the experiment is explained by the changes in the defect structure of the material under deformation. © *2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The commonly accepted concepts of the process of plastic deformation are based on the data testifying that it occurs in stages [1], and the adequacy of the models developed for describing this process implies a strict correspondence between the stages and the governing microscopic mechanisms. The determination of the boundaries of these stages and the corresponding mechanisms presents a complicated problem, especially for polycrystals, because, in most cases, reliable and informative external manifestations of the changes in the deformation mechanisms are absent. For such purposes, the data on the integral characteristics of the material, e.g., magnetization and electric resistance, can be useful. They are especially valuable, because, in contrast to microscopic studies, they can be obtained immediately in the course of mechanical tests without any additional operations related to the preparation of special samples for the analysis.

From this point of view, acoustical methods of studying the properties of solids are quite promising. In addition to the commonly used effects such as the acoustic emission from mechanically loaded samples [2] or the amplitude dependence of internal friction [3], which can be interpreted on the basis of specially developed theoretical models, it is possible to use a more easily measured characteristic, namely, the velocity of ultrasound propagation *V*. In the experiments with Al single crystals [4], it was found that the ultrasonic velocity slowly varies at the beginning of the tensile deformation of these crystals. Other experiments [5] revealed the variations of the resonance frequency proportional to *V* in Al single crystals in the process of their bending. These variations were explained by the changes that occurred in the attenuation with increasing dislocation density in the course of the plastic flow. Evidently, such an effect cannot be large, because

$V \approx (E/\rho_0)^{1/2}$, but neither the elastic modulus E nor the density ρ_0 exhibit any considerable variations in the process of the plastic flow [6]. However, with modern equipment, these variations can be measured to a fair accuracy [7].

Below, we present the data on the dependences of the velocity V on the total strain ε and the actual stress σ for deformations up to fracture (the range of deformations exceeding that studied before [4, 5]). The ultrasonic velocity was measured for transverse waves of frequency 2.5 MHz in samples subjected to tension on an Instron-1185 test machine. The measurements were performed by the method of autocirculation of ultrasonic pulses with the use of a specially designed instrument [7]. The measurement accuracy was $\sim 10^{-4}$. The distance between the piezoelectric transducers was fixed and did not change as the sample was elongated under tension. The samples had the form of a double spade with the working part of length 50 mm. The material was polycrystalline aluminum (the grain size 5–10 mm) whose stress–strain curve in the state of flow is nearly parabolic in the entire range of deformation [1, 8]. In most cases, the studies of this material revealed no particular stages of its flow.

EXPERIMENTAL RESULTS

The measurements were performed in the course of the entire process of tension at 0.5% intervals. The analysis of the dependences $V(\varepsilon)$ and $V(\sigma)$ revealed a number of interesting effects. It was found that the ultrasonic velocity considerably varies with tension, and the dependences $V(\varepsilon)$ and $V(\sigma)$ are fairly complicated. For example, the curve $V(\varepsilon)$ obtained for polycrystalline aluminum and shown in Fig. 1 is *N*-shaped with three well-defined stages [9]. Such a shape testifies to the difference between the mechanisms govern-
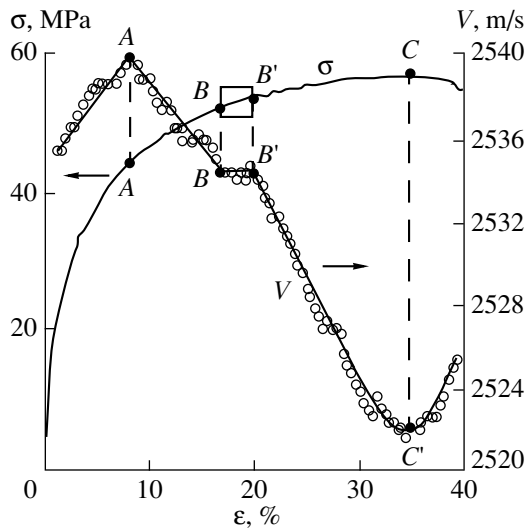
Fig. 1. Stress–strain curve of the plastic flow for an Al polycrystal and the dependence of the velocity of ultrasound propagation in it on the strain.
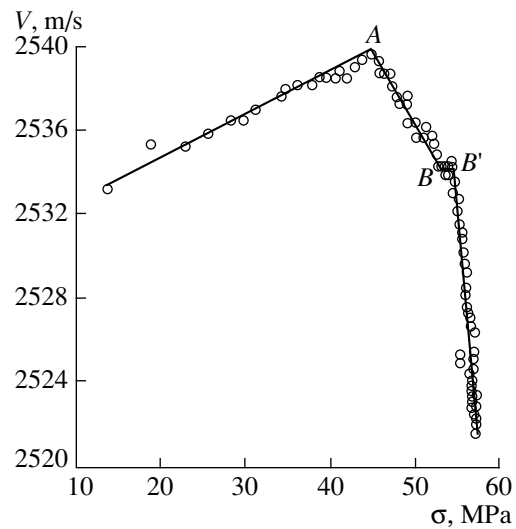
Fig. 2. Dependence of the ultrasonic velocity on the actual stress in polycrystalline Al.
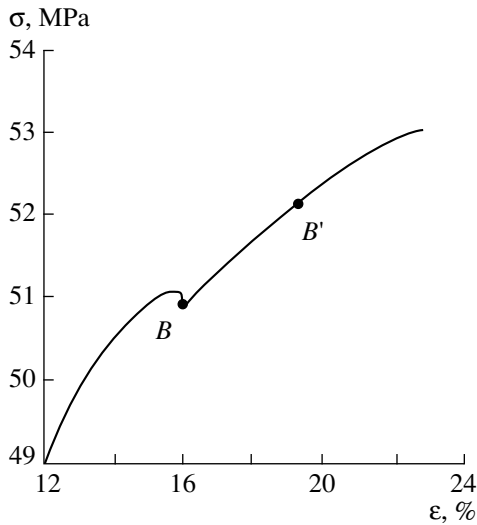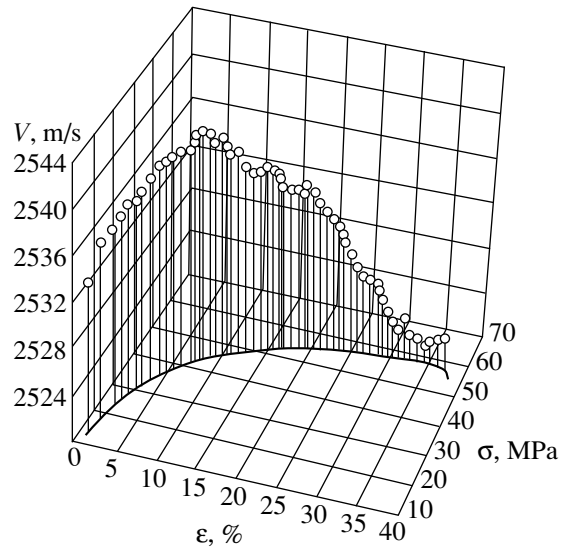

Fig. 3. Portion of the stress–strain curve of the plastic flow corresponding to a constant value of the ultrasonic velocity.

Fig. 4. Variations of the ultrasonic velocity along the stress–strain curve of the plastic flow of an Al polycrystal.

ing the relation between V and ε in the corresponding intervals of plastic deformation.

The dependence $V(\sigma)$ obtained for the same sample and shown in Fig. 2 also exhibits three stages, and, at each stage, the quantities V and σ are related by a linear law. The breakpoints indicated in the curves of Figs. 1 and 2 by the symbols A, B, B', and C divide the stress–strain curve of the flow (Fig. 1) into stages. The behavior of the experimental dependences $V(\varepsilon)$ and $V(\sigma)$ points to the fact that the stress–strain curve of the plastic flow has a more complex shape than that of the parabola $\sigma \sim \varepsilon^{1/2}$. The representation of the portion B–B' of the stress-strain curve on an enlarged scale

(Fig. 3) reveals a linear stage $\sigma \sim \varepsilon$ within this interval. Earlier [8], it was found that, at a temperature of 77 K, the stress–strain curve of the plastic flow of aluminum samples consists of two parabolic parts separated by the stage of linear strain hardening. The latter stage becomes shorter with increasing temperature and virtually vanishes at 300 K. There is reason to believe that the ultrasonic method will allow one to detect this very moment of transition between the two parabolic stages, such a detection being impossible by using only the dependence of the hardening coefficient on the strain.

Figure 4 generalizes the data on the behavior of the velocity V along the entire stress–strain curve of the

$\rho \times 10^9$, cm^{-2}
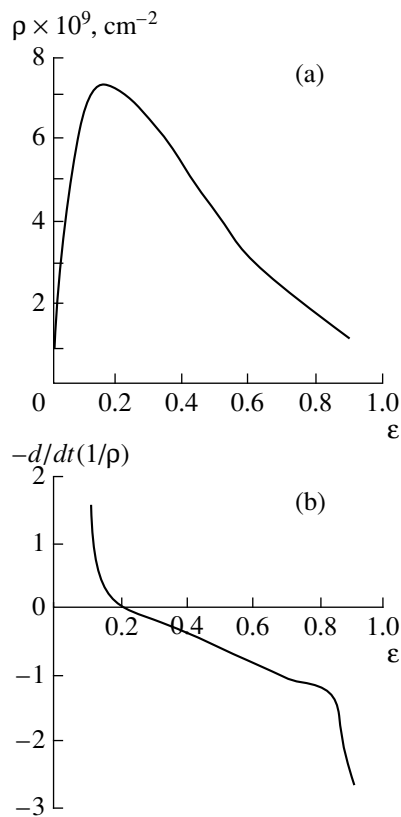
Fig. 5. Density of moving dislocations [10] and the behavior of the coefficient $D^{(\varepsilon)}$ (schematic representation) as functions of strain.

plastic flow $\sigma(\varepsilon)$. Near the end of the tension process, a drop in the deformation force is observed. It is well known that such a drop corresponds to the formation of a macroscopic neck. From Figs. 1 and 2, one can see that, in this region, the ultrasonic velocity increases. The increase in the ultrasonic velocity begins before the appearance of a noticeable neck and is an indication of the precritical state of the deformed material. This state may also be identified by the appearance of a localized maximum of the components of the plastic distortion tensor [10, 11], which can be detected well before the appearance of the neck and is evidence of the formation of a center of viscous fracture. The strain values corresponding to these two effects coincide. Hence, with the use of the two methods in parallel, the predictions of the moment of fracture will be more reliable.

DISCUSSION

From Fig. 2, it follows that the quantities V and σ are related by a linear law

$$V = V_0 + \xi\sigma, \qquad (1)$$

where the constants V_0 and ξ are different for different portions of the dependence $V(\sigma)$. Since, according to the experimental data, the coefficient ξ changes sign in

the process of the plastic flow, it should depend on some parameter of the material that behaves in a similar way in this process. In the case of plastic deformation under study, the number of such parameters is limited. However, it is known that the dependence of the density of moving dislocations ρ on the strain exhibits a peak (Fig. 5a) [12], and, hence, the derivative of ρ with respect to time t (or to strain ε, because, in the case of an active loading, $\varepsilon \sim t$) should change sign. In the case under study, it is appropriate to introduce the derivative $d/dt(1/\rho) \equiv D^{(\varepsilon)}$ having the dimensionality of the "diffusion coefficient" (m^2/s). The variation of the quantity $D^{(\varepsilon)}$ with time (Fig. 5b) agrees well with the above speculations concerning the sign of the coefficient ξ. The meaning of $D^{(\varepsilon)}$ should be discussed separately. In our previous papers [10, 11], we showed that every stage of the plastic flow has its own specific pattern of localized plastic deformation. These patterns were interpreted as autowaves of plastic deformation [10, 11]. The reaction-diffusion differential equations describing the autowaves [13] involve the coefficient $D^{(\varepsilon)}$ related to the density of moving dislocations and determining the process of the propagation of plastic deformation. The dimensionality of the coefficient ξ in (1) is $[L]^2[T][M]^{-1}$ (m^{2}s kg^{-1}). Taking into account that the velocity of the elastic wave propagation in a material is related to the density ρ_0 of this material, we apply the analysis of dimensionality and obtain the following relation:

$$\xi \sim L_i / \rho_0 D^{(\varepsilon)}. \qquad (2)$$

Here, the quantity L_i is interpreted as the spatial scale of the current internal level of plastic deformation [14] (the characteristic size of the region of the deformation nonuniformity). Then, relation (2) formally explains the discrete variation of the coefficient ξ by switching to another scale level of the process [14], while the behavior of $D^{(\varepsilon)}$ determines the change of sign of ξ in the process of plastic deformation. Evidently, the slopes of the portions of the dependence $V(\sigma)$ remain constant only when the condition $L_i/D^{(\varepsilon)} = \text{const}$ is satisfied.

Using relation (2) for the proportionality coefficient ξ and introducing the wave resistance of the medium $Z = V\rho_0$ [15], we represent equation (1) in the form

$$\Delta Z = Z - Z_0 = (L_i/D^{(\varepsilon)})\sigma. \qquad (3)$$

From this expression, it follows that the wave resistance of the material varies under the plastic deformation in response to the variations in the parameters of the internal structure of the material from one stage of the process to another. This conclusion was confirmed by the experiment in which, simultaneously with recording the stress–strain curve at small strains, not only the velocity of ultrasound was measured, but also the X-ray topographic data were obtained on the motion of large-size structure elements in single grains

of the Al polycrystal [16]. The X-ray topographs were obtained by the Fujiwara method [17] in the course of tension of a polycrystalline aluminum sample (without unloading). The results obtained by measuring the ultrasonic velocity under such conditions are presented in Fig. 6. The analysis of the behavior of the grain fragments by their Laue reflections showed that, with increasing strain at $\varepsilon < \varepsilon^*$, the misorientations of the grain fragments are enhanced; simultaneously, the ultrasonic velocity rapidly increases. As the strain exceeds the critical value ε^*, an inverse rotation of the fragments is observed, and the increase in the ultrasonic velocity becomes slower.

Phenomenologically, the variation of the ultrasonic velocity is explained by the presence of a mosaic of stressed regions in the deformed medium with their distribution and size varying in the process of the plastic flow. Let us estimate the possible velocity variations by using the model proposed in [18]. Owing to the elastic stresses, a transverse ultrasonic wave is split into two beams with the polarizations perpendicular to each other (acoustic birefringence). The beams propagate with different velocities v_1 and v_2. For these beams, the path-length difference per unit path will be [18]

$$\delta = (\omega/2\pi)(1/v_2 - 1/v_1). \tag{4}$$

On the other hand, the path-length difference depends on the difference between the principal stresses σ_2 and σ_1 in the crystal [18]:

$$\delta = (\omega/4\pi\rho_0 v_s^3)(\sigma_2 - \sigma_1), \tag{5}$$

where ω is the frequency of ultrasonic vibrations and v_s is the velocity of the propagation of transverse ultrasonic waves in an unstressed crystal.

Introducing the notation $\Delta V = v_2 - v_1$ and $v_2 v_1 \approx v_s^2$ and using expressions (4) and (5), we obtain the variation of the ultrasonic velocity

$$\Delta V \approx (\sigma_2 - \sigma_1)/2\rho_0 v_s. \tag{6}$$

Using expression (3), we obtain the variation of the wave resistance

$$\Delta Z \approx (\sigma_2 - \sigma_1)/2 v_s. \tag{7}$$

Thus, the velocity variation observed in the experiment testifies to the changes in the parameters of the stressed regions of the medium in the process of the plastic flow. This behavior can be attributed to the formation of dislocation ensembles or relatively large-scale features of the deformation structure such as an ordered (self-organizing) system of stationary mesoscopic localization centers of plastic deformation [10, 11], which is typical of the plastic deformation of aluminum [19].
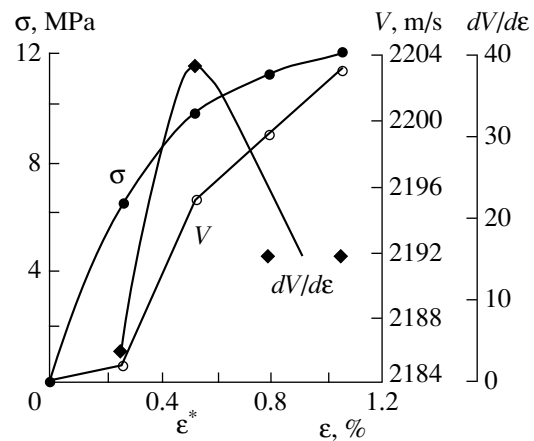
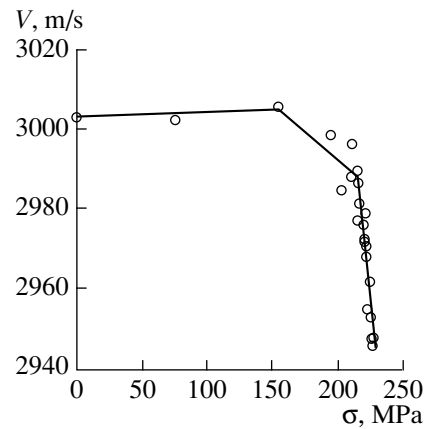**Fig. 6.** Variation of the ultrasonic velocity in the small strain region.

Fig. 7. Dependence of the ultrasonic velocity on the actual stress for a polycrystalline Fe + 3% Si alloy.

CONCLUSION

It is evident that the analysis of the variation of ultrasonic velocity under the tensile deformation reveals additional stages in the stress–strain curve of the plastic flow. These stages are presumably determined by the difference in the types of the dislocation ensembles (cells, fragments, etc. [14]) formed in the material at different stages of the plastic flow. It is believed that the change from one stage to another in the dependence $V(\sigma)$ is related to the formation of dislocation ensembles of different characteristic dimensions L_i and with different internal stress levels in the cell or in a fragment. It is this phenomenon that determines the dependence of the ultrasonic velocity on the strain or stress. Such changes in the deformation mechanisms have virtually no effect on the behavior of the stress–strain curve of the plastic flow and the coefficient of strain hardening, while their effect on the ultrasonic velocity proves to be substantial. The *in situ* measurements of the ultrasonic velocity in the course of mechanical tests of materials may provide important

additional information on the physics of the deformation processes.

The effect observed in our experiments is fairly universal. Similar variations of the ultrasonic velocity were observed by us in the course of a tensile plastic deformation of the samples of Al–Mg–Li alloy in different structural states, Zr + 1.5% Nb alloy, and iron silicide (Fe + 3% Si). As an example, we present the dependence $V(\sigma)$ for the latter material in Fig. 7. This dependence only quantitatively differs from that shown in Fig. 2. In all aforementioned cases, the stress–strain curve of the plastic flow had different stages that could not be revealed immediately from the shape of the curve.

Thus, the measurements of the velocity of ultrasound propagation in materials subjected to plastic deformation allow one to analyze the stages of the stress–strain curves of the plastic flow in the course of the experiment and to predict, for example, the transition to the state of viscous fracture prior to the appearance of its visible signs.

REFERENCES

1. A. Seeger, in *Dislocations and Mechanical Properties of Crystals. An International Conference Held at Lake Placid September 6–8, 1956* (Wiley, New York, 1957; Inostrannaya Literatura, Moscow, 1960).

2. V. S. Boĭko, R. I. Garber, and A. M. Kosevich, *Reversible Plasticity of Crystals* (Nauka, Moscow, 1991).

3. A. B. Lebedev, Fiz. Tverd. Tela (Leningrad) **35,** 2305 (1993) [Sov. Phys. Solid State **35,** 1141 (1993)].

4. A. Hikata, B. B. Chick, and C. Elbaum, Acta Metal. **10,** 423 (1962).

5. A. B. Lebedev, S. B. Kustov, and B. K. Kardashev, Fiz. Tverd. Tela (Leningrad) **34,** 2915 (1992) [Sov. Phys. Solid State **34,** 1562 (1992)].

6. T. Thomas, *Plastic Flow and Fracture in Solids* (Academic, New York, 1961; Mir, Moscow, 1964).

7. V. V. Murav'ev, L. B. Zuev, and K. L. Komarov, *Sound Velocity and the Structure of Steels and Alloys* (Nauka, Novosibirsk, 1996).

8. B. Jaoul, J. Mech. Phys. Solids **5,** 95 (1957).

9. L. B. Zuev, B. S. Semukhin, and K. I. Bushmeleva, Pis'ma Zh. Tekh. Fiz. **22,** 76 (1996) [Tech. Phys. Lett. **22** (7), 590 (1996)].

10. L. B. Zuev and V. I. Danilov, Int. J. Solids Structures **34,** 3795 (1997).

11. L. B. Zuev and V. I. Danilov, Fiz. Tverd. Tela (Leningrad) **39,** 1399 (1997) [Physics of the Solid State **39** (8), 1241 (1997)].

12. J. Gilman, in *Microplasticity*, Ed. by V. N. Geminov and A. G. Rakhshtadt (Metallurgiya, Moscow, 1972), pp. 18–37.

13. V. A. Vasil'ev, Yu. M. Romanovskiĭ, and V. G. Yakhno, *Autowave Processes* (Nauka, Moscow, 1987).

14. V. E. Panin, Yu. V. Grinyaev, V. I. Danilov, *et al.*, *Structural Levels of Plastic Deformation and Fracture* (Nauka, Novosibirsk, 1990).

15. L. Bergmann, *Ultrasound and Its Application in Science and Engineering* (Zürich, 1954; Inostrannaya Literatura, Moscow, 1956).

16. L. B. Zuev, T. M. Poletika, and B. S. Semukhin, Kristallografiya **40,** 1071 (1995) [Crystallography Reports **40,** 996 (1995)].

17. A. Guinier, *Theory and Methods of X-ray Crystallography* (Dunod, Paris, 1956; Gos. Izd. Fiz.-Mat. Lit., Moscow, 1961).

18. T. Tokuoka and Yu. Iwashimizu, Int. J. Solids Structures **4,** 383 (1968).

19. V. I. Danilov, L. B. Zuev, N. M. Mnikh, *et al.*, Fiz. Met. Metalloved., No. 3, 188 (1991).

Translated by E.M. Golyamina

=========== **RADIOPHYSICS** ===========

A New Class of Anisotropic Nonexchange Spin Waves

S. V. Tarasenko

Donetsk Physicotechnical Institute, National Academy of Sciences of Ukraine, Donetsk, 340114 Ukraine
Received October 14, 1996; in final form, July 26, 1999

Abstract—It is shown, using a thin magnetic film of a tetragonal antiferromagnet as an example, that indirect spin–spin exchange via the long-range field of quasi-static magnetoelastic deformations with allowance for the magnetocrystalline anisotropy results in qualitatively new types of nonexchange propagating spin waves, both surface and bulk. The case of an isolated magnetic film and a magnetic film with a one-sided or two-sided coating is studied. © *2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

It was shown previously [1–4] that, if the indirect spin–spin exchange in a thin magnetic film is effected through the magnetodipolar interaction, then the rigorous allowance for the magnetocrystalline anisotropy may lead to new types of propagating magnetic excitations—anisotropic dipole spin waves. However, because of the exchange weakening of the magnetodipole interaction, the efficiency of this mechanism of the formation of nonexchange spin-wave excitations is reduced sharply in antiferromagnetic films. At the same time, it was shown in [5] that, if the frequency of spin waves ω and the projection of their wave vector k onto the magnetic film plane ($k_\perp$) satisfy the elastostatic criterion [6]

$$\omega \ll sk_\perp, \tag{1}$$

where s is the minimum phase velocity of the sound wave propagation, then an indirect exchange-enhanced (in the antiferromagnets) spin–spin exchange through the long-range field of quasi-static magnetoelastic deformations may become an alternative to the magnetodipolar mechanism of the formation of nonexchange spin waves. The corresponding class of propagating nonexchange spin-wave excitations may be called elastostatic spin waves (ESSWs) [5], since condition (1) in the continuum elasticity theory corresponds to the elastostatic approximation, and the conditions of the formation of ESSWs are mainly similar to the formation of magnetostatic spin waves (MSSWs). However, the conditions for the formation and the dispersion properties of all ESSWs that were studied heretofore, both the surface and the bulk ones, were independent of the nature of the magnetocrystalline anisotropy of the real magnetic sample. Therefore, from this viewpoint, they may be called isotropic elastostatic spin waves (IESSWs).

The aim of this work is to determine the necessary conditions under which the effect of magnetocrystalline anisotropy on the character of spin–spin exchange through the field of "elastostatic phonons" may lead,

when subject to condition (1), to the formation of a new class of propagating bulk and surface spin-wave excitations in finite magnets—anisotropic spin waves (AESSWs).

BASIC RELATIONSHIPS

As an example, we consider a two-sublattice model of an easy-plane tetragonal antiferromagnet (AFM) with sublattice magnetizations $\mathbf{M}_1$ and $\mathbf{M}_2$ [7, 8]. The specific features of the spin dynamics of this class of magnetic crystals is known to be related to the existence of the Dzyaloshinskiĭ interaction, which depends on the parity of the antiferromagnetic structure with respect to the fourfold axis (0Z). If this axis is odd (4_z^-), then, in the terms of the vectors of ferromagnetism (m) and antiferromagnetism (l), this interaction $D_{ik}m_i l_k$ contains the invariants ($|\mathbf{M}_1| = |\mathbf{M}_2| = M_0$)

$$d_1(m_x l_y + m_y l_x), \quad d_1(m_x l_x - m_y l_y). \tag{2}$$

In terms of a phenomenological approach, the energy density W in the discussed model of an arbitrary two-sublattice antiferromagnet, which takes into account the interaction of the spin and elastic subsystems, may be represented by a functional of the vectors $\mathbf{m}$ and $\mathbf{l}$, their spatial derivatives, and the tensor of elastic deformations u_{ik}. In order to clearly demonstrate the effect of magnetocrystalline anisotropy on the formation of elastostatic spin waves, we assume the magnetoelastic and elastic properties of the crystal to be isotropic:

$$W = W_m + W_{me} + W_e,$$

$$W_m = 2M_0^2 \left\{ \frac{\delta}{2}\mathbf{m}^2 + \frac{\alpha}{2}[(\nabla \mathbf{l})^2] + W_a\{\mathbf{l}\} \right. \tag{3}$$

$$\left. + \frac{2}{g\delta M_0}D_{ij}m_i l_j - 2\mathbf{mh} \right\},$$

$$W_{me} = 2M_0^2 \gamma l_i l_k u_{ik}, \quad W_e = \frac{\lambda}{2} u_{ii}^2 + \mu u_{ik}^2.$$

Everywhere below, we also assume that the condition

$$|\mathbf{m}| \ll 1 \tag{4}$$

is fulfilled.

Following [8], we can show that, under condition (1), both the linear and nonlinear elastoexchange spin dynamics of the discussed model of a magnet may be described using the following set of dynamic equations that couple the vector of antiferromagnetism $\mathbf{l}$ and the vector of elastic lattice displacements $\mathbf{u}$:

$$\alpha \left[\mathbf{l} \left(\Delta \mathbf{l} - \frac{1}{c^2} \mathbf{l}_{tt} - \frac{\partial W_{me}}{\partial \mathbf{l}} \right) \right]_i - \frac{8}{g \delta M_0} (\mathbf{lh}) \frac{\partial l_i}{\partial t} - \frac{4}{\delta} (\mathbf{lh})[\mathbf{lh}]_i$$

$$+ \frac{2}{g \delta M_0} \left(\varepsilon_{ijk} l_j D_{pk} \varepsilon_{prs} l_r \frac{\partial l_s}{\partial t} + 2R \frac{\partial l_i}{\partial t} + l_i D_{pk} l_p \frac{\partial l_k}{\partial t} D_{ik} \frac{\partial l_k}{\partial t} \right)$$

$$+ \frac{2}{\delta} \varepsilon_{ijk} l_j ((\mathbf{lh})(D_{kr} + D_{rk}) l_r + h_k R - D_{rk} h_r) = 0, \tag{5}$$

$$\frac{\partial^2 W}{\partial x_k \partial u_{ik}} \equiv 0, \quad D_i \equiv D_{ik} l_k, \quad R \equiv D_{ik} l_i l_k.$$

The effective energy of magnetic anisotropy W_a for the vector $\mathbf{l}$ in this case is

$$W_a = \frac{\beta}{2} (l_x^2 + l_y^2) + \frac{1}{4} (b(l_x^2 + l_y^2)^2 + 4\tilde{b} l_x^2 l_y^2), \tag{6}$$

whereas the vector of ferromagnetism $\mathbf{m}$ may be determined from the equation

$$m = \frac{2}{\delta} \left\{ \frac{1}{g M_0} [\mathbf{l},\mathbf{l}] + \mathbf{h} - \mathbf{l}(\mathbf{lh}) \right\} + \frac{1}{2} (\mathbf{lR} - \mathbf{D}). \tag{7}$$

If we write the set of dynamic equations (5) as a set of Euler–Lagrange equations, then the density of the corresponding Lagrangian function L may be written as

$$L = M_0^2 \left\{ \frac{\alpha}{2} \left[\frac{1}{c^2} \mathbf{l}_t^2 - (\nabla \mathbf{l})^2 \right] - W_a\{\mathbf{l}\} \right.$$

$$\left. + \frac{2}{g \delta M_0} D_{ij} \left[\mathbf{l}, \mathbf{l}_t l_i l_j + l_i \left(\frac{2}{\delta} (\mathbf{R} h_i - D_{ik} h_k) \right) \right] \right\} \tag{8}$$

$$+ \frac{\rho}{2} \mathbf{u}_t^2 - W_{me} - W_e.$$

Since we are interested in the elastoexchange dynamics of a thin magnetic film, then to solve the boundary value problem the set of dynamic equations (5) and (6) should be supplemented with appropriate boundary conditions. In the case of a thin isolated anti-

ferromagnetic film of thickness d and fully free spins, such a set of boundary conditions has the form

$$\frac{\partial \mathbf{l}}{\partial \xi} = 0; \quad \sigma_{ik} n_k = 0; \quad \xi = 0, d. \tag{9}$$

If, on the contrary, the sample is completely "constrained," then the following set of boundary conditions should be used along with (5) and (6):

$$\frac{\partial \mathbf{l}}{\partial \xi} = 0; \quad u_i = 0; \quad \xi = 0, d. \tag{10}$$

Using the condition $|\mathbf{l}| = 1$, it is convenient to represent the set of dynamic equations (5) and (6) and the corresponding boundary conditions (9) or (10) in terms of polar (ϑ) and azimuthal (φ) angles, which are determined from the relation

$$l_x = \sin\vartheta\cos\varphi; \quad l_y = \sin\vartheta\sin\varphi; \quad l_z = \cos\vartheta. \tag{11}$$

In this case, as follows from (5) and (6), the easy-plane character of the equilibrium orientation of the antiferromagnetism vector corresponds to $\beta < 0$. Depending on the character of anisotropy in the basal plane (determined by the magnitude of $\tilde{b}$), the ground state corresponds to the following equilibrium orientation of the antiferromagnetism vector $\mathbf{l}$:

$$\vartheta = \pi/2, \quad \varphi = 0, \pi/2, \quad \tilde{b} > 0 \tag{12}$$

or

$$\vartheta = \pi/2, \quad \varphi = \pi/4, 3\pi/4, \quad \tilde{b} < 0. \tag{13}$$

Since in this work we are interested in the conditions for the formation of anisotropic ESSWs, we, to obtain analytical expressions, restrict ourselves to the consideration of the geometry of the propagation of spin waves such that $\mathbf{u} \perp \mathbf{k}$. From (12) and (13) and a comparison of the results obtained with the conditions for the existence of "isotropic" ESSWs [5], we obtain that new "anisotropic" types of propagating elastostatic spin waves arise at $\mathbf{u} \parallel \mathbf{l}$. In this case, however, the structure of the spectrum of these types of traveling nonexchange spin-wave excitations substantially depends on the equilibrium orientation of the antiferromagnetism vector in the *XY* plane, according to (12) and (13).

ISOLATED MAGNETIC FILM

If $\vartheta = \pi/2$ and $\varphi = 0$ (or $\varphi = \pi/2$) (see (12)), the indirect spin–spin exchange through the field of elastostatic phonons with lattice displacements $u_x \neq 0$ ($u_y \neq 0$) leads to the formation of bulk anisotropic ESSWs whose wave vector is $\mathbf{k} \in YZ$ ($\mathbf{k} \in XZ$) and the number of modes ν is an infinite denumerable many. A specific feature of this type of spin-wave excitations is that the constituting modes at a fixed ν and $0 < k_\perp < \infty$ form two nonoverlapping frequency bands. If we designate the frequency of the excitation of the "high-frequency" bulk mode by ω_+ and the "low-frequency" mode by ω_-,

then the regions of existence of the "high-frequency" (ω_+) and "low-frequency" (ω_-) bands of the spectrum of bulk anisotropic ESSWs at $0 < k_\perp < \infty$ are determined by the conditions

$$\omega_c < \omega_+ < \omega_d, \quad \omega_a < \omega_- < \omega_b, \tag{14}$$

respectively.

The frequencies ω_{a-d} are the condensation points of the spectrum of the discussed class of nonexchange spin-wave excitations at $k_\perp \longrightarrow 0$ and can be determined from the dispersion relations. An analysis shows that the frequency ranges of existence of these types of nonexchange magnetic excitations do not overlap ($\omega_b < \omega_c$) and are determined by the magnitude of the magnetic anisotropy. The character of the dispersion curve of each of the branches depends on the relative orientation of the normal to the film surface $\mathbf{n}$ and the direction of the equilibrium vector of antiferromagnetism $\mathbf{l}$. Qualitatively different situations occur at $\mathbf{n} \parallel 0Y$ and $\mathbf{n} \parallel 0Z$ ($4c^2 = g^2 M_0^2 \delta\alpha$, $\omega_{me}^2 = c^2\gamma^2 M_0^2/\mu\alpha$):

$$\frac{k_\perp^2}{(\pi\nu/d)^2} = -\frac{\mu_\parallel}{\mu_\perp}, \tag{15}$$

$$\mu_\parallel \equiv \frac{A_{11}A_{22} - \omega_{me}^2 A_{22}}{\Delta}, \quad \mu_\perp \equiv \frac{A_{11}A_{22} - \omega_{me}^2 A_{11}}{\Delta},$$

$$\mathbf{n} \parallel 0Z,$$

$$\mu_\parallel \equiv \frac{A_{11}A_{22} - \omega_{me}^2 A_{11}}{\Delta}, \quad \mu_\perp \equiv \frac{A_{11}A_{22} - \omega_{me}^2 A_{22}}{\Delta}, \tag{16}$$

$$\mathbf{n} \parallel 0Y,$$

$$\Delta = A_{11}A_{22}, \quad A_{11} = \omega^2 - \omega_{me}^2 - (|\beta| + b)c^2/\alpha,$$
$$A_{22} = \omega_{me}^2 - \omega^2 - 2\tilde{b}c^2/\alpha. \tag{17}$$

If we now investigate, using (15)–(17), the character of the dispersion curves of the thus-found nonexchange spin-wave modes, we may easily show that, at $\mathbf{n} \parallel 0Y$, the high-frequency anisotropic bulk ESSW is a wave of the forward type ($\partial\omega/\partial k_\perp > 0$; $\omega_+(k_\perp \longrightarrow 0) \longrightarrow \omega_c$; $\omega_+(k_\perp \longrightarrow \infty) \longrightarrow \omega_d$). An opposite situation takes place for the considered spectrum of anisotropic bulk ESSWs at $\mathbf{n} \parallel 0Z$. The low-frequency branch of the spectrum is simultaneously a backward-type wave ($\partial\omega/\partial k_\perp < 0$) ($\omega_-(k_\perp \longrightarrow 0) \longrightarrow \omega_b$); $\omega_-(k_\perp \longrightarrow \infty) \longrightarrow \omega_a$). The situation reverses for $\mathbf{n} \parallel 0Z$. In this case, the high-frequency modes, defined by (15)–(17) with $\nu \neq 0$, are waves of the backward type ($\omega_+(k_\perp \longrightarrow 0) \longrightarrow \omega_d$); $\omega_+(k_\perp \longrightarrow \infty) \longrightarrow \omega_c$), whereas the low-frequency modes with $\nu \neq 0$ are forward-type waves ($\omega_-(k_\perp \longrightarrow 0) \longrightarrow \omega_a$; $\omega_-(k_\perp \longrightarrow \infty) \longrightarrow \omega_b$). As to the anisotropic ESSW modes with $\nu = 0$, their excitation frequency in the considered nonexchange approximation ($c \longrightarrow 0$) is independent of the magnitude of $\mathbf{k}$ ($\mathbf{k} \in YZ$) both at

$\mathbf{n} \parallel 0Y$ and $\mathbf{n} \parallel 0Z$; with allowance for (15)–(17), it is determined by the relation

$$\mu_\parallel^{-1} = 0, \quad \mu_\perp = 0. \tag{18}$$

Let us now consider how the spectrum of the above-found bulk anisotropic elastostatic spin waves will change if the direction of the normal $\mathbf{n}$ to the magnetic film surface lies as before in the YZ plane but makes an angle ψ with the $0Z$-axis ($0 < \psi < \pi/2$). As calculations show, now the structure of the spectrum of bulk ESSWs induced by the indirect spin–spin exchange through the long-range field of elastostatic phonons with $\mathbf{u} \parallel \mathbf{l}$ is determined for a given ψ by the equation that coincides in the limiting cases ($\psi = 0$ or $\psi = \pi/2$) with (15)–(17) at $\nu \neq 0$:

$$\frac{k_\perp^2}{(\pi\nu/d)^2} = -\frac{\mu_\parallel^2}{\mu_s^2 - \mu_\parallel\mu_\perp}, \tag{19}$$

$$\mu_\parallel = \frac{A_{11}A_{22} - \omega_{me}^2 A_{11}\sin^2\psi - \omega_{me}^2 A_{22}\cos^2\psi}{\Delta},$$

$$\mu_\perp = \frac{A_{11}A_{22} - \omega_{me}^2 A_{11}\cos^2\psi - \omega_{me}^2 A_{22}\sin^2\psi}{\Delta}, \tag{20}$$

$$\mu_s = \frac{\omega_{me}^2(A_{22} - A_{11})}{\Delta}\sin 2\psi.$$

The spectrum of the zeroth mode of the spectrum of bulk anisotropic ESSWs with allowance for (19) and (20) consists of four branches:

$$(\mu_\parallel\mu_\perp - \mu_s^2)\mu_\parallel^{-2} = 0. \tag{21}$$

Thus, the range of the existence of anisotropic bulk ESSWs propagating in the YZ plane, as before, consists of two bands that satisfy conditions (14) and (15), i.e., "high-frequency" (ω_+) and "low-frequency" (ω_-) bands, but there are now four rather than two types of branches of the spectrum of bulk ESSWs (Fig. 1). At
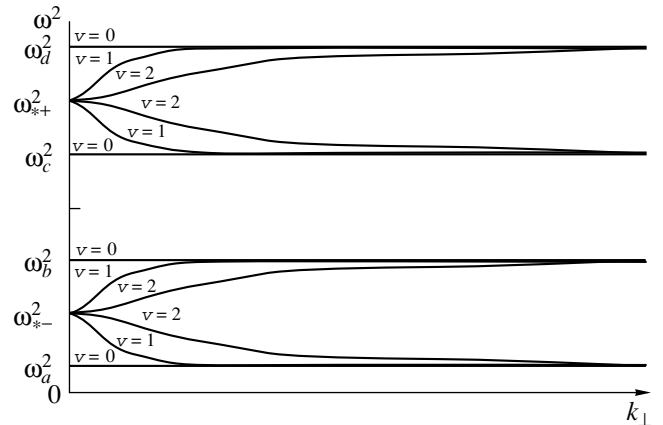

Fig. 1. $0 < \psi < \pi/2$; $\vartheta = \pi/2$; $\varphi = 0, \pi/2$; $t = l = 0$.

$\nu \neq 0$ in (19) and (20), the structure of the spectrum of propagating bulk anisotropic ESSWs changes as follows. At $k_{\perp} \longrightarrow \infty$, four points of condensation exist now in the ESSW spectrum at $k_{\perp} \longrightarrow \infty$, whose values correspond to ω_{a-d}, whereas, at $k_{\perp} \longrightarrow 0$, the number of condensation points is two, as before, although their values are $\omega_{\pm*}$. The following relationships exist between all these characteristic points of the spectrum of bulk anisotropic ESSWs with $\psi \neq 0$:

$$\omega_a < \omega_{*-} < \omega_b < \omega_c < \omega_{*+} < \omega_d. \quad (22)$$

It follows from (19) and (20) that the branches of the spectrum of anisotropic bulk ESSWs at $\omega_{*+} < \omega < \omega_d$ and $\omega_{*-} < \omega < \omega_b$ correspond to the high-frequency and low-frequency waves of the forward types $\omega_+(k_{\perp} \longrightarrow 0) \longrightarrow \omega_{*+}$, $\omega_+(k_{\perp} \longrightarrow \infty) \longrightarrow \omega_d$, $\omega_-(k_{\perp} \longrightarrow \infty) \longrightarrow \omega_b$, and $\omega_-(k_{\perp} \longrightarrow 0) \longrightarrow \omega_{*-}$, whereas, at $\omega_c < \omega < \omega_{*+}$ and $\omega_a < \omega < \omega_{*-}$, they correspond to the high-frequency and low-frequency waves of the backward types $\omega_+(k_{\perp} \longrightarrow 0) \longrightarrow \omega_{*+}$, $\omega_+(k_{\perp} \longrightarrow \infty) \longrightarrow \omega_c$, $\omega_-(k_{\perp} \longrightarrow \infty) \longrightarrow \omega_a$, and $\omega_-(k_{\perp} \longrightarrow 0) \longrightarrow \omega_{*-}$. The spectrum of bulk ESSWs with $\nu = 0$ (21) contains four branches (which are dispersionless in the nonexchange approximation), and their frequencies coincide with ω_{a-d}. The analysis shows that the zeros of (19) and (20) in the limiting cases ($\psi = 0$ or $\psi = \pi/2$) coincide with ω_{a-d} and the structure of the spectrum of propagating anisotropic bulk ESSWs completely corresponds to the above considered set of equations (15)–(18) at any ν. Additional types (apart from those found in the phase with $\varphi = 0$ or $\varphi = \pi/2$) of anisotropic ESSWs arise in cases where the equilibrium orientation of the antiferromagnet vector is determined by condition (14) at $\varphi = \pi/4$, $\vartheta = \pi/2$.

For the same relative orientations of the vectors $\mathbf{n}$ and $\mathbf{l}$ as in (15)–(17), the dispersion relations that determine the spectrum of anisotropic ESSWs with $\mathbf{k} \perp \mathbf{l} \parallel \mathbf{u}$ in the phase with $\varphi = \pi/4$, $\varphi = \pi/2$ are written as ($\omega_d^2 = 6dc^2/\alpha$)

$$(\mu_{\parallel}^2 \kappa^2 - \mu_*^2 k_{\perp}^2)\tan(\kappa d) = 0; \quad \kappa^2 \equiv -\frac{\mu_{\perp}}{\mu_{\parallel}}k_{\perp}^2,$$

$$\mu_{\parallel} = \frac{A_{11}A_{22} - \omega_{me}^2 A_{22} - \omega^2 \omega_d^2}{\Delta}, \quad \mathbf{n} \parallel 0Z, \quad (23)$$

$$\mu_{\perp} = \frac{A_{11}A_{22} - \omega_{me}^2 A_{11} - \omega^2 \omega_d^2}{\Delta}, \quad \mathbf{n} \parallel 0Z,$$

$$\mu_{\parallel} = \frac{A_{11}A_{22} - \omega_{me}^2 A_{11} - \omega^2 \omega_d^2}{\Delta}, \quad \mathbf{n} \perp 0Z,$$

$$\mu_{\perp} = \frac{A_{11}A_{22} - \omega_{me}^2 A_{22} - \omega^2 \omega_d^2}{\Delta}, \quad \mathbf{n} \perp 0Z, \quad (24)$$

$$\mu_* = i\frac{\omega \omega_d \omega_{me}^2}{\Delta}, \quad \Delta = A_{11}A_{22} - \omega^2 \omega_d^2,$$

$$A_{11} = \omega^2 - \omega_{me}^2 - (|\beta| + b + 2|\tilde{b}|)c^2/\alpha,$$
$$A_{22} = \omega^2 - \omega_{me}^2 - 2|\tilde{b}|c^2/\alpha. \quad (25)$$

In this case, as in (15)–(17), high-frequency and low-frequency bands of propagating bulk AESSWs ($\kappa^2 > 0$) that satisfy conditions (14) and (15) may be formed. In this case, the dependence of their dispersion characteristics on the orientation of the antiferromagnetism vector $\mathbf{l}$ relative to the orientation of the normal to the magnetic film surface $\mathbf{n}$ does not differ qualitatively from that considered above for the phase with $\vartheta = \pi/2$, $\varphi = 0$. However, a fundamentally new feature of the considered spectrum of exchangeless spin-wave excitations in comparison with (15)–(17) is the formation of two branches of anisotropic surface ESSWs ($\kappa^2 < 0$). It follows from (23)–(25) that these branches are due to the indirect spin–spin interaction through the field of elastostatic phonons with $\mathbf{u} \parallel \mathbf{l}$ under conditions where there exists, among magnetic interactions in the crystal, an interaction that violates the Lorentz invariance of a given magnetic system. In the case at hand, this effect is induced by the presence of the Dzyaloshinskii interaction with $D_{ik} \neq \varepsilon_{ikj}d_j$. In the exchangeless approximation used in our case, all these branches (just as the modes with $\nu = 0$ in the case of the anisotropic ESSWs of the bulk type) are dispersionless. Let us designate the frequencies of these branches of the anisotropic surface ESSWs in the order of increasing frequency $\Omega_{s\pm}$ ($\Omega_{s-} < \Omega_{s+}$). Then, we have from (23)–(25) that the following conditions are fulfilled for $0 < k_{\perp} < \infty$ at both $\mathbf{n} \parallel 0Z$ and $\mathbf{n} \perp 0Z$ with allowance for (14) and (15):

$$\omega_a < \omega_b < \Omega_{s-} < \Omega_{s+} < \omega_c < \omega_d. \quad (26)$$

Note here that the characteristic frequencies ω_{a-d} should be determined now from (23)–(25) at $k_{\perp} \longrightarrow 0$ and $k_{\perp} \longrightarrow \infty$. Let us now analyze, for the given magnetic phase ($\vartheta = \pi/2$, $\varphi = \pi/4$), the structure of the spectrum of anisotropic ESSWs propagating in the plane with the normal along l but with an arbitrary orientation of n in this plane, specified by the angle ψ measured from the 0Z-axis. The corresponding dispersion equation, which simultaneously describes the structure of the spectrum of bulk and surface ESSWs, may be written for an arbitrary ψ ($0 < \psi < \pi/2$) as

$$(\mu_{\parallel}^2 \kappa^2 - \mu_*^2 k_{\perp}^2)\tan(\kappa d) = 0;$$

$$\kappa^2 \equiv \left\{\left(\frac{\mu_s}{\mu_{\parallel}}\right)^2 - \frac{\mu_{\perp}}{\mu_{\parallel}}\right\}k_{\perp}^2; \quad (27)$$

$$\mu_\parallel = \frac{A_{11}A_{22} - \omega_{me}^2 A_{11} - \omega^2\omega_d^2 - \omega_{me}^2(A_{22} - A_{11})\cos^2\psi}{\Delta}, \quad (28)$$

$$\mu_\perp = \frac{A_{11}A_{22} - \omega_{me}^2 A_{11} - \omega^2\omega_d^2 - \omega_{me}^2(A_{22} - A_{11})\sin^2\psi}{\Delta},$$

$$\mu_* = i\frac{\omega\omega_d\omega_{me}^2}{\Delta}, \quad \mu_s = \frac{\omega_{me}^2(A_{22} - A_{11})}{\Delta}\sin 2\psi. \quad (29)$$

As in the phase with $\vartheta \equiv \pi/2$, $\varphi \equiv 0$ (see (19), (20)) at $\psi \neq 0$, the spectrum of propagating anisotropic bulk ESSWs does not differ qualitatively from that found at $\psi \neq 0$ in the phase with $\vartheta = \pi/2$, $\varphi = 0$ ($\Delta = A_{11}A_{22} - \omega^2\omega_d^2$).

In particular, it follows from (27)–(29) that, at $k_\perp \longrightarrow \infty$, there exist four condensation points ω_{a-d}, and, at $k_\perp \longrightarrow 0$, only two points ($\omega_{*\pm}$), which are interrelated by a set of inequalities analogous to (22). With allowance for the above introduced designations ($\Omega_{s\pm}$ are surface waves), at $0 < k_\perp < \infty$, the frequency ranges of existence of both surface and bulk anisotropic ESSWs in this case satisfy the set of inequalities

$$\omega_a < \omega_{*-} < \omega_b < \Omega_{s-} < \Omega_{s+} < \omega_c < \omega_{*+} < \omega_d. \quad (30)$$

As in the case of the phase with $\vartheta = \pi/2$ and $\varphi = 0$, the branches of the spectrum of anisotropic bulk ESSWs for the phase with $\vartheta = \pi/2$, $\varphi = \pi/4$ correspond to high-frequency and low-frequency waves of the forward types at $\omega_{*+} < \omega < \omega_d$ and $\omega_{*-} < \omega < \omega_b$, respectively, and to high-frequency and low-frequency waves of the backward types at $\omega_c < \omega < \omega_{*+}$ and $\omega_a < \omega < \omega_{*-}$, respectively. The spectrum of bulk ESSWs with $\nu = 0$ contains four dispersionless (in the exchangeless approximation) branches, and their frequencies coincide with ω_{a-d}. As follows from (27)–(29), in the limiting cases of $\psi = 0$ or $\psi = \pi/2$, the frequencies $\omega_{*\pm}$ coincide with ω_{a-d}, respectively, and the structure of the spectrum of anisotropic bulk ESSWs completely corresponds to the above-considered set of equations (23)–(25) (Fig. 2).

Up to this point, when considering the conditions of the formation of anisotropic types of ESSWs, we restricted ourselves to the analysis of the case of a thin isolated magnetic film. At the same time, the efficiency of the discussed "phonon" mechanism of the formation of exchangeless spin waves may be affected by changes in the elastic boundary conditions at the surface of the magnetic film. For example, this may be a one-sided or two-sided coating having a continuous acoustic contact with the surface of the magnet.

MAGNETIC FILM WITH A ONE-SIDED OR TWO-SIDED NONMAGNETIC COATING

The set of boundary conditions (9) changes in this case as follows (the parameters of the magnetic and nonmagnetic, elastoisotropic medium will be denoted by indices 1 and 2, respectively; b_s is the constant of surface magnetic anisotropy):

—for the one-sided coating,

$$\alpha\frac{\partial \mathbf{l}}{\partial \xi} + b_s\mathbf{n}(\mathbf{ln}) = 0, \quad \sigma_{ik}^{(1)}n_k = \sigma_{ik}^{(2)}n_k, \quad \xi = d, \quad (31)$$

$$\sigma_{ik}^{(1)}n_k = 0, \quad \xi = 0, \quad \sigma_{ik}^{(2)}n_k = 0, \quad \xi = d + t,$$

—for a two-sided coating,

$$\alpha\frac{\partial \mathbf{l}}{\partial \xi} + b_s\mathbf{n}(\mathbf{ln}) = 0, \quad \sigma_{ik}^{(1)}n_k = \sigma_{ik}^{(2)}n_k, \quad \xi = 0, d, \quad (32)$$

$$\sigma_{ik}^{(2)}n_k = 0, \quad \xi = d + t, \quad \xi = -l.$$

In order to study how the presence of a nonmagnetic coating affects the conditions of the formation and the dispersion properties of anisotropic ESSWs existing in an isolated magnetic film, we investigate the case of $\mathbf{k}_\perp \perp \mathbf{l} \parallel \mathbf{u}$. As calculations show, the presence of a one-sided (31) or a two-sided (32) nonmagnetic coating changes the dispersion equation for the advancing anisotropic ESSWs in the phase with $\vartheta = \pi/2$, $\varphi = 0$ ($\mathbf{k} \in YZ$) as follows:

—for the one-sided nonmagnetic coating,

$$\mu_\parallel \kappa a + k_\perp\tanh(k_\perp t)\cot(\kappa d) = 0, \quad \kappa^2 \equiv -\frac{\mu_\perp}{\mu_\parallel}k_\perp^2, \quad (33)$$

—for the two-sided nonmagnetic coating,

$$\mu_\parallel^2\kappa^2 a^2 + \mu_\parallel\kappa k_\perp a(\tanh(k_\perp t) + \tanh(k_\perp l))\cot(\kappa d)$$
$$+ k_\perp^2\tanh(k_\perp t)\tanh(k_\perp l) = 0, \quad (34)$$

where $\mu_\parallel$ and $\mu_\perp$ at $\mathbf{n} \parallel 0Z$ or $\mathbf{n} \parallel 0Y$ are determined by relations (16) and (17), and $a \equiv \mu_1/\mu_2$.

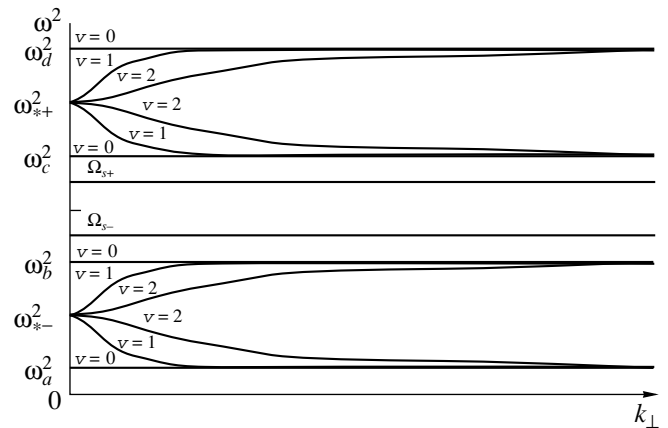**Fig. 2.** $0 < \psi < \pi/2$; $\vartheta = \pi/2$; $\varphi = \pi/4$, $3\pi/4$; $t = l = 0$.

It follows from (33) and (34) that the condensation points of the spectrum of propagating anisotropic bulk ESSWs (ω_{a-d}) are solutions to equations (33) and (34) at $k_\perp \longrightarrow 0$ and $k_\perp \longrightarrow \infty$. With their help and using (33) and (34), it may be shown that at $0 < k_\perp < \infty$, just as in the case of an isolated magnetic film, the spectrum of bulk ESSWs contains both low-frequency modes (ω_-) ($\omega_a < \omega < \omega_b$) and high-frequency modes (ω_+) ($\omega_c < \omega < \omega_d$). As before, at $\mathbf{n} \parallel 0Y$ the high-frequency branches of the spin-wave excitations refer to the waves of the forward type ($\partial\omega/\partial k_\perp > 0$), and the low-frequency branches correspond to the waves of the backward type ($\partial\omega/\partial k_\perp < 0$). If $\mathbf{n} \parallel 0Z$, then the opposite situation takes place: $\omega_+(k_\perp \longrightarrow 0) \longrightarrow \omega_d$, $\omega_+(k_\perp \longrightarrow \infty) \longrightarrow \omega_c$, $\omega_-(k_\perp \longrightarrow 0) \longrightarrow \omega_a$, and $\omega_-(k_\perp \longrightarrow \infty) \longrightarrow \omega_b$. An analysis of (33) and (34) shows that the presence of a nonmagnetic coating most strongly changes the spectrum of quasi-uniform (across the thickness of the magnetic film) bulk spin-wave modes (ω_{a-d}), which in the spectrum of the isolated film (18) were associated with modes with $\nu = 0$. Both branches of this mode of bulk exchangeless spin waves become dispersionless, and the type of the wave (forward or backward) is directly opposite to the type of the modes with greater numbers $\nu > 1$ and traveling in the same frequency range. Finally, at $\mathbf{n} \parallel 0Y$ (for ω_-) or at $\mathbf{n} \parallel 0Z$ (for ω_+) and $k_\perp \longrightarrow 0$, the range of existence of quasi-uniform (across the film thickness) anisotropic elastostatic spin-wave modes becomes dependent on the ratio of the thicknesses and elastic shear moduli of the magnetic and nonmagnetic layers.

It can be deduced from (33) and (34) that the indirect coupling via the long-range field of quasi-static magnetoelastic deformations in the presence of a nonmagnetic coating having a continuous acoustic contact with the surface of the magnetic film results in the formation of a qualitatively new type of propagating nonexchange spin-wave excitations—internal ESSWs. The spectrum of these ESSWs depends on the ratio of the thicknesses of the magnetic film and the nonmagnetic coating and in the case of (33) and (34) consists of one (Ω_i) or two ($\Omega_{i\pm}$) branches depending on whether the film carries a one- or two-sided nonmagnetic coating. At $0 < k_\perp < \infty$, the range of the existence of these internal ESSWs lies in the interval (ω_b, ω_c) in the case of both (33) (one branch) and (34) (two branches). Therefore, for $\vartheta = \pi/2$, $\varphi = 0$ at $0 < k_\perp < \infty$, the frequencies of the bulk and internal ESSWs with $\mathbf{k} \perp \mathbf{l} \parallel \mathbf{u}$ in the film of a tetragonal AFM are located in the case of both (33) and (34) as follows:

—for a two-sided nonmagnetic coating,

$$\omega_a < \omega_b < \Omega_{i-} < \Omega_{i+} < \omega_c < \omega_d, \qquad (35)$$

—for a one-sided nonmagnetic coating,

$$\omega_a < \omega_b < \Omega_i < \omega_c < \omega_d. \qquad (36)$$

Let us now consider the effect of a nonmagnetic coating on the conditions of the formation and dispersion properties of anisotropic bulk and surface exchangeless ESSWs in the phase with $\vartheta = \pi/2$ and $\varphi = \pi/4$. Using the set of dynamic equations (6) and boundary conditions (31) and (32), we may show that, at $\mathbf{k} \perp \mathbf{l} \parallel \mathbf{u}$, the indirect spin–spin coupling via the field of "elastostatic" phonons leads to the following dispersion equation for the propagating nonexchange anisotropic ESSWs ($\kappa^2 \equiv (\mu_\perp/\mu_\parallel)k_\perp^2$):

—for the one-sided nonmagnetic coating,

$$\mu_\parallel^2 \kappa^2 a^2 + \mu_\parallel k_\perp a\kappa\tanh(k_\perp t)\coth(\kappa d)$$
$$+ \mu_a s k_\perp^2 a \tanh(k_\perp t) - \mu_*^2 a^2 k_\perp^2 = 0, \qquad (37)$$

for the two-sided nonmagnetic coating,

$$\mu_\parallel^2 \kappa^2 a^2 + \mu_\parallel \kappa k_\perp a(\tanh(k_\perp t) + \tanh(k_\perp l))\coth(\kappa d)$$

$$+ \mu_a s a k_\perp^2 (\tanh(k_\perp t) - \tanh(k_\perp l)) \qquad (38)$$

$$+ k_\perp^2 \tanh(k_\perp t)\tanh(k_\perp l) - \mu_a^2 a^2 k_\perp^2 = 0,$$

where $\mu_\parallel$ and $\mu_\perp$ at $\mathbf{n} \parallel 0Z$ or $\mathbf{n} \parallel 0Y$ are determined by relations (24) and (25), and $s = 1$ or $s = -1$ correspond to the positive or negative values of the projection of the wave vector of the traveling spin wave $k_\perp$ on the film plane.

A qualitatively new feature of this spectrum of propagating anisotropic ESSWs as compared to (33) and (34) is the appearance of two branches of internal ESSWs ($\Omega_{i\pm}$). If we, as before, denote by ω_{a-d} the characteristic frequencies of the propagating bulk ESSWs defined by (37) and (38) at $k_\perp \longrightarrow 0$ and $k_\perp \longrightarrow \infty$, we can show that, at $0 < k_\perp < \infty$, the ranges of the existence of the internal and bulk anisotropic ESSWs obey the set of inequalities analogous to (26)

$$\omega_a < \omega_b < \Omega_{i-} < \Omega_{i+} < \omega_c < \omega_d. \qquad (39)$$

Important features of the considered spectrum of anisotropic surface and bulk ESSWs (37) and (38) that are associated with the presence of a nonmagnetic coating (one- or two-sided) are the lack of reciprocity with respect to the inversion of the direction of the wave propagation ($\omega(k_\perp) \neq \omega(-k_\perp)$) and the possibility of the formation of regions with a zero group velocity at $k_\perp \neq 0$.

If we now consider the case where, in the phase with $\vartheta = \pi/2$, $\varphi = \pi/4$, the normal to the film surface ($\mathbf{n} \perp \mathbf{l}$) makes an angle ψ which is not equal to zero or $\pi/2$ with the 0Z-axis, then, with allowance for (27)–(29), the dispersion equation that determines the spectrum of anisotropic ESSWs of a thin magnetic film can be written in
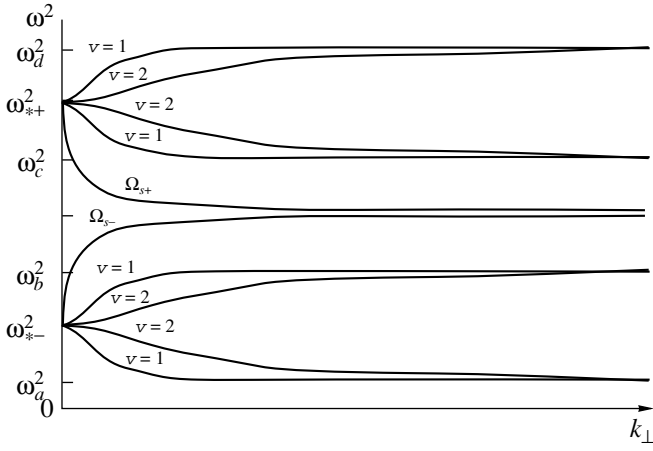
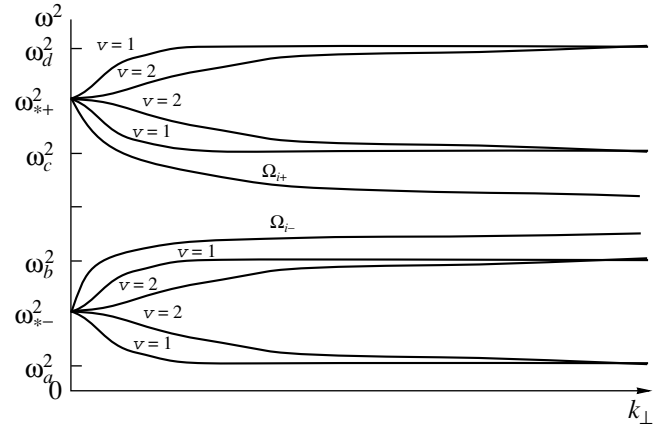**Fig. 3.** $0 < \psi < \pi/2;\ \vartheta = \pi/2;\ \varphi = 0,\ \pi/2;\ t = l = \infty.$

Fig. 4. $0 < \psi < \pi/2;\ \vartheta = \pi/2;\ \varphi = \pi/4,\ 3\pi/4;\ t = l = \infty.$

the following form ($\kappa^2 \equiv -(\mu_s^2 - \mu_\perp\mu_\parallel)k_\perp^2/\mu_\parallel^2$):

—for the one-sided nonmagnetic coating,

$$\mu_\parallel^2\kappa^2 a^2 + \mu_\parallel\kappa k_\perp a\tanh(k_\perp t)\coth(\kappa d)$$
$$+ \mu_* s k_\perp^2 a\tanh(k_\perp t) - \mu_*^2 a^2 k_\perp^2 = 0, \tag{40}$$

—for the two-sided nonmagnetic coating,

$$\mu_\parallel^2\kappa^2 a^2 + \mu_\parallel\kappa k_\perp a(\tanh(k_\perp t) + \tanh(k_\perp l))\coth(\kappa d)$$
$$+ \mu_* s k_\perp^2 a(\tanh(k_\perp t) - \tanh(k_\perp l)) \tag{41}$$
$$+ k_\perp^2 \tanh(k_\perp t)\tanh(k_\perp l) - \mu_*^2 a^2 k_\perp^2 = 0.$$

If, as before, the above-introduced designations for the condensation points of the spectrum of high-frequency (ω_+) and low-frequency (ω_-) bulk modes of ESSWs (ω_{a-d}) and for the mode of the spectrum of internal ESSWs ($\Omega_{i\pm}$) are valid, then, as follows from (40) and (41), the spectra of the corresponding exchangeless ESSWs satisfy the set of inequalities

$$\omega_a < \omega_{*-} < \omega_b < \Omega_{i-} < \Omega_{i+} < \omega_c < \omega_{*+} < \omega_d. \tag{42}$$

Here, at $\nu > 1$, the bulk forward-type ESSWs are formed at $\omega_{*+} < \omega < \omega_d$ and $\omega_{*-} < \omega < \omega_b$, whereas the bulk waves of the backward type are located in the intervals $\omega_a < \omega < \omega_{*-}$ and $\omega_c < \omega < \omega_{*+}$. In addition, at $k_\perp \equiv k_{\pm *}$, the dispersion curves of both internal ESSWs determined by equations (39)–(41) may smoothly pass into corresponding dispersion curves for the quasi-uniform (across the film thickness) modes of the anisotropic bulk ESSWs (Figs. 3, 4). For this to occur, it is necessary that the dispersion curves of the modes belonging to the spectrum of low-frequency elastostatic bulk waves (ω_-) were forward-type waves ($\partial\omega/\partial k_\perp > 0$) and the dispersion curves of the modes

belonging to the spectrum of high-frequency elastostatic bulk waves were of the backward-wave character ($\partial\omega/\partial k_\perp < 0$). The formation of the above-found modes of exchangeless internal elastostatic spin waves is impossible if the conditions $\mu_\parallel = 0$ and $a^2\mu_*^2 > 1$ (at $t = l = \infty$) are fulfilled simultaneously.

The results of the above-considered elastostatic approach to the analysis of spectra of spin-wave excitations in a thin magnetic film also remain important when taking into account the nonuniform exchange interaction.

EFFECT OF NONUNIFORM EXCHANGE COUPLING

The calculation of the dispersion equation that describes the elastoexchange spin dynamics of a thin magnetic layer (isolated or carrying a nonmagnetic coating) is intricate but represents no large difficulties. Naturally, its analysis requires the application of numerical methods, but on a qualitative level the structure of the spectrum of propagating elastoexchange spin waves can be studied by comparing (at the same values of $k_\perp$) the frequency ranges of existence of the above-found ESSWs (see (14), (22), (26), (30), (35), (36), and (42)) and of propagating exchange spin-wave excitations. It can be shown that, in the complete analogy with the exchange–dipole dynamics of thin magnetic films, a necessary condition for the formation of a nonuniform spin–spin resonance with the participation of traveling elastostatic and exchange modes is the allowance for the modes of the spectrum of exchange bulk spin waves. The interaction of these waves leads, in the range of degeneration of the spectra, to "elbowing" of the dispersion curves of the above modes and the formation in the resonance range of propagating elastoexchange spin waves. With increasing magnetic film thickness, the distances between neighboring exchange modes of the spectrum of bulk spin waves of

a magnetic film will decrease and, subject to condition (1), the above-found types of elastoexchange spin waves will gradually transform into resonance levels against the background of a continuous spectrum of bulk spin waves. In another limiting case, upon a decrease in the thickness of the magnetic film, the distance between the neighboring exchange modes will increase infinitely, which, at $d < d_*$, will make it impossible to realize the above nonuniform spin–spin resonance of the elastoexchange type in the film. If the range of existence of the above anisotropic surface ESSWs and internal ESSWs lies below that of the spectrum of the bulk spin modes, then, when solving the complete elastoexchange boundary problem, they transform into elastoexchange coupled spin-wave states. If the above-found anisotropic backward-type ESSWs can be formed below the range of the existence of exchange bulk modes of the thin magnetic film, then the sequential solution of the elastoexchange boundary problem will result in the formation of a minimum in the dispersion curve of the corresponding bulk elastoexchange spin mode propagating along the film.

CONCLUSION

Thus, the proper allowance for the magnetocrystalline anisotropy can substantially affect the character of the indirect spin-spin coupling via the field of elastostatic phonons in thin magnetic films, resulting in new types of exchangeless spin-wave excitations of anisotropic bulk, surface, and internal elastostatic spin waves that never have been studied previously.

It should be noted that the case of an antiferromagnet of tetragonal symmetry is by no means the only type suitable for the formation of the above anisotropic types of elastostatic spin waves. In particular, an analysis of thin magnetic films with an orthoferrite structure, subject to condition (1), shows that the above calculation of the spectrum of anisotropic ESSWs of a tetragonal antiferromagnet in the phase with $\vartheta = \pi/2$, $\varphi = 0$ remains valid in the case where the ground state of the uniformly magnetized film is one of the weakly ferromagnetic phases of the orthoferrite ($G_x F_z$ or $G_z F_x$). As to the spectrum of anisotropic ESSWs in the tetragonal antiferromagnet phase with $\vartheta = \pi/2$, $\varphi = \pi/4$, in the orthoferrite an analogous case is realized if the ground state of the orthoferrite film is a canted phase. For the experimental study of the above-found types of exchangeless spin-wave excitations, the condition $T_N < T_D$ between the Néel and Debye temperatures should be fulfilled in view of (1). Here, however, we should note that if a hyperfine interaction exists in the crystal ($A \neq 0$, where A is the hyperfine field constant), then, as was shown in [9], the results of analyzing the conditions of the formation of elastostatic spin waves in the electron subsystem will also be completely applicable to the spectrum of nuclear spin-wave excitations to the accuracy of the substitution of the frequency of spin waves ω by an effective frequency ω_* defined by the condition

$$\omega_*^2 = \frac{\omega^2 \omega_T^2}{\omega_n^2 - \omega^2} + \omega^2, \quad (43)$$

where ω_T is the dynamical shift of the nuclear magnetic resonance (NMR) frequency, and ω_n is the unshifted NMR frequency.

Since the NMR frequency is smaller by a factor of several orders of magnitude than the frequency of the uniform AFMR, the conditions for the exchangeless approximation can be fulfilled much more easily for nuclear anisotropic ESSWs even in those crystals in which the analogous approximation for electron ESSWs is unrealistic. With the above in mind, one of the possible objects of searching for anisotropic ESSWs discovered in this work may be a film (isolated or with a nonmagnetic coating) of a tetragonal antiferromagnet MnF_2 in the spin-flop phase, for which nuclear spin waves were experimentally studied previously [10].

REFERENCES

1. B. Schneider, Phys. Status Solidi B **51** (1), 325 (1972).
2. A. S. Beregov, Izv. Vyssh. Uchebn. Zaved., Radioélektron. **24** (27), 9 (1981).
3. O. L. Galkin and P. E. Zil'berman, Pis'ma Zh. Tekh. Fiz. **10** (17), 1077 (1984).
4. V. B. Afinogenov, P. E. Zil'berman, G. T. Kazakov, *et al.*, Pis'ma Zh. Tekh. Fiz. **12** (16), 996 (1986).
5. A. L. Sukstanskiĭ and S. V. Tarasenko, Zh. Éksp. Teor. Fiz. **105** (4), 928 (1994).
6. Yu. A. Sirotin and M. P. Shaskol'skaya, *Introduction to Crystal Physics* (Nauka, Moscow, 1979).
7. E. A. Turov, *Physical Properties of Magnetically Ordered Crystals* (Izd. Akad. Nauk SSSR, Moscow, 1963).
8. B. A. Ivanov and G. A. Oksyuk, *Nonlinear Dynamics of Antiferromagnets Exhibiting Weak Antiferromagnetism.* Preprint No. 84-8P, ITF (Institute of Theoretical Physics, Kiev, 1984).
9. E. P. Stefanovskiĭ and S. V. Tarasenko, Fiz. Nizk. Temp. **19** (7), 779 (1993).
10. A. R. King, V. Jaccarino, and S. M. Rezende, Phys. Rev. Lett. **37**, 533 (1976).

Translated by S. Gorin

================ **RADIOPHYSICS** ================

Theory of a Microwave-Pulse Compressor Based on a Barrel-Shaped Cavity with Helical-Corrugated Surface

Yu. Yu. Danilov, S. V. Kuzikov, and M. I. Petelin

Institute of Applied Physics, Russian Academy of Sciences, ul. Ul'yanova 46, Nizhniĭ Novgorod, 603600 Russia
Received August 19, 1998

Abstract—The radiation damping ratio and the transmission coefficient were calculated for a passive compressor of microwave pulses that is built around a length of an oversized waveguide with a helical-corrugated internal surface. © *2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

For passive compression of microwave pulses [1], a ring cavity designed on an axially symmetric metal structure (see figure) that consists of sections having different diameters and connected by tapered waveguide segments was proposed [2]. The cavity itself is a slightly expanding part of a tube. A wave passing through the system and an operating mode of the cavity are coupled by a helical-corrugated internal surface [3, 4]. In this work, we develop an approximate theory of such a compressor.

OPERATING MODES IN THE CAVITY AND THE WAVEGUIDE

Let the waveguide *H*-mode

$$\mathbf{E}_w = \mathrm{Re}\{C_j(z)\mathbf{E}_j\exp(i\omega t)\},$$

$$\mathbf{H}_w = \mathrm{Re}\{C_j(z)\mathbf{H}_j\exp(i\omega t)\},$$

$$\mathbf{E}_{\perp j} = -ik\left(\frac{R}{\mu_{mn}}\right)^2[\nabla_{\perp}H_{zj}, \mathbf{z}^0], \tag{1}$$

$$\mathbf{H}_{\perp j} = -ih_j\left(\frac{R}{\mu_{mn}}\right)^2\nabla_{\perp}H_{zj},$$

$$H_{zj} = J_{mj}\left(\frac{\mu_{mn}}{R}r\right)\exp(-im_j\varphi - ih_jz)$$

(where $k = \omega/c$, $h_j = \sqrt{k^2 - (\mu_{mn}/R)^2}$, μ_{mn} is the *n*th root of the derivative of an *m*th-order Bessel function, and $R(z)$ is the radius of the undisturbed cavity surface) and the cavity *E*-mode (optimal from the viewpoint of electric strength)

$$\mathbf{E}_r = \mathrm{Re}\{B_S\mathbf{E}_S\exp(i\omega t)\},$$

$$\mathbf{H}_r = \mathrm{Re}\{B_S\mathbf{H}_S\exp(i\omega t)\},$$

$$\mathbf{H}_{\perp s} = ik\left(\frac{R}{\nu_{mn}}\right)^2[\nabla_{\perp}E_{zs}\mathbf{z}^0], \tag{2}$$

$$E_{zs} = J_{ms}\left(\frac{\nu_{mn}}{R}r\right)\exp(im_s\varphi)F(z)$$

(where $h_s(z) = \sqrt{k^2 - (\nu_{mn}/R)^2}$ and ν_{mn} is the *n*th root of the *m*th-order Bessel function) be the operating modes.

We assume that the cavity length far exceeds the wavelength $\lambda = 2\pi c/\omega$, which allows us to neglect the radial and azimuth components of the electric field of the operating single-humped quasi-cutoff mode in comparison with the longitudinal component. For the latter, the *z* dependence is described by a slowly varying real function $F(z)$, satisfying the nonuniform-string equation [5, 6]

$$\frac{d^2F(z)}{dz} + h_s^2(z)F(z) = 0$$

and exponentially decaying outside the cavity.

COUPLING BETWEEN THE CAVITY AND WAVEGUIDE OPERATING MODES

The corrugation-induced disturbance is taken into account by introducing surface magnetic current $\mathbf{i}^m$ [3, 5]

$$\mathbf{i}^m = \frac{1}{4\pi}[\mathbf{n}, i\omega l[\mathbf{n}, \mathbf{H}] + c\nabla(E_n l)], \tag{3}$$

where $\mathbf{n}$ is the normal unit vector directed inward to the metal, $l(z) = b(z)\cos(\overline{m}\varphi + \overline{h}z)$ is the corrugation profile, and *b* is the corrugation depth.

According to formula (3), the magnetic current induced on the corrugated surface by the cavity operating mode (2) has a harmonic corresponding to the syn-
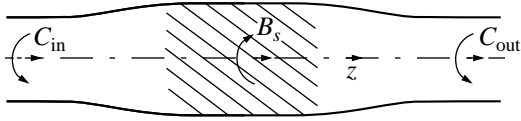
Model of the cavity.

chronous waveguide mode (1) under the condition (cf. [3])

$$m_j - m_s = \bar{m}, \quad h_j \approx \bar{h} = \frac{2\pi}{d}, \tag{4}$$

where $\bar{m}$ and d are the number of starts and the corrugation pitch, respectively.

The reradiation into spurious modes is suppressed by the utilization of modes with opposite directions of rotation (opposite signs of m_j and m_s) in the cavity and the waveguide.

CAVITY FREE OSCILLATIONS FOR RADIATION INTO THE WAVEGUIDE: AN ENERGY APPROACH

Find the complex amplitude of a waveguide mode excited by the cavity eigenmode field (2) from the equation [7]

$$\frac{dC_j(z)}{dz} = -\frac{1}{N_j}\oint \mathbf{i}^m \mathbf{H}_j^* d\sigma, \tag{5}$$

where $d\sigma = Rd\varphi$, $N_j = -c\,\mathrm{Re}\int [\mathbf{E}_j, \mathbf{H}_j^*]\,\mathbf{z}^0 dS_\perp/2\pi$ is the norm of the waveguide operating mode and $dS_\perp = rdrd\varphi$.

Substituting formulas (1) and (2) into (3) and then (3) into (5), neglecting the derivatives of the slowly varying functions ($\partial C_j(z)/\partial z$, $\partial R/\partial z$, and $\partial b/\partial z$) and the fast-oscillating terms, and taking into account condition (4), we obtain from equation (5), at $C_{in} = C_j(-\infty) = 0$, the amplitude $C_{out} = C_j(\infty)$ of the wave radiated from the cavity,

$$C_{out} = K_1 I_1 B_s,$$

where $K_1 = km_j\mu_{mn}^2 J_{ms}'(\nu_{mn})/2(\mu_{mn}^2 - m_j^2)\nu_{mn}RJ_{mj}(\mu_{mn})$ and $I_1 = \int_{-\infty}^{+\infty} b(z)F(z)dz$, and the power radiated from the cavity into the waveguide,

$$P_j = \frac{|C_{out}|^2 |N_j|}{4}. \tag{6}$$

From formula (6), we find the radiation damping ratio $\omega_s'' = P_j/2W_s$ (where $W_s = |B_s|^2 N_s/2$ is the mean energy stored in the cavity and determined by (2) and

$N_s = \int |\mathbf{E}_s|^2 dV/4\pi = \int |\mathbf{H}_s|^2 dV/4\pi$ is the norm of the cavity operating mode):

$$\omega_s'' = \frac{\omega M}{(\mu_{mn}/m_j)^2 - 1}, \tag{7}$$

where $M = h_j I_1^2/8R^2 I_r$ and $I_r = \int_{-\infty}^{+\infty} F^2(z)dz$. Hence, the radiation Q-factor is given by the cavity

$$Q_s = \frac{(\mu_{mn}/m_j)^2 - 1}{2M}.$$

Similar formulas are also obtained for the case when the operating mode of the cavity is the E-mode:

$$\omega_s'' = \frac{\omega k^2 M}{h_j^2}, \quad Q_s = \frac{1}{2M}.$$

CAVITY TRANSMISSION COEFFICIENT

The transmission of a lossless ring cavity is described by the universal formula [8]

$$T(\omega) = \frac{C_{out}}{C_{in}} = \frac{\omega - \omega_{s0} + i\omega_s''}{\omega - \omega_{s0} - i\omega_s''}, \tag{8}$$

where ω_{s0} is the undisturbed eigenfrequency of the cavity and ω_s'' is the radiation damping ratio, for which, in this case, formula (7) is valid.

We can arrive at the same result by solving the self-consistent set of equations for excitation of the waveguide [equation (5)] and the cavity:

$$B_s = -\frac{i}{2(\omega - \omega_{s0})N_s}\oint \mathbf{i}^m \mathbf{H}_s^* d\sigma. \tag{9}$$

Substituting (1) and (2) into (3) and then (3) into (5) and (9) and neglecting the derivatives of the slowly varying functions ($\partial C_j(z)/\partial z$, $\partial R/\partial z$, and $\partial b/\partial z$) and the fast-oscillating terms, we obtain, in view of (4), the set of equations for the amplitude $C_j(z)$ of a wave passing through the cavity and for the amplitude B_s of the cavity eigenmode:

$$C_j(z) = C_{in} + K_1\bar{I}_1 B_s, \tag{10a}$$

$$B_s = \frac{iK_2 I_2 C_j(z)}{\omega - \omega_{s0}}, \tag{10b}$$

where $C_{in} = C_j(-\infty)$ is the amplitude of the wave exciting the cavity, $K_2 = \omega m_j h_j \nu_{mn} J_{mj}(\mu_{mn})/4k \times \mu_{mn}^2 RJ_{ms}'(\nu_{mn})I_r$, $\bar{I}_1 = \int_{-\infty}^{z} b(z')F(z')dz'$, and $I_2 = \int_{-\infty}^{\infty} b(z)C_j(z)F(z)dz$.

Passing in formula (10a) to the amplitude of the transmitted wave, $C_{out} = C_j(\infty)$; substituting (10a) into

(10b); and integrating by parts, we obtain the set of equations for C_{out} and B_s:

$$C_{out} = C_{in} + K_1 I_1 B_s, \tag{11a}$$

$$(\omega - \omega_{s0})B_s = iK_2 I_1 \left(C_{in} + \frac{1}{2} K_1 I_1 B_s \right). \tag{11b}$$

From system (11), the universal transmission $T = C_{out}/C_{in}$ in the form of (8) can be derived.

CONCLUSION

Since the transmission of the device shown in the figure is described by relationship (8), which is applicable to any ring cavity, this device can be applied to compress microwave pulses with the standard method described in [1, 8].

ACKNOWLEDGMENTS

REFERENCES

1. P. B. Wilson, *Application of High-Power Microwaves*, Ed. by A. Gaponov-Grekhov and V. Granashtein (Artech House, Boston, 1994).

2. Yu. Yu. Danilov, S. V. Kuzikov, V. G. Pavelyev, *et al.*, in *On CD ROM Proc. of 7th Int. Workshop on Linear Colliders* (Zvenigorod, Russia, 1997).

3. N. F. Kovalev, I. M. Orlova, and M. I. Petelin, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **11** (5), 783 (1968).

4. A. L. Gol'denberg, G. S. Nusinovich, and A. B. Pavel'ev, *Gyrotrons* (Institute of Applied Physics, USSR Academy of Sciences, Gorky, 1980).

5. B. Z. Katsenelenbaum, *Theory of Irregular Waveguides with Slowly Varying Parameters* (USSR Academy of Sciences, Moscow, 1961).

6. S. N. Vlasov, G. M. Zhislin, I. M. Orlova, *et al.*, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **12** (8), 1236 (1969).

7. L. A. Vaĭnshteĭn, *Electromagnetic Waves (*Radio i Svyaz', Moscow, 1988).

8. M. I. Petelin and L. M. Tai, in *AIP Conference Proceedings* (1995), No. 337, p. 303.

Translated by A.S. Kondrat'ev

====================== **RADIOPHYSICS** ======================

Tunnel Josephson Junctions as Highly Sensitive Comparators in Stroboscopic Converters

I. N. Askerzade

Institute of Physics, Academy of Sciences of Azerbaijan, Baku, 370143 Azerbaijan
Received July 1, 1998; in final form, August 3, 1999

Abstract—The dynamics of switching to the resistive state of several types of comparators based on tunnel Josephson junctions is analyzed by the method of numerical modeling. The expediency of utilizing the scheme of the balance comparator in the mode of single-quantum pulses for attaining higher performance as the hysteresis of the voltage–current characteristics (VCCs) decreases is demonstrated. © 2000 MAIK "Nauka/Interperiodica".

INTRODUCTION

Superconducting stroboscopic converters are designed according to a conventional scheme [1], in which single Josephson junctions or interferometers created on their basis play the role of the Esaki diodes. The first stroboconverters were designed using two tunnel junctions, one of them acting as a generator of short pulses and the other as a comparator. Such a design of stroboconverters was proposed by Faris and allowed one to improve substantially the parameters of known semiconductor stroboconverters even in the first experiments [2].

Since the time resolution of stroboconverters with a comparator based on the single Josephson junction is determined through the strobopulse duration, further improvement of converters is connected with the increase of density j_c of the critical current. The best results were reported in [3]: at $j_c = 30$ kA/cm^2, the pulse duration was 3.7 ps, the time resolution was 2.1 ps, and the sensitivity $\delta I \approx 1$ mkA/Hz$^{1/2}$. In the opinion of the authors of papers [3] and [4], by further increasing the density of the current, one may attain a subpicosecond resolution. Therefore, the study of the dynamics of comparators employing tunnel junctions with high density j_c of the current is a topical problem.

In order to enhance the noise immunity of stroboconverters, a balance scheme was proposed, in which two Josephson junctions are coupled using the principle of the Gauto pair [5] and the comparator is strobed by the voltage overfall. It is also of interest to investigate the balance property of the Faris scheme and to compare the characteristics of both schemes.

In this paper, I present the results of numerical modeling of the dynamics of processes for several types of comparators based on the Josephson junctions in a wide range of variation of their parameters. I also compare their performance as applied to the possibility of designing the Josephson converters with a record time resolution and sensitivity.

SCHEME OF THE FARIS COMPARATOR AND CALCULATION TECHNIQUE

The scheme of the comparator with the Faris generator is shown in Fig. 1. Josephson junction J_4 is connected to a two-contact interferometer with junctions J_1 and J_2 via resistor R and forming unit J_3. The interferometer is switched by choosing feeding current I_e and control current I_{cont}. A sharp transition, which is obtained as a result of switching the interferometer to the resistive state, is enhanced by small junction J_3. A short strobopulse (a short pulse of current) is formed and sent to comparator J_4, where it is added to signal current I_s and feedback current I_{fb}.

In order to investigate the dynamics, the performance of this system was simulated numerically using the software Personal Superconductive Circuit Analyzer (PSCAN) developed at the laboratory of cryoelectronics of the Department of Physics of the Moscow State University [6]. PSCAN is intended for the analysis of circuits with Josephson junctions. The base subroutine of PSCAN, which simulates the system behavior with a variable time step, calculates, within the
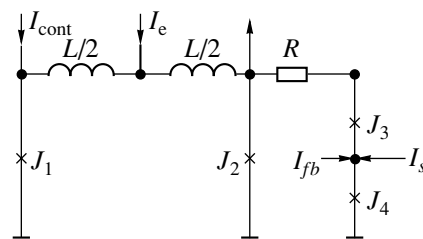
Fig. 1. The comparator based on tunnel Josephson junction J_4 connected to a two-contact interferometer (J_1, J_2) via resistor R and forming unit J_3.

framework of the chosen model, Josephson phases φ at all nodes of the scheme; the time step is chosen so that the maximum phase increment at every point of the scheme would not exceed $\pi/100$. All calculations are performed using the microscopic Wertchamer model, according to which each junction is characterized by five parameters [6]: critical current I_0; slot voltage V_g; the characteristic voltage $V_c = I_c R_N$, where R_N is the normal resistance of the junction; and the capacity parameter $\beta = 2\pi I_c R_N^2 C/\Phi_0$, where C is the junction capacity. The junction is also characterized by Q-factor κ of the tunnel junction, which is defined as the ratio of the quasiparticle resistance in the preslot domain to the normal resistance of the Josephson junction, $\kappa = R_j/R_N$.

NUMERICAL RESULTS

A typical strobing pulse consists of two main parts: the triangular body of the pulse and the tail in the form of the Josephson oscillations that follows the body. The best parameter that characterizes such a shape of the pulse is the ratio of the difference between the amplitude of the triangular part and the oscillation amplitude to the amplitude of the body, $\delta a = \delta A/A$. The dynamics of the two-contact interferometer is governed by the values of feeding current I_e and capacity parameter β; therefore, it is convenient to perform the analysis in the (β, i_e) plane, where $i_e = I_e/I_c$. The results of calculations presented in Fig. 2 by solid lines show that, for typical values of parameters ($\kappa = 9$, $R \sim R_{N1,2} = 1$, and $l = 3$, where l is the geometric inductivity of the interferometer in units $\Phi_0/2\pi I_c$), decreasing β deteriorates δa and the working range with respect to current i_e narrows down. For small β ($3 < \beta < 8$), the left boundary of the working domain in the (β, i_e) plane can be extended by setting $R = 2$ (dashed curves in Fig. 2). For other values of $\kappa = 3$, an abrupt shift of the boundary to the right is observed, which means that the quality of tunnel junctions exerts a pronounced effect on the system dynamics. Figure 3 shows the lines of equal front in the (β, i_e) plane.

Geometric inductivity l of the two-contact interferometer is one the most important parameters that influence the dynamics. It is known [7] that, when the voltage $\bar{V} = (\hbar/2e)\omega_0$, a narrow almost vertical peak appears on the voltage–current characteristic (VCC) of the two-contact interferometer, which is connected with intense forced oscillations with the frequency $\omega_0 = (C_1 + C_2)/LC_1C_2$ in the system formed by the inductance of the interferometer and capacitance of junctions. The height of this peak depends on Φ_e and may reach $0.6I_{c+}$, where I_{c+} is the critical current of the two-contact interferometer. The amplitude of oscillations excited in the circuit when the interferometer is switched to the resistive state depends on the ratio of the feeding current to the peak amplitude and on the speed of passing over the width of this peak $\varepsilon = d\omega/dt$.
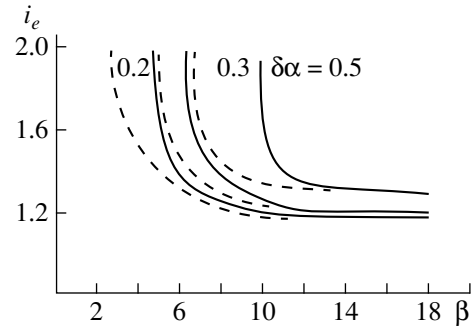

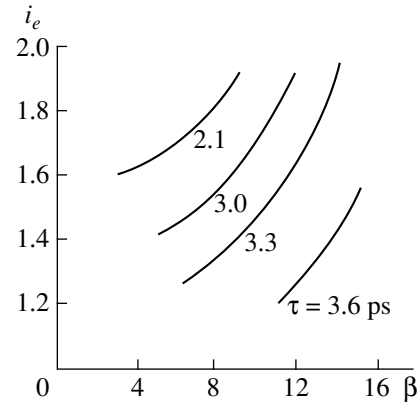**Fig. 2.** Variation of parameter δa of the strobing pulse in the (β, i_e) plane.

Fig. 3. Lines of equal duration of the strobopulse front in the (β, i_e) plane.

According to the theory of oscillations, the swing of the amplitude in the contour is determined, in the linear approximation, by the formula [7]

$$\frac{\Delta B}{B} \simeq \frac{\omega_0^2}{Q^2}\frac{1}{\varepsilon}, \tag{1}$$

where ω_0 is the eigenfrequency of the oscillatory circuit and Q is its Q-factor.

Substituting into (1) the Q-factor of the system recalculated taking into account the Josephson inductance and using the expression for ε, we obtain the relationship

$$\frac{\Delta B}{B} \simeq \frac{1}{\kappa^2\beta(l+1)} \tag{2}$$

for a symmetric interferometer. As follows from this expression, at small κ, β, and l, intense eigenoscillations can be easily excited in the contour, which hampers useful dynamics of the process. Since the values of parameters κ and β are governed by the technology of manufacturing the junctions, it seems convenient to decrease the amplitude of the excited oscillations by increasing inductivity l. However, the results of calculations show that the strobopulse duration increases
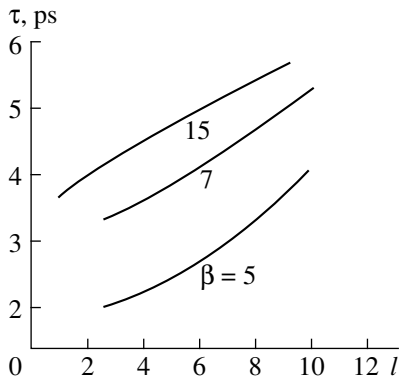
Fig. 4. The strobopulse duration vs. inductivity l of the interferometer.

together with l (Fig. 4). Therefore, I think that the values of l chosen in [4] are optimal.

Resistance R also exerts a pronounced effect on the shape of the obtained strobopulse. The value of the resistance is chosen using the following considerations: since the rate of increase of the current through a small junction is determined by the resistance, this rate should provide the fulfillment of the inequality

$$\tau_0 + \tau_D \leq \tau_N, \qquad (3)$$

where τ_0 is the time during which critical current I_{cM} of the junction is attained, $\tau_0 = I_{cM}/\dot{I}$; τ_D is the time of delay in the case of a linear increase of the current through the small junction [7]; and τ_N is the time of switching to the resistive state (duration of the front of the rectangular voltage pulse when the interferometer is switched), $\tau_N = R_N C$.

Explain the physical sense of inequality (3): during the time of existence of the front of the rectangular voltage, the formation of the front and top of the strobopulse must terminate; i.e., the current through the small junction must reach the critical value, and the delay interval must pass, after which the voltage on the junction attains a noticeable level. Taking into account that $\dot{I} = V_g/RR_N C$ and introducing dimensionless quantities $r = R/R_N$, $\xi = I_{cM}/I_c$, and $v_g = V_g/V_c$, rewrite relationship (3) to obtain

$$r \leq \left(1 - \frac{\omega_c \tau_D}{\beta}\right)\frac{v_g}{\xi}, \qquad (4)$$

where $\omega_c \tau_D$ is determined as

$$\omega_c \tau_D = \begin{cases} 4.64\sqrt{\beta}\left(\xi\frac{\sqrt{\beta}r}{v_g}\right)^{1/5} & \text{for } \beta \gg 1, \\ 2.9\xi(\beta r/v_g)^{1/3} & \text{for } \beta \lesssim 1. \end{cases}$$

The analysis of these formulas shows that in order to provide the validity of (4) for decreasing β, it is neces-

sary to increase r. This statement is confirmed by the results of calculations obtained for medium attenuation ($\beta \sim 3$, Fig. 2).

The calculations have demonstrated that when $\beta < 3$, parameter δa characterizing the shape of the pulse deteriorates ($\delta a \leq 0.1$) and the working range with respect to the current strongly decreases ($\Delta I/I_{c+} \sim 0.1$). One should not use such strobopulses, because rapidly increasing signal I_s, when added to relatively strong oscillations of the tail, may switch the comparator to the resistive state, which may cause errors in measurements of the current value of the signal. Thus, one may conclude that it is inexpedient to employ the Faris generator at small β ($\beta < 3$).

SCHEME WITH SINGLE-QUANTUM PULSES

Another version of utilizing tunnel junctions with high densities of the current for generating short pulses is connected with peculiarities of the dynamics of a two-contact interferometer when it is switched from the main superconducting state to the neighboring superconducting state. Taking I_e to be lower than threshold value I_Q [7] and choosing the appropriate I_{cons}, one can obtain a single-quantum pulse with the area

$$\int V dt = \Phi_0 = \frac{\hbar}{2e} = 2 \text{ mV ps}, \qquad (5)$$

amplitude $\sim 2V_c$, and duration $\sim \pi/\omega_c$. Current I_{const} is directed so as to provide the passage of the quantum through the right junction. In this version, it is not necessary to use a small junction, which simplifies the scheme. One should choose the resistance according to the following condition: the inverse current generated after switching the comparator to the resistive state must be less than the critical current of the right junction of the interferometer,

$$r \geq \frac{v_g}{1 - i_e/2}, \quad i_e < i_Q. \qquad (6)$$

The necessity of employing an external clocked feeding source, which complicates the scheme even more, constitutes the main disadvantage of the generators of short pulses described above. A resistive single-quantum scheme proposed in [8] is free of this drawback; a similar scheme was used in [9] as a strobopulse generator. Another important direction of improving the parameters of stroboconverters with tunnel junctions is the enhancement of their immunity to different kinds of noise and temperature fluctuations. With this purpose, a balance scheme was proposed, in which the Josephson junctions are coupled using the principle of the Gauto pair [5, 9]. Such a design enables one to exclude distortions connected with nonideality of the strobing signal. The time resolution of a stroboconverter is determined by the duration of the front of the input voltage overfall. The dynamics of the processes in one of the schemes employing an internal clock gener-

ator (a relaxator with the Josephson junction) and a balance comparator with tunnel junctions were investigated in [10].

The scheme of the Faris generator is universal in the following sense: replacing small junction *3* (Fig. 1) with the junction which is equivalent to the comparator (or vice versa), one obtains a balance scheme. The dynamics of the balance comparator is governed by the value of signal current I_s and the rate of increase of the current through junctions, which is determined by the front of the input strobing voltage overfall $\dot{I}$. It is important to construct the switching characteristic of the balance comparator in the ($\dot{I}$, I_s) plane, i.e., to determine the boundary, above which only one junction is switched. Calculations show that I_s is a linear function of $1/R$. Thus, as the rate $\dot{I} = V_g/RR_NC \sim 1/R$ increases, the signal current should be increased in order to satisfy the following condition: during the time of switching the first junction, the current through the second junction must not reach the critical value, i.e.,

$$I_s/\dot{I} \gtrsim \tau_N; \quad i_s \gtrsim v_g/r. \tag{7}$$

It is interesting to note that, according to the results of calculations, formula (4) obtained for resistance R remains valid, which means that the system dynamics does not change when the Faris scheme is transformed to a balance scheme.

Thus, one may state that the scheme of the generator shown in Fig. 1 is universal; namely, the Faris generator, the generator of single-quantum pulses, and the balance comparator constitute particular cases of this general scheme.

CONCLUSION

I have performed numerical modeling of several types of comparators based on tunnel Josephson junctions. I have determined the ranges of useful work of comparators versus the scheme parameters. I have demonstrated that, when density j_c of the critical current reaches large values corresponding to $\beta < 3$, utilization of the Faris scheme aimed at creation of a highly sensitive and fast stroboconverter becomes inexpedient, because the strobopulse parameters deteriorate. A transition to single-quantum schemes and the use of the balance principle for designing comparators provide a better time resolution [8, 9] and high sensitivity.

REFERENCES

1. É. H. Hermanis and V. G. Karklin'sh, *Discrete Stroboscopic Converters* (Zinatne, Riga, 1977).
2. S. M. Faris, Appl. Phys. Lett. **36**, 1005 (1980).
3. P. Wolf, B. J. van Zeghboroeck, and U. Deutsch, IEEE Trans. Magn. **23**, 226 (1985).
4. S. R. Whiteley, G. K. G. Hohenwarter, and S. M. Faris, IEEE Trans. Magn. **23**, 899 (1985).
5. P. A. Bakhtin, P. E. Kandyba, V. I. Makhov, *et al.*, Patent No. 1008908 (1983).
6. S. V. Polonsky, V. K. Semenov, and K. K. Likharev, in *Proceedings of ISEC-91* (Glasgow, Scotland, 1991), p. 160.
7. K. K. Likharev, *Introduction to the Dynamics of Josephson Junctions* (Moscow, 1985).
8. O. A. Mukhanov, V. K. Semenov, and K. K. Likharev, IEEE Trans. Magn. **23**, 759 (1985).
9. V. K. Kornev and V. K. Semenov, in *Proceedings of ISEC-87* (Tokyo, Japan, 1987), p. 131.
10. I. N. Askerzade and V. K. Kornev, Radiotekh. Élektron. **39**, 869 (1994).

=== **ELECTRON AND ION BEAMS, ACCELERATORS** ===

One Class of Electrostatic Fields Perfectly Conserving the Parallelism of Planar Homogeneous Beams of Charged Particles

L. G. Glikman and Yu. V. Goloskokov

Institute of Nuclear Physics, National Nuclear Center, Almaty, 480082 Kazakhstan

Received June 16, 1997; in final form, October 19, 1998

Abstract—Electrostatic systems whose field is a superposition of two two-dimensional fields with the plane of symmetry (midplane) in common are considered. It is assumed that these fields overlap in the region through which a charged particle beam passes. The basic property of these systems is that they perfectly (without angular aberrations) conserve the parallelism of charged particle beams that are homogeneous in terms of energy-to-charge ratio and pass in the midplane of the field. A four-electrode system where each of the electrodes consists of two plates symmetric with respect to the midplane is an example of the new-class electrostatic systems. © 2000 MAIK "Nauka/Interperiodica".

INTRODUCTION

It is known that plane-parallel homogeneous (in terms of energy-to-charge and mass-to-charge ratios) beams of charged particles perfectly (without aberration) retain their parallelism after passing through two-dimensional or conic static electromagnetic fields (see, for example, [1, 2]). In these cases, the particles move in the midplane, which is the plane of symmetry (antisymmetry) for the electric (magnetic) fields. The two-dimensional fields are described by scalar potentials independent of one of the Cartesian coordinates, and the conic fields, by potentials that depend only on the angular variables in the spherical coordinate system. We demonstrated [3] that there is one more class of electrostatic fields that conserve the parallelism of planar charged-particle beams homogeneous in terms of energy-to-charge ratio. In what follows, the beams homogeneous in terms of energy-to-charge ratio will be referred to as homogeneous for brevity.

In this paper, we analyze in more detail the electronic and optical properties of fields of the new class. In particular, it is shown that the four-electrode electrostatic system described in [3] can effectively be used as a small-size electrostatic prism, a large-deflection-angle mirror, and a lens with the straight optical axis that perfectly conserves the parallelism of a wide planar beam of charged particles with a large spread in energy.

ELECTRONIC AND OPTICAL PROPERTIES OF A SUPERPOSITION OF TWO TWO-DIMENSIONAL FIELDS WITH THE MIDPLANE IN COMMON

Let the mutual midplane of two two-dimensional fields coincide with the plane $z = 0$ of a Cartesian coordinate system x, y, z and these fields be described by potentials $\varphi_1(x, z)$ and $\varphi_2(y, z)$, respectively. In the nonrelativistic approximation, the Hamilton–Jacobi equation for the motion of a charged particle in the midplane of the superposition of these fields that is described by the scalar potential $\varphi = \varphi_1 + \varphi_2$ has the form

$$\frac{1}{2m}\left[\left(\frac{\partial S_0}{\partial x}\right)^2 + \left(\frac{\partial S_0}{\partial y}\right)^2\right] + e\Phi_1(x) + e\Phi_2(y) = E. \quad (1)$$

Here e, m, and S_0 are, respectively, the charge, mass, and constant total energy of the particle; $\varphi_1(x, 0) \equiv \Phi_1(x)$; $\varphi_2(y, 0) \equiv \Phi_2(y)$; and $S = -Et + S_0$ is the truncated action function, where S is the action function and t is time.

The variables in equation (1) are separable. Substituting S_0 in the form of the sum $S_0(x, y) = S_1(x) + S_2(y)$ into equation (1), we obtain

$$\frac{1}{2m}\left(\frac{dS_1}{dx}\right)^2 + e\Phi_1(x) - E$$
$$= -\frac{1}{2m}\left(\frac{dS_2}{dy}\right)^2 - e\Phi_2(y) = \lambda, \quad (2)$$

where λ is an arbitrary constant.

From equalities (2), we determine the momenta $P_x = m\dot{x} = dS_1/dx$ and $P_y = m\dot{y} = dS_2/dy$ and then the total integral of the Hamilton–Jacobi equation. Whereupon, the equation for particle trajectory is found in quadra-

tures by the well-known method,

$$\int_{x_0}^{x} \frac{(\mathrm{sgn}\dot{x}(x))dx}{\sqrt{W_0 \sin^2\theta_0 + e(\Phi_{10} - \Phi_1)}}$$

$$- \int_{y_0}^{y} \frac{(\mathrm{sgn}\dot{y}(y))dy}{\sqrt{W_0 \cos^2\theta_0 + e(\Phi_{20} - \Phi_2)}} = 0, \quad (3)$$

and the relationship between the coordinate x and time t is obtained:

$$t - t_0 = \sqrt{\frac{m}{2}} \int_{x_0}^{x} \frac{(\mathrm{sgn}\dot{x}(x))dx}{\sqrt{W_0 \sin^2\theta_0 + e(\Phi_{10} - \Phi_1)}}. \quad (4)$$

Here, W is the kinetic energy of the particle, θ is the angle between the particle velocity and the y-axis, and the subscript 0 denotes the initial values of the variables. The subscript $i0$ means that the ith potential ($i = 1, 2$) is calculated at the initial point of the trajectory. The angle θ at any point of the trajectory of a particle moving in the midplane can be determined from equalities (1)–(3):

$$\sin^2\theta = \frac{W_0 \sin^2\theta_0 + e(\Phi_{10} - \Phi_1)}{W_0 + e(\varphi_0 - \Phi_1 - \Phi_2)}. \quad (5)$$

It follows from this equality that a homogeneous parallel beam of charged particles moving in the midplane perfectly (without angular aberrations) retains its parallelism after passing through the field. This is one of the basic properties of the electron–optical systems under consideration. The required electronic and optical properties of this field in the direction perpendicular to the midplane can be ensured, as in the case of the two-dimensional field, by properly choosing the electrode potentials. In particular, we can choose conditions under which a three-dimensional homogeneous beam retains its parallelism (telescopy conditions). The parameters characterizing the properties of such a system in the direction perpendicular to the midplane are calculated from general formulas for midplane electron–optical systems [4, 5].

Comparing the right-hand side of equality (5) with the expression for $\sin^2\theta$ that was derived for the known electrostatic system consisting of two electron–optical elements whose two-dimensional fields do not overlap in the region where a charged particle beam passes [1, 6], we easily find that these expressions coincide. Therefore, the angular spread in energy, as well as the angular and linear magnifications in the midplane, is also the same in these systems.

Below is a simple example of the proposed system.

ELECTRON–OPTICAL SYSTEM WITH THE FIELD IN THE FORM OF A SUPERPOSITION OF TWO TWO-DIMENSIONAL ELECTROSTATIC FIELDS

We consider a four-electrode system whose field-generating surfaces are parallel to the midplane and $d/2$ distant from it (Fig. 1). Each of the electrodes consists of two equicharged plates symmetric with respect to the midplane. It is assumed that the interelectrode gaps are very small and can be neglected in an analysis of the electron–optical properties. In the cases of practical interest, this gap is usually $\approx 0.1d$. When projected onto the midplane, the plates of the first, second, third, and fourth electrodes occupy the quadrants $x < 0$, $y < 0$; $x > 0$, $y < 0$; $x > 0$, $y > 0$; and $x < 0$, $y > 0$, respectively. The electrode potentials are denoted as V_1, V_2, V_3, and V_4, respectively. It is easy to conclude that the potential φ for two two-electrode systems with the two-dimensional field can then be represented in the form of the sum of two potentials $\varphi_1(x, z)$ and $\varphi_2(y, z)$ if condition

$$V_3 - V_2 = V_4 - V_1 \quad (6)$$

is fulfilled.

The potential distribution in the first two-electrode system is found by solving the Dirichlet problem for the two-dimensional Laplace equation with the boundary conditions $\varphi_1 = V_1$ at $x < 0$ and $\varphi_1 = V_2$ at $x > 0$ on the planes $z = \pm d/2$. For the second two-electrode system, the boundary conditions have the form $z = \pm d/2$ $\varphi_2 = 0$ at $y < 0$ and $\varphi_2 = V_3 - V_2$ at $y > 0$ on the planes $z = d/2$. For each of the two-electrode systems, the potential distributions in terms of elementary functions are well known (see, for example, [5, 6]). The potential distribution in our four-electrode system can be represented in the form

$$\varphi = \varphi_1 + \varphi_2 = \frac{V_1 + V_3}{2} + \frac{V_2 - V_1}{\pi}\arctan\frac{\sinh(\pi x/d)}{\cos(\pi z/d)}$$

$$+ \frac{V_3 - V_2}{\pi}\arctan\frac{\sinh(\pi y/d)}{\cos(\pi z/d)}. \quad (7)$$

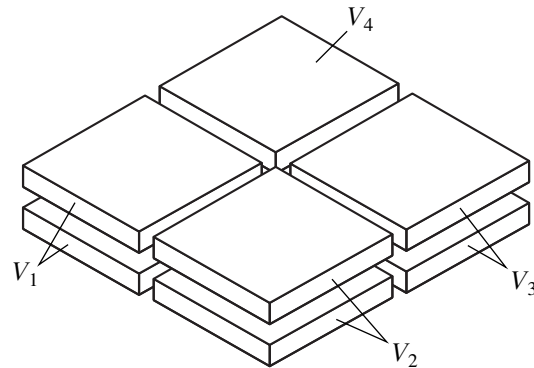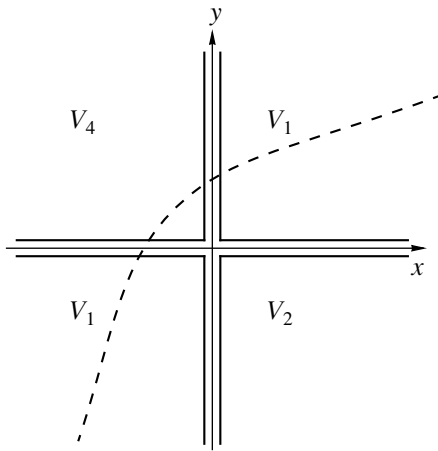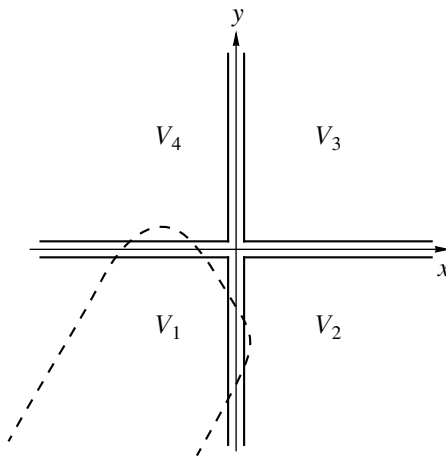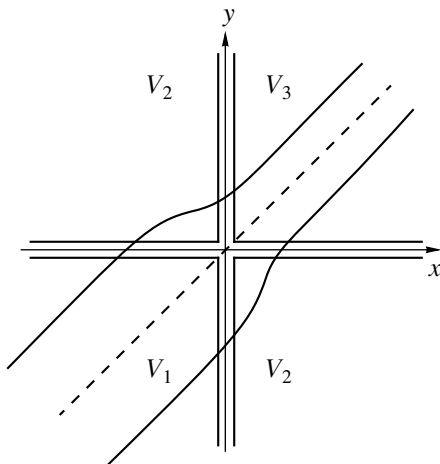The function $\varphi - (V_1 + V_3)/2$ is even with respect to

Fig. 1. Four-electrode electrostatic system.

Fig. 2. Axial beam trajectory in the four-electrode electrostatic system used as a prism.

Fig. 3. Axial beam trajectory in the four-electrode electrostatic system used as a mirror.

Fig. 4. Beam trajectories in the four-electrode electrostatic system used as a direct-axis lens.

the variable z and odd with respect to x and y. Its values on the plates of the first, third, second, and fourth electrodes are equal to $(V_1 - V_3)/2$, $(V_3 - V_1)/2$, $(V_2 - V_4)/2$, and $(V_4 - V_2)/2$, respectively.

The two-dimensional field of either system rapidly decays outside the interelectrode gaps. For example, if it is required that the field strength of the four-electrode system not exceed 0.01% of its maximum at the boundary of the field region, the field of the system in the midplane can be considered to be located in the cross-shaped region

$$\left| \frac{x}{d} \right| \le 3, \quad \left| \frac{y}{d} \right| \le 3.$$

Outside this region, the particle trajectories can be viewed as rectilinear for most applications. The beam refraction at the entrance to and the exit from the field can be disregarded. As for two-dimensional fields generated by electrodes separated by straight gaps (see, for example, [1, 6]), one can easily select the shapes and sizes of the plates such that the field in the region through which a beam passes coincides with the calculated value. The problem of beam introduction into and extraction from the system is readily solved if the potentials of the electrodes upstream and downstream of the beam coincide with those of the object and image spaces, respectively. The design solution to this problem depends on what objective is to be pursued by an electron–optical system. If the system is used, for example, as a prism in a dihedral electrostatic prismatic energy analyzer, the potentials of the object and image spaces coincide with the potentials of the electrodes of the collimating and focusing lenses adjacent to the prism. Similar design solutions are found elsewhere [1, 6].

The dashed lines in Figs. 2–4 are the axial trajectories of charged particles in the midplane of the proposed four-electrode system when it is used as a prism with equal potentials of the object and image spaces (Fig. 2), a mirror (Fig. 3), and a lens (Fig. 4). Figure 4 displays also two adjacent trajectories (solid lines).

In the prism (Fig. 2), two two-dimensional fields are superimposed in the region through which a beam passes. For this reason, such a prism can be considerably smaller than the known prism with nonoverlapping two-dimensional fields [1, 6]. At the same time, the basic electron–optical property of the former prism remains: when deflected, a homogeneous charged-particle beam moving in the midplane of prism's field totally retains its parallelism.

In the mirror (Fig. 3), the deflection angle is equal to 180° at any θ_0. Such a mirror can be made also with the use of two nonoverlapping two-dimensional fields. However, the superposition of the fields again enables us to reduce the mirror sizes while conserving the parallelism of a reflected planar homogeneous beam.

In the specific case when $V_2 = V_4 = (V_1 + V_3)/2$, there is one more plane of field symmetry, $x = y$, which is per-

pendicular to the midplane. In this case (Fig. 4), the system can be used as a lens with the direct axis, which is the line of intersection of the planes of symmetry; i.e., it intersects the y axis at an angle of 45°. In the midplane, this lens is a telescopic system without intermediate focus. The angular magnification of the lens, as follows from equality (5), is equal to W_0/W_b, where W_b is the kinetic energy of the particles leaving the field of the lens. It follows from the Lagrange–Helmholtz formula that the linear magnification of the lens is equal to $\sqrt{W_b/W_0}$. A parallel beam of charged particles moving in the midplane and entering the field of the lens at an angle of 45° to the y axis remains plane-parallel regardless of its width; i.e., it leaves the field of the lens at the same angle to the y axis. The telescopy condition in the midplane is fulfilled for any W_b/W_0 value, but the linear and angular magnifications depend on this ratio. Lenses with such properties cannot be implemented in a system with nonoverlapping two-dimensional fields.

CONCLUSION

In this study, we covered electron–optical systems on which various devices with high focusing properties can be built. Among them are mass and energy analyzers, systems for charged particle beam transport, etc. Charged-particle spectrometers are expected to have high resolution and sensitivity if the parallelism-conserving systems are used in combination with low-aberration focusing elements, such as transaxial electrostatic lenses [6].

REFERENCES

1. V. M. Kel'man and S. Ya. Yavor, *Electron Optics* [in Russian] (Nauka, Leningrad, 1968).
2. L. G. Glikman, Zh. Tekh. Fiz. **54,** 1986 (1984) [Sov. Phys. Tech. Phys. **29,** 1166 (1984)].
3. L. G. Glikman and Yu. V. Goloskokov, Pis'ma Zh. Tekh. Fiz. **24** (19), 57 (1998) [Tech. Phys. Lett. **24,** 772 (1998)].
4. S. P. Karetskaya and L. V. Fedulina, Zh. Tekh. Fiz. **52,** 740 (1982) [Sov. Phys. Tech. Phys. **27,** 469 (1982)].
5. S. P. Karetskaya, L. G. Glikman, L. G. Beizina, *et al.*, Adv. Electron. Electron Phys. **89,** 391 (1994).
6. V. M. Kel'man, S. P. Karetskaya, L. V. Fedulina, *et al.*, *Electron–Optical Elements of Prismatic Charged Particle Spectrometers* [in Russian] (Nauka, Alma-Ata, 1979).

Translated by R.T. Tyapaev

ELECTRON AND ION BEAMS, ACCELERATORS

Generalized Equations for the Dynamics of the Resistive Firehose Instability of an REB with Time-Dependent Radius and Current

A. S. Manuĭlov

*Smirnov Research Institute of Mathematics and Mechanics, St. Petersburg State University,
St. Petersburg, 198904 Russia*

Received August 10, 1998

Abstract—Generalized equations are derived that describe the linear stage of the resistive firehose instability of a relativistic electron beam whose radius and current change along the pulse. Such factors as reverse current, the perturbations of the plasma channel, and the evolution of the plasma conductivity due to impact ionization, avalanche ionization, and recombination are taken into account. © *2000 MAIK "Nauka/Interperiodica"*.

1. New areas for applications of relativistic electron beams (REBs) make further research on the dynamics of electron-beam transport in gas–plasma media worthwhile [1–13]. Among the problems related to the guiding of REBs, of special interest is the problem of investigating large-scale instabilities of the beams, the most dangerous of which is the resistive firehose instability (RFI). This is the mode with azimuthal wavenumber $m = 1$, which manifests itself as growing bending oscillations of a beam [1–8].

Here, we generalize the linear theory of the RFI of an REB, which was developed previously in [2, 3], to the case of a beam with time-dependent radius and current. We also assume that the radial profiles of the plasma current, the beam current, and the current flowing in the conducting channel are all arbitrary and take into account the effect of the bending oscillations of an REB on the conductivity of the background plasma.

2. We consider a paraxial axisymmetric monoenergetic REB propagating in a gas–plasma medium with scalar conductivity $\sigma_{ch0}(\xi)$ along the z-axis of a cylindrical coordinate system (r, θ, z), where $\xi = v_z t - z$ is the distance from the beam front to the beam cross section of interest, v_z is the z-component of the beam electron velocity, and t is the time.

We restrict ourselves to treating highly conducting background media such that the main part of an REB is completely charge-neutralized, $4\pi\sigma_{ch0}R_b/c \gg 1$, where R_b is the characteristic beam radius and c is the speed of light. We also assume that the equilibrium current densities of the beam and of the background plasma, $J_{b0}(r)$ and $J_{p0}(r)$, and the ohmic conductivity of the plasma, $\sigma_{ch0}(r)$, are all smooth functions of r.

3. We use the traditional model of a "rigid" beam, which implies that the transverse displacement of an REB causes no deformation of the radial profile of the beam current density [2, 3]. However, unlike papers [2, 3], we assume that the plasma conductivity σ_{ch} changes due to the transverse displacement of the beam and the processes of impact ionization, avalanche ionization, and recombination [10]. We also assume that the net beam current $I_b(\xi)$ and the characteristic beam radius $R_b(\xi)$ both depend on time. We switch from the two independent variables (t, z) to (ξ, z) to obtain the following equations, describing the dynamics of the RFI of a paraxial quasisteady REB in the linear stage:

$$\frac{\partial^2 Y}{\partial z^2} = \frac{c\pi}{I_b^2}\left(\frac{I_b}{I_A}\right)\int_0^\infty dr\, r\left[J_{b1}\frac{\partial A_{z0}}{\partial r} + \frac{J_{b0}}{r}\frac{\partial}{\partial r}(rA_{z1})\right], \quad (1)$$

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial A_{z0}}{\partial r}\right) = -\frac{4\pi}{c}(J_{b0} - J_{p0}), \quad (2)$$

$$\frac{\partial}{\partial r}\left[\frac{1}{r}\frac{\partial}{\partial r}(rA_{z1})\right] - \frac{4\pi}{c}\sigma_{ch0}\frac{\partial A_{z1}}{\partial \xi}$$
$$-\frac{4\pi}{c}\sigma_{ch1}\frac{\partial A_{z0}}{\partial \xi} = -\frac{4\pi}{c}J_{b1}. \quad (3)$$

Here, Y is the amplitude of the transverse oscillations of the beam, J_{b0} and J_{b1} are the monopole and dipole components of the beam current density, J_{p0} is the density of the equilibrium reverse current, A_{z0} and A_{z1} are the equilibrium and perturbed z-components of the vector potential of an electromagnetic field, σ_{ch1} is the perturbed plasma conductivity, and $I_A = c^3\beta\gamma m/e$ is the limiting Alfvén current with e and m the charge and mass of an electron, γ the Lorentz factor, and $\beta = v_z/c$.

In the model of a rigid beam, we can write

$$J_{b1} = -Y\frac{\partial J_{b0}}{\partial r}, \quad (4)$$

$$A_{z1} = -D\frac{\partial A_{z0}}{\partial r}, \tag{5}$$

$$\sigma_{ch1} = -Y_{ch}\frac{\partial \sigma_{ch0}}{\partial r}, \tag{6}$$

where D and Y_{ch} are the transverse displacements of the symmetry axis of the collective magnetic field of a beam–plasma system and of the axis of the plasma channel, respectively.

With allowance for (2), (4), and (5), from (1) we can readily obtain the equation

$$\frac{\partial^2 Y}{\partial z^2} = k_s^2(\xi)(D - Y), \tag{7}$$

where

$$k_s^2(\xi) = 4\pi^2\left(\frac{I_b}{I_A}\right)\int_0^\infty drr\frac{J_{b0}(r, \xi)J_{n0}(r, \xi)}{I_b^2} \tag{8}$$

is the squared wavenumber of the beam electron hose oscillations and $J_{n0} = J_{b0} - J_{p0}$ is the net equilibrium (monopole) current density in a beam–plasma system.

To derive the expression for the net current I_n in the system,

$$I_n(\xi) = \int_0^\infty dr 2\pi r J_{n0}(r, \xi), \tag{9}$$

we take the product of the monopole component of Ampère's law (2) with rJ_{n0} and integrate the resulting equation over r from 0 to ∞. After some manipulations, we obtain the following equation for the net current $I_n(\rho, \xi) = \int_0^\rho d\eta 2\pi\eta J_{n0}(\eta, \xi)$ flowing in a tube of radius ρ in a beam–plasma system:

$$\xi_0\frac{\partial I_n}{\partial \xi} + I_n + I_n^2 F(\xi) = I_b^*, \tag{10}$$

where

$$\xi_0(\xi) = \frac{\displaystyle\int_0^\infty drr J_{n0}(r, \xi)\sigma_{ch0}(r, \xi)A_{z0}(r, \xi)}{\displaystyle\int_0^\infty drr[J_{n0}(r, \xi)]^2} \tag{11}$$

is the monopole skin depth and

$$F(\xi) = \frac{\displaystyle\int_0^\infty drr J_{n0}(r, \xi)\sigma_{ch0}(r, \xi)\frac{\partial}{\partial \xi}\left[\frac{A_{z0}(r, \xi)}{I_n(\xi)}\right]}{\displaystyle\int_0^\infty drr[J_{n0}(r, \xi)]^2}, \tag{12}$$

$$I_b^*(\xi) = I_n(\xi)\frac{\displaystyle\int_0^\infty drr J_{b0}(r, \xi)J_{n0}(r, \xi)}{\displaystyle\int_0^\infty drr[J_{n0}(r, \xi)]^2}, \tag{13}$$

$$A_{z0}(r, \xi) = -\frac{2}{c}\int_{R_c}^r \frac{d\rho}{\rho}I_n(\rho, \xi). \tag{14}$$

Equation (10) differs from the corresponding equations derived in [8, 9] in that it contains an additional (third) term on the right-hand side and the quantities ξ_0 and I_b^* are specified in a more general form. We integrate equation (3) multiplied by $\partial A_{z0}/\partial r$ over r from 0 to ∞. Going through some lengthy manipulations and taking into account (4)–(6) yields

$$\xi_1\frac{\partial}{\partial \xi}(DI_n) + DI_n + DI_n^2 S(\xi) + Y_{ch}\frac{\partial I_n}{\partial \xi}(\xi_0 - \xi_1)$$
$$+ Y_{ch}I_n^2[F(\xi) - S(\xi)] = YI_b^*, \tag{15}$$

where

$$\xi_1 = \frac{c}{4\pi}\frac{\displaystyle\int_0^\infty drr\sigma_{ch0}(r, \xi)\left(\frac{\partial A_{z0}}{\partial r}\right)}{\displaystyle\int_0^\infty drr[J_{n0}(r, \xi)]^2} \tag{16}$$

is the dipole skin depth, the functions $F(\xi)$ and ξ_0 are defined by (12) and (11), and

$$S(\xi) = \frac{c}{4\pi}\frac{\displaystyle\int_0^\infty drr\sigma_{ch0}(r, \xi)\left(\frac{\partial A_{z0}}{\partial r}\right)\frac{\partial}{\partial \xi}\left[\frac{1}{I_n(\xi)}\frac{\partial A_{z0}}{\partial z}\right]}{\displaystyle\int_0^\infty drr[J_{n0}(r, \xi)]^2}. \tag{17}$$

Unlike the familiar results of [8, 9], equation (15) contains terms with the coefficients $F(\xi)$ and $S(\xi)$, which, in our model, stem from $\partial I_b/\partial \xi \neq 0$ and $\partial R_b/\partial \xi \neq 0$. Additionally, the parameter I_b^* is specified in a more general form, and all of the coefficients can be evaluated for arbitrary radial profiles of $J_{b0}(r, \xi)$, $J_{p0}(r, \xi)$, and $\sigma_{ch0}(r, \xi)$.

The equation describing the dynamics of the transverse oscillations of the axis Y_{ch} of the plasma channel can be derived from the evolutionary equation for the plasma conductivity,

$$\frac{\partial \sigma_{ch}}{\partial \xi} = \Psi J_b(r, \xi) + \frac{\alpha_{av}\sigma_{ch}}{c} - \beta_r\sigma_{ch}^2. \tag{18}$$

Here, $\Psi \simeq 3 \times 10^6$ [cm/(A s)] is the rate of impact ionization of the background gas by the beam electrons; α_{av} is the avalanche ionization rate, which depends on $|E|/\rho$, where E is the collective electric field of the beam–plasma system and ρ is the background gas density [10]; and $\beta_r \simeq 7 \times 10^{-15}(\rho/\rho_n)$ [s/cm] with ρ/ρ_0 being the ratio of the gas density to its value under normal conditions.

With allowance for (4)–(6), the dipole component of the evolutionary equation (18) for the plasma conductivity has the form

$$\frac{\partial}{\partial\xi}\left(Y_{ch}\frac{\partial\sigma_{ch0}}{\partial r}\right) = \Psi Y\frac{\partial J_{b0}}{\partial r} + Y_{ch}\frac{\alpha_{av}}{c}\frac{\partial\sigma_{ch0}}{\partial r} - 2\beta_r\sigma_{ch0}Y_{ch}\frac{\partial\sigma_{ch0}}{\partial r}. \tag{19}$$

We integrate (19) multiplied by $r\partial A_{z0}/\partial r$ over r from r to ∞ and perform the necessary manipulations with allowance for (2) to arrive at

$$\frac{\partial Y_{ch}}{\partial\xi} + Y_{ch}[G(\xi) + \Gamma(\xi) - \Lambda(\xi)] = YL(\xi), \tag{20}$$

$$G(\xi) = \frac{\int_0^\infty dr\frac{\partial\sigma_{ch0}}{\partial r}rJ_{n0}(r,\xi)}{T(\xi)}, \tag{21}$$

$$\Gamma(\xi) = \frac{\int_0^\infty dr\beta_r(r,\xi)\sigma_{ch0}^2(r,\xi)rJ_{n0}(r,\xi)}{T(\xi)}, \tag{22}$$

$$\Lambda(\xi) = \frac{1}{c}\frac{\int_0^\infty dr\alpha_{av}(r,\xi)\sigma_{ch0}(r,\xi)rJ_{n0}(r,\xi)}{T(\xi)}, \tag{23}$$

$$L(\xi) = \frac{\int_0^\infty dr\Psi(r,\xi)J_{b0}(r,\xi)rJ_{n0}(r,\xi)}{T(\xi)}, \tag{24}$$

$$T(\xi) = \int_0^\infty dr\sigma_{ch0}(r,\xi)rJ_{n0}(r,\xi). \tag{25}$$

Equation (20) with (21)–(25) generalizes the equations obtained in [8] to the case in which both recombination and avalanche ionization are taken into account and all the parameters of the problem are time-dependent.

Hence, we have derived the equations that describe the dynamics of the RFI and generalize the familiar results of [8] to the case in which the beam radius and beam current are both time-dependent. We have obtained general formulas for evaluating monopole and dipole skin depths and the equation describing the transverse dynamics of a plasma channel with nonzero conductivity governed by the processes of impact ionization, avalanche ionization, and recombination.

REFERENCES

1. E. P. Lee and R. K. Cooper, Part. Accelerat. **7**, 83 (1976).
2. E. P. Lee, Phys. Fluids **21**, 1327 (1978).
3. H. S. Uhm and M. Lampe, Phys. Fluids **23**, 1574 (1980).
4. E. P. Nadezhdin and G. A. Sorokin, Fizika Plazmy **9**, 989 (1983) [Sov. J. Plasma Phys. **9**, 576 (1983)].
5. E. P. Nadezhdin and G. A. Sorokin, Fizika Plazmy **14**, 619 (1988) [Sov. J. Plasma Phys. **14**, 365 (1988)].
6. E. K. Kolesnikov and A. S. Manuĭlov, Zh. Tekh. Fiz. **60** (3), 40 (1990) [Sov. Phys. Tech. Phys. **35**, 298 (1990)].
7. E. K. Kolesnikov and A. S. Manuĭlov, Zh. Tekh. Fiz. **61** (12), 43 (1991) [Sov. Phys. Tech. Phys. **36**, 1351 (1991)].
8. M. Lampe, W. Sharp, R. F. Habbard, *et al.*, Phys. Fluids **27**, 2921 (1984).
9. J. E. Brandenburg, Report No. SAND-84-1026 (Sandia National Laboratory, 1985).
10. A. W. Ali, Laser Part. Beams **6** (1), 105 (1988).
11. E. K. Kolesnikov and A. S. Manuĭlov, Zh. Tekh. Fiz. **67** (7), 108 (1997) [J. Tech. Phys. **42**, 819 (1997)].
12. E. K. Kolesnikov and A. S. Manuĭlov, Radiotekh. Électron. **35**, 218 (1990).
13. R. F. Fernsler, R. F. Hubbard, and M. Lampe, J. Appl. Phys. **75**, 3278 (1994).

Translated by O. E. Khadin

SURFACES, ELECTRON AND ION EMISSION

Mechanism of Ionization of Organic Compounds on Tip Emitters

N. N. Grishin

Tananaev Institute of Chemistry and Technology of Rare Elements and Mineral Raw Materials,
Russian Academy of Sciences, Apatity, Murmansk oblast, 184200 Russia
Received May 28, 1998; in final form, April 13, 1999

Abstract—The current–voltage characteristics of organic compounds at a tip platinum emitter have been studied in 0.2–1.0 V/Å fields. It is shown that the parameter $(V - \varphi)$ does not affect the formation of the desorption ion barrier in the field ionization of atoms and molecules under a high potential gradient providing electron tunneling to the Fermi level of the emitter material and, therefore, the probability of the ion desorption cannot explicitly depend on this parameter. © 2000 MAIK "Nauka/Interperiodica".

Earlier, it was assumed that the theory of surface ionization would be applicable to all the instances of field ionization under a high potential gradient, including the ionization of inert-gas atoms in ionic projectors. This assumption was confirmed by the well-known experimental data on ionization of alkali metals in fields with a strengths up to 0.07 V/Å [1].

We studied the current–voltage characteristics of organic compounds on a tip platinum emitter ($\varphi = 532$ eV) at the field strength ranging from 0.2 to 1.0 V/Å. The results obtained allowed us to estimate the relative contributions of tunneling (field) and thermal effects to the ionization mechanism of these objects.

Consider the dependence of the degree of surface ionization on the electric-field strength using the method described in the well-known monograph on surface ionization [1]. Introduce the following notation: I is the ionic current from an emitter; U is the applied potential difference between the emitter and the counterelectrode; e is an electron or an electron charge; V is the energy of a molecule (atom) ionization far from the emitter surface; V' is the energy of a molecule (atom) ionization in the vicinity of the emitter surface; φ is the work function; x is the distance from an emitter; x_0 is the distance from the emitter at the maximum of the potential barrier; x_k is the critical recharge distance; x_p is the equilibrium distance of an adsorbed particle from the surface; F is the electric-field strength; F_k is the strength of the field providing the compensation of the barrier of the ion desorption caused by the image forces; k is the Boltzmann constant; T is the temperature; E_{pn} is the polarization energy of a molecule (atom) at $F \neq 0$; E_0 is the energy of the Fermi level; μ is the dipole moment; μ_0, μ_+, and m_e are the thermodynamic potentials of a molecule (atom), an ion, and an electron, respectively; l_+ and l_0 are the isothermal heat of evaporation of an ion and an uncharged molecule (atom) at

$F \neq 0$; l_{0+} and l_{00} are the isothermal heat of evaporation of an ion and an uncharged molecule (atom) at $F = 0$; q_0 is the work done to move an adsorbed particle located at x_p to x_k at $F = 0$; q is the work done to move an absorbed particle located at x_p to x_k at $F \neq 0$; λ_+ and λ_0 are the work done to move a desorbed ion and an uncharged molecule (atom) located at x_k to $x = \infty$ at $F \neq 0$; λ_{0+} and λ_{00} are the work done to move a desorbed ion and uncharged molecule (atom) located at x_k to $x = \infty$ at $F = 0$; α_e is the molecule (atom) polarizability; ν_+ and ν_0 are the flows of ions or uncharged molecules (atoms) from the unit area of the adsorbent surface per unit time; $\nu = \nu_+ + \nu_0$ is the total particle flow from the surface; $\alpha = \nu_+/\nu_0$ is the degree of surface ionization; and $\beta = \nu_+/\nu$ is the surface ionization coefficient defined as

$$\beta = \frac{1}{1 + 1/\alpha}. \tag{1}$$

In the theory of surface ionization, the dependence of the degree of surface ionization on the field strength and the temperature is obtained proceeding from the reaction
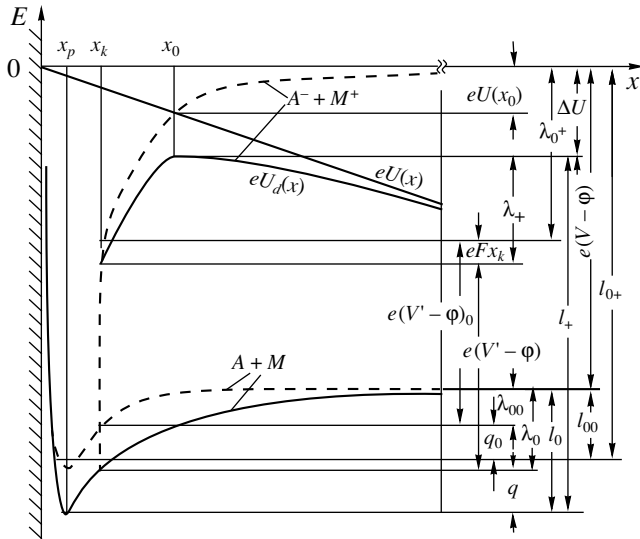
$$M \rightleftharpoons M^+ + e. \tag{2}$$

The condition for thermodynamic equilibrium is written in the form

$$\mu_0 = \mu_+ + \mu_e, \tag{3}$$

whence follows the expression for the degree of the surface ionization

$$\alpha = \frac{\nu_+}{\nu_0} = \frac{(1 - R_+)Q_+}{(1 - R_0)Q_0} \exp\frac{l_0 - l_+}{kT}, \tag{4}$$

Fig. 1. Curves of potential energy for the systems $A + M$ and $A^- + M^+$ in an applied accelerating field F (solid lines) and at $F = 0$ (dashed lines).

where the preexponential factor takes into account the statistical sums of a neutral (Q_0) and an ion (Q_+) and also the reflection from the barrier (R_+, R_0).

To illustrate the further transformations, consider the schematic potential curves for the adsorbent–neutral ($A + M$) and the adsorbent–positive ion ($A^- + M^+$) systems in a nonzero ($F \neq 0$) and the zero ($F = 0$) fields (Fig. 1). It is seen from Fig. 1 that

$$l_0 = q + \lambda_0, \tag{5}$$

$$l_+ = q + \lambda_0 + e(V - \varphi) - \Delta U, \tag{6}$$

$$l_0 - l_+ = e(\varphi - V) + \Delta U, \tag{7}$$

Table 1. Field strength (F_0), distance from the surface (x_0), and the work of the image forces ($e^2/4x_0$) at the maximum of the ion-desorption barrier

x_0, Å	F_0, V/Å	$e^2/4x_0$, erg $\times 10^{12}$
3.0	0.4	1.9
3.4	0.3	1.7
4.0	0.22	1.4
4.3	0.19	1.3
5.0	0.14	1.1
5.8	0.1	1.0
10.0	0.03	0.6

where

$$\Delta U = \frac{e^2}{4x_0} + eFx_0 \tag{8}$$

at the point of intersection

$$\frac{e^2}{4x_0} = eFx_0, \tag{9}$$

whence

$$x_0 = \frac{e}{2}\frac{1}{\sqrt{eF}}. \tag{10}$$

Substituting x_0 from (10) into (8), we have

$$\Delta U = e\sqrt{eF}. \tag{11}$$

Table 1 lists the x_0 and $e^2/4x_0$ values for various F. Equation (11) confirms the decrease of the Schottky ion-desorption potential barrier. Then, formula (4) takes the form well known from the theory of surface ionization

$$\alpha = \frac{(1-R_+)Q_+}{(1-R_\nu)Q_0}\exp\frac{e(\varphi - V + \sqrt{eF} + \Delta x)}{kT}, \tag{12}$$

$$F < F_{cr}.$$

Here, Δx is the correction for the work of the polarization forces, which takes into account the additional interactions with the surface. At $F > F_{cr}$, the theory of surface ionization yields

$$\alpha = \frac{(1-R_+)Q_+}{(1-R_\nu)Q_0}\exp\frac{e\left(\varphi - V - \frac{1}{e}\lambda_{0+} + Fx_k\right)}{kT}, \tag{13}$$

$$F > F_{cr}.$$

Since the tunneling transition of an electron from a molecule (atom) into the emitter metal is an isoenergetic process, formula (4) takes the form

$$\alpha = \frac{(1-R_+)Q_+}{(1-r_\nu)Q_0}\exp\frac{l_0 - \lambda_+}{kT}. \tag{14}$$

In this case, the neutral is desorbed from a lower vibrational level, whereas an ion, from the critical recharge distance, because, at distances from the surface less than x_k, it can exist only as a neutral. Strictly speaking, this is valid only at $T = 0$, whereas, at higher temperatures, one also has to take into account the Fermi distribution. The quantity l_0 in (14) can be represented as

$$l_0 = l_{00} + E_{pn}. \tag{15}$$

Some values of the polarization energy are listed in Table 2.

From the diagram shown in Fig. 1, it follows that

$$\lambda_+ = \lambda_{0+} + eFx_k - \Delta U \tag{16}$$

or that

$$\lambda_+ = \frac{e^2}{4x} + eFx_k - e\sqrt{eF}. \tag{17}$$

Finally, we have

$$l_0 - \lambda_+ = l_{00} + 1/2\alpha_e F^2 + \mu F \\ - eFx_k + e\sqrt{eF} - e^2/4x_k. \tag{18}$$

For atoms and nonpolar molecules, we obtain

$$l_0 - \lambda_+ = l_{00} + 1/2\alpha_e F^2 - eFx_k + e\sqrt{eF} - e^2/4x_k. \tag{19}$$

Formula (14) is valid in the case where the probability of the electron penetration through the potential barrier equals unity. If it is not so, the exponent in (14) should be preceded by an appropriate probability factor.

The probability of ionization $P(t)$ for time t in the theory of field ionization (which considers the phenomenon of electron penetration through the potential barrier) is given by the expression [2]

$$P(t) = 1 - \exp(-t/\tau), \tag{20}$$

where $\tau = 1/D\nu$, and ν is the frequency of electron collisions with the barrier equal to 10^{16} s^{-1}.

If the time t is short in comparison with the duration of the preionization state, we have

$$P(t) = t\nu D. \tag{21}$$

If the potential barrier of an electron is represented by an equilateral triangle of height $(V - 2\sqrt{e^3 F})$ and the base $(V - \varphi)/eF$, the probability of the electron penetra-
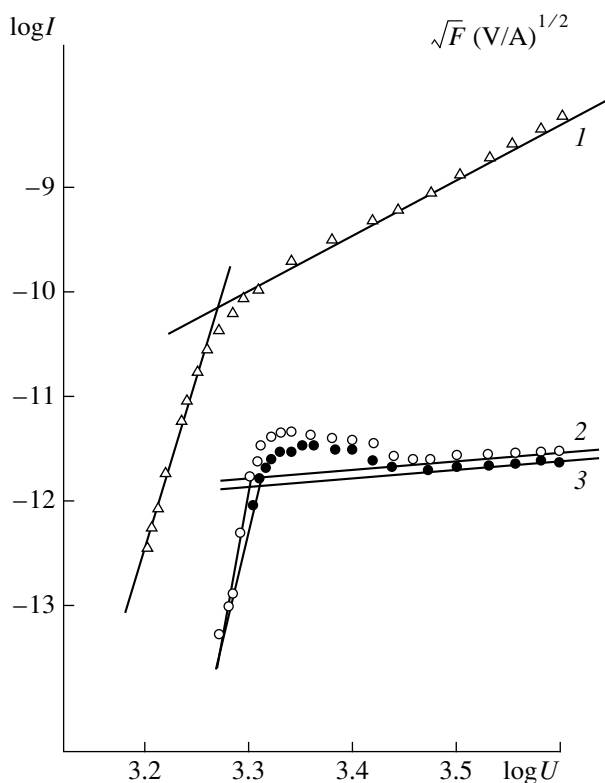

Fig. 2. Logarithmic current–voltage characteristics of a tip platinum emitter for methanol. (*1*) Total ionic current within an angle of ion accumulation $\sim\pi$ (rad), (*2*) ionic current within an angle of ion accumulation $\pi/6$ (rad), (*3*) ionic current at the exit of a mass spectrometer calculated from the total intensity of all the spectrum lines.

tion through the barrier D is given by the expression

$$D = \exp\left[-4.55 \times 10^7 \sqrt{V - 7.6F^{1/2}} \frac{V - \varphi}{F}\right]. \tag{22}$$

Then formula (14) takes the form

$$\alpha = \frac{(1 - R_+)Q_+}{(1 - R_0)Q_0} P(t) \exp \frac{l_{00} + 1/2\alpha_e F^2 + \mu F - eFx_k - e\sqrt{eF} - e^2/4x_k + \Delta x}{kT}. \tag{23}$$

Unlike (12), expression (23) contains no term describing the contribution $(V - \varphi)$ to the ion-desorption barrier.

Now, estimate the contributions of the terms in the numerator of the exponent depending on the field strength for ethanol (Table 3).

The critical recharge distance was estimated from a simplified Muller equation [2], which is quite justified because of the smallness of the polarization-energy contribution,

$$eFx_k = (V - \varphi) - e^2/4x_k. \tag{24}$$

Similar to (12), the values of the exponential factor in (23) are inconsistent with the experimental data on field ionization for organic compounds, in particular, with the fields necessary for appearance of molecular ions. Now, compare the theoretical versions of (12) and (23) with the experimental data (Fig. 2).

Using the data on the field ionization of inert gases, the ionic current from the emitter can be represented by the product of two functions—that of influx of the particles to the emitter surface (Z) and that of the ionization probability (β)

$$I = Z\beta. \tag{25}$$

The function Z was determined from the experimental points on a mildly sloping segment of the current–voltage characteristic ($U > U'$) in the form

$$\log Z = K + n \log U, \qquad (26)$$

where $\beta = 1$. The correlation coefficient exceeded 0.99. Then, the function β on the segment $U < U'$ was determined numerically by dividing I into Z,

$$\beta = I/Z. \qquad (27)$$

The experimental β values are listed in Table 4. We transformed the scale of applied voltages into the scale of field strengths using the current–voltage characteristic for methanol, whose characteristic was obtained by Beckey [4] in the logarithmic "current–field strength" coordinates.

The values of the field strength corresponding to the inflection points of the current–voltage characteristics determined for various substances (in V/Å) are 0.69 for

CH_3OH, 0.64 for C_2H_5OH, 0.48 for C_6H_6, 0.42 for C_7H_8, and 0.40 for $(C_2H_5)_2NH$ (Table 2). The critical ionization distances for all the compounds calculated by (24) with the use of the above F-values were almost the same, 7.2 Å (Table 4), which is quite understandable, because x_k characterizes the width of the barrier that should be overcome by an electron in the tunneling transition into the metal, whereas an electron cannot "recognize" any concrete object (potential well) from which it would perform this transition.

Figure 3 shows the dependence of the field strength sufficient for attaining the value $x_k = 7.2$ Å on the parameter $(V - \varphi)$ (line 2) and the dependence of the field strength providing the attainment of the value x_0 on the same parameter (line 1). These dependences lead to a conclusion that, for all the ionization objects (molecules and atoms) whose values (line 2) lie under line 1 ($V - \varphi > 0.35$ eV), the critical recharge distance should lie on the falling branch of the ion-desorption barrier.

Table 2. Polarization energy of molecules in an intense electric field $E_{pn} = \mu F + 1/2\alpha_e F^2$, erg $\times 10^{13}$

| Compound	CH_3OH (Methanol)		C_2H_5OH (Ethanol)		C_6H_6 (Benzene)		C_7H_8 (Toluene)		$(C_2H_5)_2NH$ (Diethylamine)	
Polarizability [3] $\alpha_e \times 10^{-24}$, cm^3	3.29		5.41		10.32		12.3		10.2	
Dipole moment [3] $\mu[D]$	1.7		1.69		0		0.375		0.92	
Field strength F, B/Å	1	0.69*	1	0.64*	1	0.48*	1	0.42*	1	0.4*
μF	5.7	4.0	5.6	3.6	–	–	1.2	0.5	3.1	1.2
$1/2\alpha_e F^2$	1.8	0.9	3.0	1.3	5.7	1.3	6.8	1.2	5.6	0.9
E_{pn}	7.5	4.9	8.6	4.9	5.7	1.3	8.0	1.7	8.7	2.1

* Position of the inflection point of the current–voltage characteristic.

Table 3. The values of the individual terms of the exponent in equation (23) for ethyl alcohol, in erg $\times 10^{12}$ ($\delta E = l_{00} + 1/2\alpha_e F^2 + \mu F - eFx_k + e\sqrt{eF} - e^2/4x_k$)

	$F_1 = 0.1$ V/Å	$F_2 = 0.3$ V/Å	$F_3 = 0.5$ V/Å	$F_4 = 1$ V/Å
μF	0.06	0.17	0.28	0.56
$1/2\alpha_e, F^2$	0.003	0.03	0.08	0.04
x_k, Å	51.5	17.15	10.27	5.11
$\dfrac{e^2}{4x_k}$	0.11	0.33	0.56	1.13
eFx_k	8.23	8.22	8.21	8.16
$e\sqrt{eF}$	1.9	3.3	4.1	5.8
$\lambda_{00} = \Delta H\ (295)$		0.7		
δE	−4.92	−3.65	−2.91	−1.13
$\exp\dfrac{\delta E}{kT}$	4×10^{-54}	3×10^{-40}	2×10^{-32}	8×10^{-13}
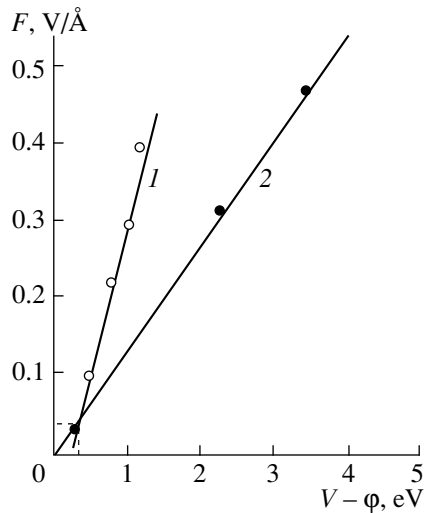
Fig. 3. Field strengths providing (*1*) the equality $x_k = x_0$ and (*2*) the values $x_k = 7.2$ Å as functions of the ionization parameter.
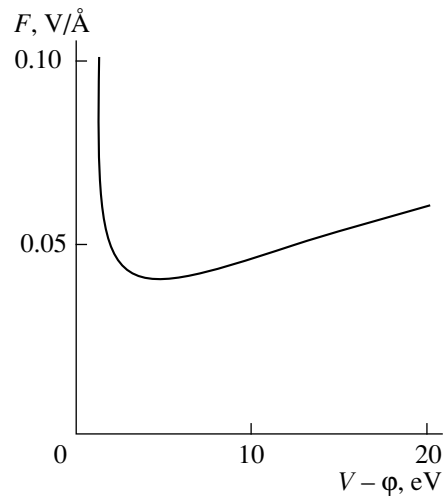
Fig. 4. Curve separating the regions where the tunneling and the thermal ionization mechanisms for the tip platinum emitter act.

Since there is no ion-desorption barrier, the exponential factor is eliminated, and formula (23) takes the form

$$\alpha = \frac{(1-R_+)Q_+}{(1-R_0)Q_0}P(t) \qquad (28)$$

or

$$\alpha = AD, \qquad (29)$$

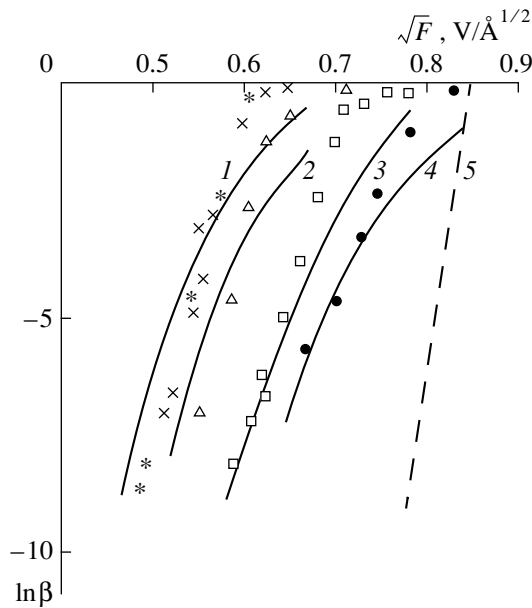where the factor A takes into account all the F-dependent terms, and D is given by (22).

Equating the exponents in (22) and (12), we arrive at the equation for the ratio of the ionization parameters $(V-\varphi)$ and F providing their equal contributions to ionization. If V and φ are given in electron volts and F is in volts per angstrom, the equation at 300 K takes the form

$$0.455(V - 7.6F^{1/2})^{1/2}(V-\varphi)$$
$$= 36.4(V-\varphi)F - 1.38F^{3/2}. \qquad (30)$$

Figure 4 shows this dependence for field ionization

Table 4. Function β_F calculated from the experimental data for some organic compounds

Methanol CH$_3$OH			Ethanol C$_2$H$_5$OH			Benzene C$_6$H$_6$			Toluene C$_7$H$_8$			Diethylamine (C$_2$H$_5$)$_2$NH		
ln$I$	F, V/Å	$\sqrt{F}$	ln$I$	F, V/Å	$\sqrt{F}$	ln$I$	F, V/Å	$\sqrt{F}$	ln$I$	F, V/Å	$\sqrt{F}$	ln$I$	F, V/Å	$\sqrt{F}$
−5.68	0.44	0.67	−8.19	0.36	0.60	−6.90	0.32	0.56	−7.01	0.27	0.52	−8.68	0.24	0.49
−4.60	0.49	0.70	−7.30	0.38	0.62	−4.60	0.34	0.59	−4.92	0.30	0.55	−8.15	0.25	0.50
−3.27	0.53	0.73	−6.60	0.40	0.63	−2.56	0.37	0.61	−3.14	0.32	0.56	−6.69	0.28	0.53
−2.30	0.56	0.75	−6.42	0.40	0.63	−1.04	0.40	0.63	−2.92	0.32	0.57	−4.63	0.30	0.55
−1.00	0.61	0.78	−5.01	0.43	0.65	−0.63	0.42	0.65	−0.83	0.35	0.60	−4.17	0.32	0.56
0.0	0.69	0.83	−3.84	0.45	0.67	0.0	0.48	0.69	−0.22	0.40	0.63	−2.51	0.34	0.58
			−1.2	0.49	0.70				−0.17	0.42	0.65	−0.41	0.37	0.61
			−0.69	0.50	0.71							−0.21	0.40	0.64
			−0.28	0.54	0.73									
			−0.23	0.56	0.75									
			0.0	0.64	0.80									

$A = 1.3 \times 10^3$ 3.6×10^3 1.1×10^3 1.5×10^3 6.6×10^2

ln$A = 7.2$ 8.2   7.0   7.3   6.5

x_k, Å at $\beta \longrightarrow 1$

7.3   7.27   7.12   7.13   7.35

Fig. 5. Ionization probability as a function of the applied field. Theoretically calculated curves for (*1*) diethylamine and toluene (almost coincide), (*2*) benzene, (*3*) ethanol, (*4*) methanol, and (*5*) the curve of the surface ionization coefficient. Experimental data for ● methanol, □ ethanol, △ benzene, × toluene and * diethylamine.

on a platinum emitter. As is seen, the dependence is nonmonotonic: for atoms and molecules with high ionization energies, the dependence of F on $(V - \varphi)$ is a slightly varying function decreasing with a decrease in $(V - \varphi)$. Beginning with $(V - \varphi) \approx 2$ eV, it starts increasing, thus showing an increase in the relative contribution of thermal ionization. Then, it asymtotically tends to the value $(V - \varphi) = 1.66$ eV. The ionization objects with the $(V - \varphi)$ values much lower than the above value are ionized above the barrier according to (12), whereas, at the $(V - \varphi)$ values exceeding the above value, they are ionized during tunneling according to formula (29). The ratio of the parameters F and $(V - \varphi)$ above the curve correspond to the prevalent tunneling mechanism. At $(V - \varphi) < 1.6$ eV at 300 K, the above-the-barrier mechanism becomes more probable than tunneling.

In accordance with the logic of the derivation of (23), the current from the emitter is determined by the well-known formula of the theory of surface ionization [1]:

$$I = e\nu_{+} = e\beta\nu = \frac{e\nu}{1 + \dfrac{1}{\alpha}}, \qquad (31)$$

whereas the ionic current normalized to the molecule inflow, by the formula

$$I = e\beta = \frac{e}{1 + \dfrac{1}{\alpha}}. \qquad (32)$$

Function β from (32) corresponds to the β values listed in Table 4. The function α is given by (29); A can be determined as the proportionality coefficient between the β values calculated theoretically by (32) and obtained experimentally (Table 4). Thus, at $\beta = 0.5$, we have $\alpha = 1$ and $\ln\beta = -0.69$. Then, using the data from Table 4, one can determine the field strength at which the experimentally obtained β value would equal 0.5 and, using (22), also the D value at this field strength and, finally, using (29), also the value of A. The A value can also be chosen by checking the consistency of all the experimental data, e.g., with the aid of the least squares method (Table 4).

Since our aim is to verify the adequacy of the theoretical models of the field and surface ionization to the experimental data, it is convenient to represent the data obtained as functions of $\sqrt{F}$, i.e., as the characteristic function taken from the theory of surface ionization [formula (12)]. Then, at low values of the surface-ionization coefficient ($\beta < 0.5$) and in view of formulas (12) and (32), the quantity $\ln I$ is a linear function of $\sqrt{F}$ and has a constant slope coefficient at a fixed temperature.

Figure 5 shows function (32) theoretically calculated using D from (22) and the constant coefficient A from Table 4 (solid curves) and the experimentally determined ionization coefficients β listed in Table 4 (various symbols). The dashed line shows the surface ionization coefficient calculated in the same way from (32), where the α value was calculated by (12). In the latter case, the coefficient A attained an incredible value of 5.2×10^{38}, which corresponds to e^{90}. This fact alone indicates the inadequacy of formula (12) to the real physicochemical situation and confirms the assumption that the factor $(V - \varphi)$ is not a component of the barrier of ion desorption from the emitter surface at $(V - \varphi) > 1.66$ eV.

The experimentally obtained β values fit the theoretical curves of the tunneling probability (Fig. 5) with only one essential stipulation—in the vicinity of the saturation of the current–voltage characteristic ($\beta \longrightarrow 1$), the experimental β-values are higher than the theoretically calculated ones in all the cases, which indicates that this fact is not of an accidental nature.

A similar phenomenon was also observed earlier. Thus, studying the field ionization for adsorbed hydrogen layers and measuring the ionic current from individual faces of an emitter, Gomer [5] revealed alternating maxima and minima despite the fact that dependence of the total ionic current had the conventional form. He attributed these fluctuations in the current to the adsorption effect.

To confirm this assumption, we studied the current–voltage characteristics within a complete solid angle of ion collection ($\sim \pi$, rad), the current–voltage characteristics determined during recording of ionic currents in front of the exit slit of the field ion source of the mass-

spectrometer (angle $\pi/6$), and the current–voltage characteristics at the exit of the mass spectrometer measured as sums of the intensities of all the lines of the mass spectrum. The data obtained for methanol (Fig. 2) are well consistent with the Gomer data [5].

Performing experiments at different temperatures, inlet pressures, field strengths, and amplitudes and duration of desorbing pulses with simultaneous recording of the corresponding mass spectra, we confirmed the effect of adsorption on the slope of the current–voltage characteristics. Thus, the deviation of the experimental points from the theoretically calculated curves in Fig. 5 is quite regular.

The results obtained can be interpreted in the following way. During field ionization of atoms and molecules under a high potential gradient providing electron tunneling to the Fermi level of the emitter material, the parameter $(V - \varphi)$ does not participate in the formation of the ion-desorption barrier, and therefore the probability of ion desorption does not explicitly depend on this parameter.

For the ionization objects with $(V - \varphi) > 0.35$ eV, the critical recharge distance lies on the decreasing branch of the desorption barrier, and therefore, if the ionization proceeds by the tunneling mechanism, a desorption barrier is absent. At $(V - \varphi) > 1.66$ eV and at the temperature 300 K, this condition is fulfilled.

In the fields that cannot provide electron tunneling to the emitter material, $(V - \varphi) < 1.66$ eV, only surface ionization occurs [formula (12)]. In this case, the external field promotes the reduction of the Schottky potential barrier of ion desorption.

The discrepancy between (12) and the experimental data on ionization in an electric field with a high potential gradient is explained by the fact that the theory of surface ionization is developed for thermodynamically equilibrium systems, whereas the tunneling ionization is a typical irreversible process.

REFERENCES

1. É. Ya. Zandberg and N. I. Ionov, *Surface Ionization* (Nauka, Moscow, 1969).

2. E. Muller and T. Tson, *Autoionic Microscopy* (Metallurgiya, Moscow, 1972).

3. *Handbook of Chemistry and Physics*, Ed. by D. R. Lide, 74th ed. (1993–1994).

4. H. D. Beckey, *Principles of Field Ionization and Field-Desorption Mass Spectrometry* (Pergamon Press, Oxford, 1972), p. 325.

5. R. Gomer, J. Chem. Phys. **52** (3), 2227 (1970).

Translated by L. Man

===== **THEORETICAL AND MATHEMATICAL PHYSICS** =====

On the Theory of Small-Amplitude Vibrations
of Curved-Surface Diaphragms

G. F. Glinskiĭ

St. Petersburg State University of Electrical Engineering, St. Petersburg, 197376 Russia

Received September 23, 1998

Abstract—An equation is derived describing small-amplitude vibrations of an arbitrary curved diaphragm, whose surface is considered as a two-dimensional Riemannian space. The derivation is based on the variational principle, from which the motion equation and conservation law follow in a form invariant with respect to arbitrary transformations of coordinates on the diaphragm surface. It has been shown that the wave equation, along with the two-dimensional Laplace–Beltrami operator, includes an additional term proportional to the scalar curvature of the diaphragm surface. As an example, the equations are considered for a spherical diaphragm and a catenoid-shaped diaphragm with a minimal surface of revolution. © *2000 MAIK "Nauka/Interperiodica".*

INTRODUCTION

The problem of determining the natural and forced vibrations of curved diaphragms arises when certain matters are considered concerning applied acoustics, hydroacoustics, aero- and hydrodynamics. However, formulation of the problem itself and solving it go far beyond the scope of purely practical matters and are themselves of scientific interest.

The surface of an arbitrarily curved diaphragm may be considered as a two-dimensional (2D) Riemannian space embedded in an ordinary (3D) Euclidean space. Parametric definition of this surface uniquely defines a curvilinear coordinate system and metric-tensor components in a 2D Riemannian space. If we are interested in small deviations of the curved diaphragm from equilibrium, then the problem must be reduced to solving the wave equation within the curved 2D space under consideration. The shape assumed by the curved diaphragm after the boundary conditions and external static forces are specified will be referred to as an equilibrium shape. For example, the diaphragm may be stretched over a rigid contour representing a curve in a 3D space, which is equivalent to imposing the corresponding boundary conditions. If no external forces are present, then the equilibrium shape of the diaphragm must correspond to its minimal surface energy. For the ideal uniform diaphragms with no flexural rigidity considered in this paper, this corresponds to their minimal surface areas. Thus, in the absence of external forces, an ideal diaphragm must assume the minimal-surface shape. Since the diaphragm surface area and its other integral characteristics are independent of the method of parameterization of its surface, all equations obtained must be invariant under arbitrary transformations of coordinates (parameters); i.e., they must have a generally covariant form. The latter may be obtained using the technique of covariant or absolute differentiation.

In this paper, a unified invariant approach is developed to solve the problem of small-amplitude vibrations of an arbitrarily curved diaphragm. The basis for this approach is the variational principle, from which the motion equation and conservation law follow immediately in a generally covariant form. As degrees of freedom characterizing small deviations of the diaphragm from the equilibrium, the diaphragm displacements along the normal to the surface at every point are considered. With respect to an arbitrary transformation of coordinates on the diaphragm surface, such a quantity behaves as a pseudoscalar. It is shown in this paper that the vibration equation, along with the Laplace–Beltrami operator, includes an additional term proportional to the scalar curvature of the diaphragm surface. As an example, the equations are considered for a spherical diaphragm and catenoid-shaped diaphragm having a minimal surface of revolution. The results obtained can be used for calculating the natural and forced vibrations of diaphragms of arbitrary shape and also for analyzing the stability of interface-containing systems.

THE CURVED-DIAPHRAGM SURFACE
AS A TWO-DIMENSIONAL RIEMANNIAN SPACE

Consider a curved diaphragm in a plane 3D space. Let **x** be a radius vector of an arbitrary point of this space. The diaphragm surface can be represented in a parametric form using the following vector equation:

$$\mathbf{x} = \mathbf{x}(\xi^1, \xi^2), \tag{1}$$

where ξ^α ($\alpha = 1, 2$) are arbitrary independent parameters.

If Cartesian coordinates are chosen in the 3D space under consideration, then this equation is equivalent to three equations for the projections of the radius vector on the corresponding coordinate axes $x^i = x^i(\xi^1, \xi^2)$ ($i = 1, 2, 3$).[1] The parameters ξ^α in these equations play the role of curvilinear coordinates on the diaphragm surface, and families of the curves $\xi^1 = $ const and $\xi^2 = $ const determine the corresponding coordinate grid. Therefore, the diaphragm surface can be considered as a 2D Riemannian space with a specified curvilinear coordinate system ξ^α.

Consider two infinitely near points $\mathbf{x}$ and $\mathbf{x} + d\mathbf{x}$ on the diaphragm surface. According to (1),

$$d\mathbf{x} = \frac{d\mathbf{x}}{d\xi^\alpha}d\xi^\alpha = \mathbf{a}_\alpha d\xi^\alpha,$$

where $\mathbf{a}_\alpha = d\mathbf{x}/d\xi^\alpha$ represents two vectors ($\alpha = 1, 2$) that lie in the plane tangent to the diaphragm surface at the point $\mathbf{x}$ and form a local generally nonorthogonal basis at this point. The squared distance between these two points is determined by the expression

$$dl^2 = d\mathbf{x}^2 = \mathbf{a}_\alpha \cdot \mathbf{a}_\beta d\xi^\alpha d\xi^\beta = G_{\alpha\beta}d\xi^\alpha d\xi^\beta. \quad (2)$$

It follows from here that the quantities

$$G_{\alpha\beta} = \mathbf{a}_\alpha \cdot \mathbf{a}_\beta = \frac{dx^i dx^i}{d\xi^\alpha d\xi^\beta}, \quad (3)$$

symmetrical with respect to the index permutation, represent components of a metric tensor of the Riemannian space under consideration [1]. They are also termed coefficients of the first quadratic form of the surface. If the surface metric is known, the diaphragm surface area is determined by the integral

$$\sigma = \int \sqrt{G}d\xi^1 d\xi^2, \quad (4)$$

where $G = \det\|G_{\alpha\beta}\|$.

When arbitrary transformations of the curvilinear coordinates are performed, $\xi^\alpha \mathbf{a}_\alpha$ are transformed as vector components. However, their ordinary derivatives are not components of any tensor; only covariant (absolute) derivatives, which we denote by the symbol ∇_α, are transformed as tensor components. Thus, the covariant derivative of $\mathbf{a}_\beta$ with respect to ξ^α is determined by the expression

$$\nabla_\alpha \mathbf{a}_\beta = \frac{\partial \mathbf{a}_\beta}{\partial \xi^\alpha} - \Gamma^\gamma_{\alpha\beta}\mathbf{a}_\gamma, \quad (5)$$

where $\Gamma^\gamma_{\alpha\beta} = (1/2)G^{\gamma\sigma}(\partial G_{\sigma\alpha}/\partial\xi^\beta + \partial G_{\sigma\beta}/\partial\xi^\alpha - \partial G_{\alpha\beta}/\partial\xi^\sigma)$ are the coefficients of connection or Christoffel symbols [1].

Along with the basis vectors $\mathbf{a}_\alpha$, at each point of the diaphragm surface, we can determine the normal unit vector that is perpendicular to them:

$$\mathbf{n} = \frac{[\mathbf{a}_1 \times \mathbf{a}_2]}{\sqrt{G}}. \quad (6)$$

It can be easily verified that the vector satisfies the following conditions:

$$\mathbf{n} \cdot \mathbf{a}_\alpha = 0 \ (\alpha = 1, 2), \quad (7)$$

$$\mathbf{n} \cdot \mathbf{n} = 1. \quad (8)$$

When an arbitrary coordinate transformation is performed, ξ^α is multiplied by ± 1 depending on the sign of the Jacobian of transformation; i.e., in the 2D Riemannian space under consideration, it is a pseudoscalar. This becomes apparent if its components are written as

$$n^i = \frac{1}{2}e^{ikl}a^k_\alpha a^l_\beta \frac{\varepsilon^{\alpha\beta}}{\sqrt{G}},$$

where e^{ikl} and $\varepsilon^{\alpha\beta}$ are, respectively, the 3D and 2D Levi–Civita symbols ($e^{123} = 1$, $\varepsilon^{12} = 1$).

Since the covariant and ordinary derivatives coincide for the scalars and pseudoscalars, the following relation is valid:

$$\nabla_\alpha \mathbf{n} = \frac{\partial \mathbf{n}}{\partial \xi^\alpha}. \quad (9)$$

Along with the coefficients of the first quadratic form $G_{\alpha\beta}$, the coefficients of the second quadratic form $b_{\alpha\beta}$, which are components of the symmetric pseudotensor, play an important role in the theory of surfaces. They appear in the so-called derivational formulas of Gauss and Weingarten [1]:

$$\nabla_\alpha \mathbf{a}_\beta = b_{\alpha\beta}\mathbf{n}, \quad (10)$$

$$\nabla_\alpha \mathbf{n} = -G^{\beta\gamma}b_{\alpha\beta}\mathbf{a}_\gamma. \quad (11)$$

To determine $b_{\alpha\beta}$, we will use formula (10). Multiplying scalarwise the left and right sides of this equation by $\mathbf{n}$, we obtain

$$b_{\alpha\beta} = \frac{\partial^2 \mathbf{x}}{d\xi^\alpha d\xi^\beta}\mathbf{n}. \quad (12)$$

Here we use relations (5), (7), and (8) and take into account that $\partial \mathbf{a}_\beta/\partial\xi^\alpha = \partial^2\mathbf{x}/\partial\xi^\alpha\partial\xi^\beta$.

This formula allows explicit calculation of the coefficients $b_{\alpha\beta}$ if the surface is specified parametrically (formula (1)). We may represent it in a form that is more suitable for practical calculations if we take into account that in the right side of (12), in accordance with

---

[1] Here and below, the roman indices $i, k, l, \ldots$ assume values 1, 2, and 3; Greek indices $\alpha, \beta, \gamma, \ldots$ assume values 1 and 2; for doubly occurring indices, summation is performed from 1 to 3 and from 1 to 2, respectively, unless otherwise stated. Since the metric of the 3D space is Euclidean, we will make no distinction between covariant and countervariant components of the 3-tensors and all roman indices will be superscribed.

(6), a mixed vector product is present that can be written in the form of the determinant

$$b_{\alpha\beta} = \frac{1}{\sqrt{G}} \left\| \begin{matrix} \dfrac{\partial^2 x^1}{\partial \xi^\alpha \partial \xi^\beta} & \dfrac{\partial^2 x^2}{\partial \xi^\alpha \partial \xi^\beta} & \dfrac{\partial^2 x^3}{\partial \xi^\alpha \partial \xi^\beta} \\[3mm] \dfrac{\partial x^1}{\partial \xi^1} & \dfrac{\partial x^2}{\partial \xi^1} & \dfrac{\partial x^3}{\partial \xi^1} \\[3mm] \dfrac{\partial x^1}{\partial \xi^2} & \dfrac{\partial x^2}{\partial \xi^2} & \dfrac{\partial x^3}{\partial \xi^2} \end{matrix} \right\|. \tag{13}$$

The coefficients of the second quadratic form allow us to calculate the Riemann–Christoffel curvature tensor

$$R_{\alpha\beta\gamma\sigma} = b_{\alpha\gamma}b_{\beta\sigma} - b_{\alpha\beta}b_{\beta\gamma}$$

and the scalar curvature

$$R = G^{\alpha\gamma}G^{\beta\sigma}R_{\alpha\beta\gamma\sigma}. \tag{14}$$

Here $G^{\alpha\beta}$ are the countervariant components of the metric tensor $G^{\alpha\beta}G_{\beta\gamma} = \delta_\gamma^\alpha$, and δ_γ^α are components of the unit tensor. The mean H and Gaussian K curvatures, usually used in the theory of surfaces, are determined by the relations

$$H = \frac{1}{2}G^{\alpha\beta}b_{\alpha\beta}, \tag{15}$$

$$K = \frac{b}{G} = \frac{1}{2}R, \tag{16}$$

where $b = \det\|b_{\alpha\beta}\|$.

They are related to the principal radii of curvature R_1 and R_2 by the known formulas

$$H = \frac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R_2}\right), \quad K = \frac{1}{R_1 R_2}.$$

With respect to arbitrary transformations of coordinates in the 2D Riemannian space, H behaves as a scalar and K is a scalar.

THE VARIATIONAL PRINCIPLE AND THE EQUILIBRIUM SHAPE OF THE DIAPHRAGM

When calculating the equilibrium shape of the diaphragm, we proceed from the principle that, with no external forces present, the diaphragm assumes the shape with which its surface energy U_0 is minimal. For ideal diaphragms, this energy is proportional to the diaphragm area (formula (4)):

$$U_0 = T\int \sqrt{G}d\xi^1 d\xi^2. \tag{17}$$

Here, T is the surface tension of the diaphragm, and the integration is performed over the region limited by the

rigid contour, over which the diaphragm is stretched. The energy is minimal on the condition that the first variation of functional (17) with respect to $\mathbf{x}(\xi^1, \xi^2)$ vanishes:

$$\delta U_0 = T\int \delta\sqrt{G}d\xi^1 d\xi^2$$

$$= T\int \sqrt{G}G^{\alpha\beta}\frac{\partial \mathbf{x}}{\partial \xi^\beta}\left(\frac{\partial}{\partial \xi^\alpha}\delta\mathbf{x}\right)d\xi^1 d\xi^2 \tag{18}$$

$$= -T\int \frac{\partial}{\partial \xi^\alpha}\left(\sqrt{G}G^{\alpha\beta}\frac{\partial \mathbf{x}}{\partial \xi^\beta}\right)\delta\mathbf{x}d\xi^1 d\xi^2 = 0.$$

Here we use the formula $\delta G = GG^{\alpha\beta}\delta G_{\alpha\beta}$ and, when integrating by parts, i.e., applying the Green theorem, we take into account that, for a rigidly fixed diaphragm, the integration-boundary variation $\delta\mathbf{x}(\xi^1, \xi^2) = 0$. Owing to the arbitrariness of $\delta\mathbf{x}(\xi^1, \xi^2)$, a nonlinear equation follows from (18), which determines the equilibrium shape of the diaphragm; i.e., the equation of the minimal surface area

$$\frac{1}{\sqrt{G}}\frac{\partial}{\partial \xi^\alpha}\left(\sqrt{G}G^{\alpha\beta}\frac{\partial \mathbf{x}}{\partial \xi^\beta}\right) = 0. \tag{19}$$

In the presence of external forces due, for example, to the difference of pressures on the two sides of the diaphragm Δp_0, the variation in the surface energy δU_0 should be equal to the work performed by these forces in displacing the diaphragm by $\delta\mathbf{x}(\xi^1, \xi^2)$ [2]:

$$\delta U_0 = -\int \Delta p_0 \mathbf{n}\delta\mathbf{x}\sqrt{G}d\xi^1 d\xi^2. \tag{20}$$

In this case, the equilibrium shape of the diaphragm is determined from the solution of the equation with a right-hand side

$$\frac{1}{\sqrt{G}}\frac{\partial}{\partial \xi^\alpha}\left(\sqrt{G}G^{\alpha\beta}\frac{\partial \mathbf{x}}{\partial \xi^\beta}\right) = \frac{\Delta p_0}{T}\mathbf{n}. \tag{21}$$

The latter can be written componentwise as

$$\frac{1}{\sqrt{G}}\frac{\partial}{\partial \xi^\alpha}\left(\sqrt{G}G^{\alpha\beta}\frac{\partial x^i}{\partial \xi^\beta}\right) = \frac{\Delta p_0}{T}n^i \quad (i = 1, 2, 3). \tag{22}$$

Let us show that only one of the three equations of system (22) is independent. For example, if $x^3(\xi^1, \xi^2)$ satisfies equation (22), then two other equations for $x^1(\xi^1, \xi^2)$ and $x^2(\xi^1, \xi^2)$ are identically true. In order to do this, we will multiply the left and right sides of equation (10) by $G^{\alpha\beta}$:

$$\nabla_\alpha G^{\alpha\beta}\mathbf{a}_{\alpha\beta} = G^{\alpha\beta}b_{\alpha\beta}\mathbf{n}.$$

Here, we take into account that the components of the metric tensor can be placed under the sign of covariant differentiation. In view of (5), this equation can also be

represented as

$$\frac{1}{\sqrt{G}}\frac{\partial}{\partial\xi^\alpha}\left(\sqrt{G}G^{\alpha\beta}\frac{\partial\mathbf{x}}{\partial\xi^\beta}\right) = G^{\alpha\beta}b_{\alpha\beta}\mathbf{n}. \qquad (23)$$

It follows from this that the mean curvature of the minimal surface $H = 0$. Multiplying scalarly both sides of (23) by $\partial\mathbf{x}/\partial\xi^\gamma$ and taking into account orthogonality condition (7), we arrive at the identity

$$\frac{1}{\sqrt{G}}\frac{\partial}{\partial\xi^\alpha}\left(\sqrt{G}G^{\alpha\beta}\frac{\partial\mathbf{x}}{\partial\xi^\beta}\right)\frac{\partial\mathbf{x}}{\partial\xi^\gamma} = 0.$$

Assuming that $x^3(\xi^1, \xi^2)$ satisfies equation (21), we will write this identity in the following form:

$$\sum_{i=1}^{2}\frac{1}{\sqrt{G}}\frac{\partial}{\partial\xi^\alpha}\left(\sqrt{G}G^{\alpha\beta}\frac{\partial x^i}{\partial\xi^\beta}\right)\frac{\partial x^i}{\partial\xi^\gamma}+\frac{\Delta p_0}{T}n^3\frac{\partial x^3}{\partial\xi^\gamma} = 0.$$

In accordance with orthogonality condition (7),

$$n^3\frac{\partial x^3}{\partial\xi^\gamma} = -\sum_{i=1}^{2}n^i\frac{\partial x^i}{\partial\xi^\gamma}.$$

As a result, we obtain

$$\sum_{i=1}^{2}\left[\frac{1}{\sqrt{G}}\frac{\partial}{\partial\xi^\alpha}\left(\sqrt{G}G^{\alpha\beta}\frac{\partial x^i}{\partial\xi^\beta}\right)-\frac{\Delta p_0}{T}n^i\right]\frac{\partial x^i}{\partial\xi^\gamma} = 0. \qquad (24)$$

Since it is assumed that $\det\|\partial x^i/\partial\xi^\gamma\| \neq 0$ ($i = 1, 2$ and $\gamma = 1, 2$), the only solutions to the system of two homogeneous equations (24) for the bracketed quantities are zero solutions. This condition is equivalent to identically true equation (22) for $i = 1, 2$.

THE EQUATION FOR SMALL-AMPLITUDE VIBRATIONS OF CURVED-SURFACE DIAPHRAGMS

As dynamic variables characterizing small deviations of the diaphragm from its equilibrium, we will consider time-dependent displacements of the diaphragm along the normal to its surface $u(\xi^1, \xi^2, t)$. We assume that, under the equilibrium condition ($u = 0$), the diaphragm surface is specified parametrically by the function $\mathbf{x} = (\xi^1, \xi^2)$ that, generally, satisfies equation (21). Let $\mathbf{x}(\xi^1, \xi^2)$ be the coordinate of an arbitrary point on the surface of a diaphragm, which is in the equilibrium state. In this case, as a result of a diaphragm deformation (a small displacement along the normal), this point acquires the coordinate $\mathbf{x}' = \mathbf{x} + u\mathbf{n}$. The radius vector connecting two infinitely near points on the surface of the deformed diaphragm is defined by the expression $d\mathbf{x}' = d\mathbf{x} + ud\mathbf{n} + \mathbf{n}du$, and the squared distance between them is

$$dl'^2 = d\mathbf{x}'^2 = d\mathbf{x}^2 + 2ud\mathbf{x}d\mathbf{n} + u^2d\mathbf{n}^2 + du^2. \qquad (25)$$

Here, we use relations (7) and (8), from which it follows, in particular, that $\mathbf{n}d\mathbf{x} = 0$ and $\mathbf{n}d\mathbf{n} = 0$. According to equations (9) and (11),

$$\frac{\partial\mathbf{n}}{\partial\xi^\alpha} = -b_{\alpha\beta}G^{\beta\gamma}\mathbf{a}_\gamma$$

and, therefore,

$$d\mathbf{x}d\mathbf{n} = -b_{\alpha\beta}d\xi^\alpha d\xi^\beta, \quad d\mathbf{n}^2 = b_{\alpha\gamma}b_{\beta\sigma}G^{\gamma\sigma}d\xi^\alpha d\xi^\beta.$$

These relations allow us to determine the metric of the deformed-diaphragm surface $\tilde{G}_{\alpha\beta}$. Substituting them into (25) and taking into account (2) results in

$$dl'^2 = \tilde{G}_{\alpha\beta}d\xi^\alpha d\xi^\beta,$$

where

$$\tilde{G}_{\alpha\beta} = G_{\alpha\beta} - 2b_{\alpha\beta}u + b_{\alpha\gamma}b_{\beta\sigma}G^{\gamma\sigma}u^2 + \frac{\partial u}{\partial\xi^\alpha}\frac{\partial u}{\partial\xi^\beta}.$$

Here, $G_{\alpha\beta}$ and $b_{\alpha\beta}$ are the coefficients of the first and second quadratic forms of a nondeformed, i.e., equilibrium-state, diaphragm. The surface energy of a deformed diaphragm is determined by the equation

$$U = T\int\sqrt{\tilde{G}}d\xi^1 d\xi^2,$$

where $\tilde{G} = \det\|\tilde{G}_{\alpha\beta}\|$.

Assuming u to be small, let us develop $\sqrt{\tilde{G}}$ as a series in u and restrict our consideration to the terms that are quadratic with respect to displacement. As a result, we have

$$U \approx T\int\sqrt{G}d\xi^1 d\xi^2$$

$$+ \frac{1}{2}T\int\left(G^{\alpha\beta}\frac{\partial u}{\partial\xi^\alpha}\frac{\partial u}{\partial\xi^\beta} - 4Hu + Ru^2\right)\sqrt{G}d\xi^1 d\xi^2.$$

The first term on the right side of this expression determines the energy of a nondeformed diaphragm U_0 (formula (17)), the second, the energy resulting from its deformation; H and R are, respectively, the mean and scalar curvatures of a nondeformed diaphragm (formulas (15), (14), and (16)). In the absence of external static forces, when the diaphragm has a minimal-surface shape whose mean curvature is $H = 0$, this expression involves no term that is linear with respect to u.

Thus, as a result of deformation, the surface energy of the diaphragm changes by

$$\Delta U = U - U_0$$

$$= \frac{1}{2}T\int\left(G^{\alpha\beta}\frac{\partial u}{\partial\xi^\alpha}\frac{\partial u}{\partial\xi^\beta} - 4Hu + Ru^2\right)\sqrt{G}d\xi^1 d\xi^2.$$

If the equilibrium pressures on both sides of the diaphragm differ by Δp_0, then, upon deforming, the dia-

phragm does the work

$$\Delta W = \int \Delta p_0 u \sqrt{G} d\xi^1 d\xi^2.$$

Here, we use formula (20), in which we have assumed $\delta \mathbf{x} = \mathbf{n}u$. According to (21) and (23), the mean curvature H of the diaphragm and pressure difference Δp_0 are related by the expression (the Laplace formula [2])

$$\Delta p_0 = 2TH.$$

Therefore, the total change in the diaphragm energy caused by its deformation is

$$\Delta E = \Delta U + \Delta W$$
$$= \frac{1}{2} T \int \left(G^{\alpha\beta} \frac{\partial u}{\partial \xi^\alpha} \frac{\partial u}{\partial \xi^\beta} + Ru^2 \right) \sqrt{G} d\xi^1 d\xi^2. \qquad (26)$$

The absence of a term that is linear with respect to u in this expression is not accidental and reflects the fact that, even when external static forces are present, the minimum of the energy ΔE must correspond to the diaphragm equilibrium when $u = 0$.

When deriving the equation describing small-amplitude vibrations of a curved diaphragm, we will proceed from the principle of least action. As the action, we will consider the functional

$$S = \int\limits_{t_1}^{t_2} \iint \mathfrak{L} \sqrt{G} d\xi^1 d\xi^2 dt, \qquad (27)$$

where the Lagrangian is

$$\mathfrak{L} = \frac{1}{2} \rho \left(\frac{\partial u}{\partial t} \right)^2 - \frac{1}{2} T \left(G^{\alpha\beta} \frac{\partial u}{\partial \xi^\alpha} \frac{\partial u}{\partial \xi^\beta} + Ru^2 \right) - \Delta p u. \quad (28)$$

The first term on the right side of the expression defines the density of the kinetic energy of the diaphragm (ρ is the surface density of the diaphragm); the second term defines the density of its potential energy (formula (26)); the third term takes into account the effect of the external forces due to the difference of pressures on both sides of the diaphragm Δp, which is excessive with respect to the effect exerted on the equilibrium diaphragm Δp_0. Varying (27) with u and assuming $\delta S = 0$, we arrive at the following Euler–Lagrange equation, which describes small-amplitude vibrations of the curved diaphragm:

$$\frac{1}{\sqrt{G}} \frac{\partial}{\partial \xi^\alpha} \left(\sqrt{G} G^{\alpha\beta} \frac{\partial u}{\partial \xi^\beta} \right) - Ru - \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2} = \frac{\Delta p}{T}. \quad (29)$$

Here we introduce the notation $v^2 = T/\rho$ and, when deriving the equation itself, we take into account that, at the limits of integration, $\delta u = 0$. Equation (29) extends the known equation for small-amplitude vibrations of a plane diaphragm to an arbitrary shaped diaphragm. The metric-tensor components $G_{\alpha\beta}$ and scalar curvature R in this equation are determined by the equilibrium shape of the diaphragm (formulas (3), (13), and

(16)). The invariance of equation (29) with respect to arbitrary transformations of the parameters ξ^α is obvious, because R is a scalar and the left-side Laplace–Beltrami operator can be written in the explicitly covariant form

$$\frac{1}{\sqrt{G}} \frac{\partial}{\partial \xi^\alpha} \left(\sqrt{G} G^{\alpha\beta} \frac{\partial u}{\partial \xi^\beta} \right) = \nabla_\alpha \nabla^\alpha u.$$

It should be noted that the character of the solution of the wave equation obtained depends on the sign of R and, therefore, is different for diaphragms with positive (e.g., sphere) and negative (pseodosphere) curvature.

ENERGY DENSITY AND ENERGY-FLUX DENSITY OF THE FIELD OF DISPLACEMENTS: CONSERVATION LAW

In the absence of external dynamic forces ($\Delta p = 0$), the total energy of the vibrating diaphragm, i.e., the energy of the field of displacements $u(\xi^1, \xi^2, t)$, must not vary with time. In the field theory, the energy-conservation law follows immediately from the differential conservation law (equation of continuity) relating temporal variation of the field-energy density W to spatial variation of the energy-flux density S^α. This law is in turn the consequence of the invariance of action under infinitely small time shifts [3]. For the $u(\xi^1, \xi^2, t)$ field under consideration, which is specified on the surface of a curved diaphragm; i.e., in a 2D Riemannian space, it is appropriate to write it in a generally covariant form

$$\frac{\partial W}{\partial t} + \nabla_\alpha S^\alpha = 0$$

or, on expanding the covariant derivative, as

$$\frac{\partial W}{\partial t} + \frac{1}{\sqrt{G}} \frac{\partial}{\partial \xi^\alpha} (\sqrt{G} S^\alpha) = 0. \qquad (30)$$

If the field Lagrangian is specified, then the energy density and the energy-flux density can be determined using the known relations [3]

$$W = \frac{\partial \mathfrak{L}}{\partial (\partial u / \partial t)} \frac{\partial u}{\partial t} - \mathfrak{L}, \quad S^\alpha = \frac{\partial \mathfrak{L}}{\partial (\partial u / \partial \xi^\alpha)} \frac{\partial u}{\partial t}.$$

Substitution of the Lagrangian (28), in which $\Delta p = 0$, into these expressions gives

$$W = \frac{1}{2} \left[\rho \left(\frac{\partial u}{\partial t} \right)^2 + T \left(G^{\alpha\beta} \frac{\partial u}{\partial \xi^\alpha} \frac{\partial u}{\partial \xi^\beta} + Ru^2 \right) \right],$$

$$S^\alpha = -T G^{\alpha\beta} \frac{\partial u}{\partial \xi^\beta} \frac{\partial u}{\partial t}.$$

By direct testing, we ascertain that the obtained W and S^α satisfy the conservation law (30) on the condition that $u(\xi^1, \xi^2, t)$ is a solution to field equation (29) with a zero right side.

A SPHERICAL DIAPHRAGM

Consider a spherical diaphragm with the center of the sphere at the origin of coordinates and the radius of curvature R_0. As independent parameters ξ^1 and ξ^2, which determine the position of any point on the surface of this diaphragm, it is convenient to use spherical angles Θ and $\varphi(\xi^1 = \Theta, \xi^2 = \varphi)$. In this case, the surface of the diaphragm is specified by the parametric equations

$$x^1 = R_0 \cos\varphi \sin\Theta, \quad x^2 = R_0 \sin\varphi \sin\Theta,$$

$$x^3 = R_0 \cos\Theta.$$

Relations (3) and (13) allow one to determine the coefficients of the first $G_{\alpha\beta}$ and second $b_{\alpha\beta}$ quadratic forms of this surface. Direct calculations give the following results:

$$G_{\alpha\beta} = \begin{bmatrix} R_0^2 & 0 \\ 0 & R_0^2 \sin^2\Theta \end{bmatrix}, \quad b_{\alpha\beta} = \begin{bmatrix} -R_0 & 0 \\ 0 & -R_0 \sin^2\Theta \end{bmatrix}.$$

From these relations and formula (16), it follows that the scalar curvature of the spherical surface $R = 2/R_0^2$. In accordance with this, equation (29) describing small-amplitude vibrations of a spherical diaphragm assumes the form

$$\frac{1}{\sin\Theta}\frac{\partial}{\partial\Theta}\left(\sin\Theta\frac{\partial u}{\partial\Theta}\right) + \frac{1}{\sin^2\Theta}\frac{\partial^2 u}{\partial\varphi^2}$$

$$-2u - \frac{R_0^2}{v^2}\frac{\partial^2 u}{\partial t^2} = R_0^2\frac{\Delta p}{T}.$$

As one can see, this equation admits solutions corresponding to purely radial vibrations if Θ and φ boundary conditions are not specified and the diaphragm is a sphere.

A CATENOID-SHAPED DIAPHRAGM

Consider a diaphragm with the shape of a minimal surface of revolution, namely, that of a catenoid, whose axis is directed along the axis x^3 The availability of the axis of revolution allows one to use the cylindrical coordinates ρ and φ as parameters determining the positions of points on the surface of the diaphragm $(\xi^1 = \rho, \xi^2 = \varphi)$. In this case, the diaphragm surface is specified by the parametric equations

$$x^1 = \rho\cos\varphi, \quad x^2 = \rho\sin\varphi, \quad x^3 = a\,\text{arch}\frac{\rho}{a} + b,$$

which are the solutions to the equations of the minimal surface (19). The constants a and b are found from the boundary conditions, which fix the position and equilibrium shape of the diaphragm. For the coefficients of the first and second quadratic forms of the catenoid surface, we have, respectively,

$$G_{\alpha\beta} = \begin{bmatrix} \dfrac{\rho^2}{\rho^2 - a^2} & 0 \\ 0 & \rho^2 \end{bmatrix}, \quad b_{\alpha\beta} = \begin{bmatrix} -\dfrac{a}{\rho^2 - a^2} & 0 \\ 0 & a \end{bmatrix}.$$

The scalar value of the surface under consideration is negative and depends on ρ: $R = -2(a^2/\rho^4)$. Substitution of the obtained values into (29) results in the following equation describing the small-amplitude vibrations of the diaphragm:

$$\frac{\sqrt{r^2 - 1}}{r^2}\frac{\partial}{\partial r}\sqrt{(r^2 - 1)\frac{\partial u}{\partial r}} + \frac{1}{r^2}\frac{\partial^2 u}{\partial\varphi^2}$$

$$+ \frac{2}{r^4}u - \frac{a^2}{v^2}\frac{\partial^2 u}{\partial t^2} = a^2\frac{\Delta p}{T},$$

where $r = \rho/a$ is a dimensionless parameter $(1 \le r < +\infty)$.

REFERENCES

1. P. K. Rashevskiĭ, *Riemannian Geometry and Tensor Analysis* (Nauka, Moscow, 1967).
2. L. D. Landau and E. M. Lifshits, *Hydrodynamics* (Nauka, Moscow, 1986).
3. L. D. Landau and E. M. Lifshits, *Field Theory* (Nauka, Moscow, 1973).

Translated by I.P. Nikishin

Effect of Shock Compression of Solid Insulators on the Injection of Valence Electrons in Strong Magnetic Fields

Yu. N. Vershinin, D. S. Il'ichev, and P. A. Morozov

Institute of Electrophysics, Ural Division, Russian Academy of Sciences,
ul. Komsomol'skaya 34, Yekaterinburg, 620049 Russia
Received December 8, 1998

Abstract—A correlation was shown to exist between the velocity of impulsing discharge from the anode, shock-wave parameters, electric-field strength, elastic properties, and the probability of ionization in the valence band of insulating solids. Using NaCl and KCl crystals as examples, a quantitative estimation of these dependences was performed for the interval of discharge speeds of 5×10^3–10^6 m/s. © *2000 MAIK "Nauka/Interperiodica".*

It was indicated in [1, 2] that the process of electric breakdown in insulators can be affected by shock compression. The development of shock waves in this process was usually related to the arc stage of discharge, which follows the short-circuiting of the disruptive gap through the breakdown path. The last mechanism is known to be the basis for the spark-discharge (electroimpulsing) methods of boring and grinding of rocks [2].

At the same time, intense shock waves may also be generated at the stage of propagation of a discharge filament. This follows from the supersonic velocities v_{ad} of the impulsing anode discharge in condensed insulators [3], which are accompanied by the supersonic propagation of the front of the first-order insulator–plasma phase transition [4].

The development of plasma in the anode-discharge filament is a consequence of electron transformation from the valence-electron state into a quasi-free state. The process of their intense injection will occur in this case not only under the action of strong electric fields, but also under conditions of strong compression of the insulator by a shock wave generated by the front of the phase transition.

The above-mentioned features of pulsed anode discharge suggest the existence of a dependence between the discharge velocity v_{ad}, the parameters of the shock wave, the elastic properties of the insulator, and the probability of the electrostatic ionization $\omega = n/N$, s^{-1} (n is the number of electron–hole pairs formed per unit volume per unit time, and N is the number of valence electrons per unit volume). The possibility for such a relation to exist is, in particular, evidenced by the dependence of the radius r_0 of the anode discharge filament on not only the energy gap, but also the compressibility K of solid insulators [5].

Upon compression of crystalline insulators and semiconductors, the energy gap width $E_g(p) = E_g^*$ is known [6] to change as

$$E_g^* = E_g + a_p p \qquad (1)$$

depending on the magnitude and sense of the pressure coefficient a_p. In direct-transition substances, including alkali-halide crystals, the coefficient a_p is proportional to the compressibility K

$$a_p = -\gamma K \qquad (2)$$

and is negative. Upon isothermal compression in a range of $9 \times 10^{-12} \le K \le 200 \times 10^{-12}$, the coefficient γ changes in narrow limits of $-2.3 \times 10^{-19} < \gamma < -1.3 \times 10^{-19}$ J and has an average value of -2×10^{-19} J [7].

Upon shock-wave compression, the average value of γ increases to -3×10^{-19} J. The negative value of the pressure coefficient a_p indicates that the energy gap of the substance decreases upon compression. Correspondingly, it is that region of a direct-transition semiconductor that is subject to a shock wave that will be, under other conditions equal, the primary source of injected electrons. If the width of this region is l, then the degree of ionization of the insulator upon shock compression will be

$$x_e = \omega(l/v_{ad}). \qquad (3)$$

The investigation of a one-dimensional crystal-lattice model [8] showed that the width of a shock wave depends on its velocity and at large velocities approaches a constant value. The intensity of the shock wave in this case decreases by a factor of e on distances of about $\Delta l = (2-4)d_0$, where d_0 is the lattice parameter.

The probability of ionization ω and injection currents in solids were studied in numerous works, which were generalized, e.g., in monographs such as [9, 10].

For the case of tightly bound electrons, the corresponding equations were suggested in [11, 12].

On the assumption that these equations remain also valid under shock-compression conditions, we may write [12]

$$\omega = n/N = \frac{(eEd^*)^2}{2\pi\hbar E_g^*}\exp\left(-\frac{E_g^*}{eEd^*}\ln 1/\alpha\right). \quad (4)$$

Here, α is the ratio of the energy-gap width to the valence-band width; and E_g^* and d^* are the effective values of the energy gap and lattice parameter, respectively, upon shock compression. The values $d^* = d_0(p)$ can be calculated from experimental adiabats of concrete insulators [13, 14], which are usually approximated by expressions such as

$$p = a[(\rho^*/\rho_0)^b - 1], \quad (5)$$

from which we obtain

$$d^* = d_0\Big/\left(\frac{p}{a}+1\right)^{1/3b}. \quad (6)$$

In (5) and (6), d_0 and ρ_0 are the lattice parameter and the density of the unperturbed insulator, and a and b are empirical coefficients. Their values for some insulators for a pressure range of $10^{10} < p < 10^{11}$ Pa are given in Table 1.

In the model of an ionizing shock wave, the conductivity of the insulator in front of the leading edge of the wave is $\sigma = 0$. In this case, the field strength and the critical degree of ionization of the insulator at the phase surface may be assumed, to the first approximation, to be constant in the entire range of acting stresses and, correspondingly, velocities v_{ad} [5].

As was noted in [3], the velocity v_{ad} in crystalline insulators depends on two initial conditions of the discharge development: on the critical breakdown voltage U_0 at the time instant t_0 and on the rate of pulse rise $dU/dt(t_0)$. In such insulators, the minimum velocities v_{ad} usually exceed the longitudinal sound velocity in the corresponding direction by a factor of 1.25–1.3.

In alkali-halide crystals, the maximum velocities ($v_{ad} \geq (3–5) \times 10^5$ m/s) are achieved upon the action of nanosecond pulses with a front length τ of less than 5 ns and amplitude voltage $U \geq 150$ kV. No explicit dependence of v_{ad} on τ was observed in this case (see Fig. 1).
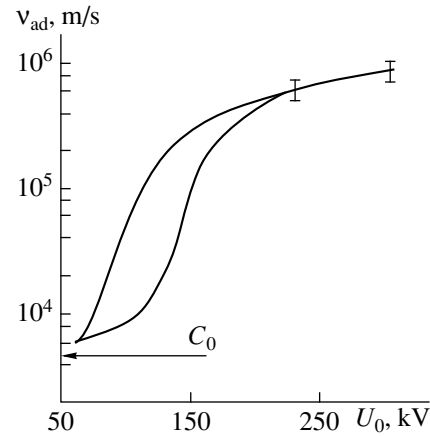

Fig. 1. Variation of the propagation velocity of filament of the pulsed anode discharge, v_{ad}, in NaCl. The lower generating line corresponds to the condition $dU/dt = 0$.

The use of the above relationships for the estimation of the interconnection of parameters that affect the process of ionization of a solid insulator is illustrated using NaCl and KCl crystal as examples (see Table 2).

In both cases, the disruptive distance was $h \geq 10^{-2}$ m; the width of the shock-compression region, $\Delta l = 3d_0$; and the coefficient of proportionality γ in (2), -3×10^{-19} J. The pressure p_{min} was determined in the detonation approximation as

$$p_{min} = \frac{\rho_0 v_{ad,\,min}^2}{k+1} \quad (7)$$

at constants $k = 3.83$ for NaCl and 3.55 for KCl calculated by the method described in [14]. Under these conditions, we have

$$\omega_m = \omega_{min}\frac{v_{ad,\,n}}{v_{ad,\,min}}, \quad (8)$$

where ω_{min} is the probability of ionization at $v_{ad,\,min}$, and ω_n is the probability of ionization at $v_{ad,\,n} > v_{ad,\,min}$.

The general appearance of the surfaces of the variation of the ionization probability ω, pressure p, and field strength E is given in logarithmic coordinates in Fig. 2 for NaCl as an example; the region of parameters corresponding to the adopted initial data is marked by a solid line in it.

Figure 3 displays the dependence of the pressure during shock compression on the velocity v_{ad} and related changes in the effective energy gap E_g^* at the

Table 1

Coefficients	LiF	NaCl	KCl	KBr	NaI	CsI
a, 10^9 Pa	11.73	4.31	1.819	1.710	5.245	4.554
b	5.082	4.993	5.475	5.282	4.033	4.015
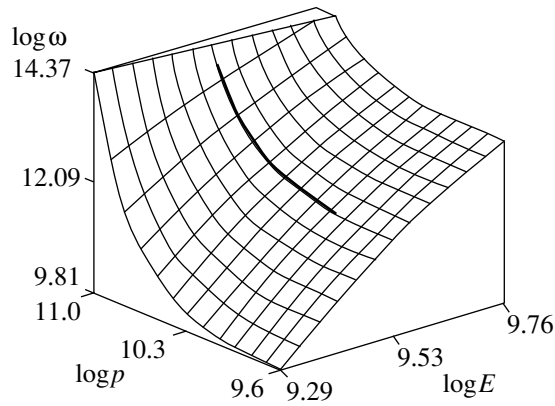
Fig. 2. The (ω, p, E) surface in logarithmic coordinates.

Fig. 3. Variation of E_g^* and p as functions of v_{ad}: (●) NaCl, and (+) KCl.

degrees of ionization $x_e(NaCl) = 0.256$ for NaCl and 0.252 for KCl. The results obtained do not contradict those previously reported. Thus, the quantum-mechanical estimates of the electrical conductivities of KI, CsCl, CsBr, and CsI crystals have shown [6] that at pressures $p = (1.6–2.7) \times 10^{10}$ Pa the effective energy gap E_g^* is equal to $(0.8–0.6)E_{g,0}$. From Fig. 3, it follows that in this pressure range we have $E_g^* = (0.82–0.64)E_{g,0}$.

Note that, at $v_{ad} = $ const, E_g^* is almost independent of the pressure p and the individual properties of the insulator. This effect needs a special investigation. On a qualitative level, it can be explained as follows.

As was noted in [3], an important parameter that determines the dynamic characteristics of the anode discharge in solid dielectrics is the compressibility of the crystal. Probably, the above velocity dependence of the effective energy gap $E_g^* (v_{ad})$ is also a consequence of a complex influence of the elastic properties of the insulator on the process of discharge propagation. Thus, the velocity v_{ad} at $U_0 = $ const, $dU/dt = $ const is directly proportional to the ratio $K/E_{g,0}$ [3]. It can easily be shown that, for the homologous series of alkali halide crystals, we have $E_{g,0} \sim K^{-1}$ [7], and at $v_{ad} = $ const, we obtain $p \sim K$ (Fig. 3). Then, the velocity v_{ad} is

Table 2

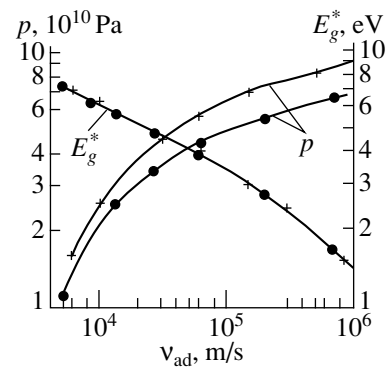| Parameters of the process	Units of measure	NaCl	KCl
Critical breakdown voltage U_0	kV	57–320	45–320
Discharge velocity v_{ad}	10^3 m/s	6–850	5–1150
Electric field strength E_0 [5]	10^9 V/m	3.34	2.61
Energy gap $E_{g,0}$	10^{-19} J	13.6	13.44
Compressibility K [7]	10^{-12} Pa	42.73	54.91
Pressure p_{min} at $v_{ad, min}$	10^{10} Pa	1.6	1.1

proportional to K^2, and expression (1) can be rewritten in the form

$$E_g^* \sim x v_{ad}^{-1/2} + y v_{ad}, \qquad (9)$$

where x and y are the coefficients of proportionality.

The adopted relation $v_{ad} \sim K^2$ does not contradict experimental data. Indeed, at $U_0 = 220$ kV and $dU/dt = 0$, we have $v_{ad} = (5.8–6.0) \times 10^5$ m/s for NaCl and $(9.0–9.4) \times 10^5$ m/s for KCl. The ratio of these velocities is equal to 1.55–1.62, and the ratio of the corresponding compressibilities is 1.65.

The above calculations indicate the necessity of taking into account the impact compressibility of solid insulators when analyzing pulsed breakdown of such materials. It should, however, be emphasized that these results have only an estimative, semiquantitative character. The reason is that a correct quantitative account of the compressibility effect on the process of electron injection in solid insulators is connected with the necessity of carrying out a whole number of special experimental and theoretical investigations. First of all, this refers to the study of pressure effects on the electron-band structure of wide-bandgap insulators.[1] This will permit us, in particular, to more rigorously determine the values of the pressure coefficients a_p. It may be expected that dependence (1) will be nonlinear in a wide range of pressures. We also should estimate the potential of equation (5) at high levels of injection with degrees of ionization close to unity. It cannot be ruled out that, under these conditions, these equations will have another structure and, at large velocities v_{ad} and corresponding pressures, the degree of ionization x_e and the charge numbers of the ions will be more than unity.

---

[1] A review of analogous work on semiconductors and narrow-bandgap insulators was given by H. Drickamer [6, ch. 12].

REFERENCES

1. V. A. Chuenkov, Izv. TPI **91**, 45 (1956).

2. A. A. Vorob'ev, G. A. Vorob'ev, E. K. Zavadovskaya, *et al.*, *Pulse Breakdown and Fracture of Insulators and Rocks* [in Russian] (Tomsk. Gos. Univ., Tomsk, 1971).

3. Yu. N. Vershinin, Zh. Tekh. Fiz. **59** (2), 158 (1989).

4. Yu. N. Vershinin, G. A. Mesyats, A. L. Mironov, *et al.,* in *Proceedings IEEE Int. Puls. Pow. Conference* (San-Diego, CA., 1991), p. 353.

5. Yu. N. Vershinin, Dokl. Ross. Akad. Nauk **347** (5), 614 (1996).

6. *Solids under Pressure*, Ed. by W. Paul and D. Warshauer (McGraw-Hill, New York, 1963).

7. *A Handbook of Physical Quantities* [in Russian], Ed. by I. I. Grigor'ev and E. I. Meĭlikhov (Energoatomizdat, Moscow, 1991).

8. R. Mabvi, G. E. Duvall, and S. C. Lowell, Int. J. Mech. Sci. **11**, 1 (1969).

9. K. Kao and W. Hwang, *Electron Transport in Solids* (Pergamon, Oxford, 1981).

10. M. A. Lampert and P. Mark, *Current Injection in Solids* (Academic, New York, 1970).

11. P. Feuer, Phys. Rev. **88** (1), 92 (1952).

12. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **33** (4/10), 994 (1957).

13. L. V. Al'tshuller, M. N. Pavlovskiĭ, L. V. Kuleshova, *et al.*, Fiz. Tverd. Tela (Leningrad) **5** (1), 279 (1963).

14. F. A. Baum, K. P. Stanyukevich, and B. I. Shekhter, *Physics of Explosion* (GIFML, Moscow, 1959).

Translated by S. Gorin

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

The Design of a Λ-HFS Magnetometer

A. K. Vershovskiĭ, A. S. Pazgalev, and E. B. Aleksandrov

Vavilov State Optical Institute, All-Russia Scientific Center,
Birzhevaya liniya 12, St. Petersburg, 199034 Russia
Received March 27, 1998

Abstract—The design of a new quantum magnetometer for geomagnetic fields, an outgrowth of the idea of the so-called HFS magnetometer that uses microwave transitions between hyperfine sublevels of alkaline atoms, is presented. In contrast to the ordinary HFS magnetometer, which measures the frequency difference of two independent transitions, the new Λ-HFS magnetometer is based on the excitation of two transitions with a common level. The design under discussion retains the principal advantages of the prototype (negligible systematic errors and the absence of dead zones for a single sensor), with the electronic circuits being simplified significantly and the requirements on the short-term frequency stability of the reference oscillator being relaxed substantially. An analysis of the new scheme is based on the solution of the equation for the density matrix of a multilevel system without constraints on the power of the microwave fields. © *2000 MAIK "Nauka/Interperiodica"*.

INTRODUCTION

The so-called HFS magnetometer [1, 2] proposed in the 1970s, which uses microwave transitions between magnetic sublevels of the ground state of an alkaline metal that belong to different hyperfine components, stands out among the many types of resonance, optically pumped magnetometer (OPM). Such a magnetometer (in its most perfect balanced modification) is remarkable for the absence of systematic errors at a level of random errors of the order of 1 pT. In addition, this is a unique instrument without dead zones for a single sensor. This circumstance makes the HFS magnetometer completely nonorientable, which is the most valuable property when used on movable platforms. However, the electronic part of the HFS magnetometer is much more complex than that of the conventional OPMs. The reason is that, first, a microwave frequency is measured (to within a few hundredths of a hertz) in this magnetometer rather than a radio frequency, as in ordinary magnetometers, which makes the requirements on the stability of the grid of reference frequencies three or four orders of magnitude more stringent. Second, the frequencies of two independent microwave transitions (whose subsequent subtraction removes the effect of light and collisional atomic-level shifts) in a balanced magnetometer are measured simultaneously. The presence of two automatic-frequency-control loops almost doubles the volume of the electronic circuits. The enhanced complexity of the electronic equipment has prevented the HFS magnetometer from coming onto the market for twenty years, and only now has the successful industrial development of two versions of a potassium instrument been reported.

In this paper, we discuss a new version of a balanced HFS magnetometer which retains the basic advantages of the prototype but has only one automatic-frequency-control loop and does not require an ultrahigh frequency stability of the microwave field. To this end, we propose to excite two transitions with a common upper level that form a Λ scheme with the inherent effect of coherent population trapping [3, 4] rather than two independent microwave transitions.

QUALITATIVE DESCRIPTIONS OF THE DESIGN

Figure 1 schematically shows the sublevels of the $n^2S_{1/2}$ ground state of an alkaline metal atom in an external magnetic field (the most common case of an
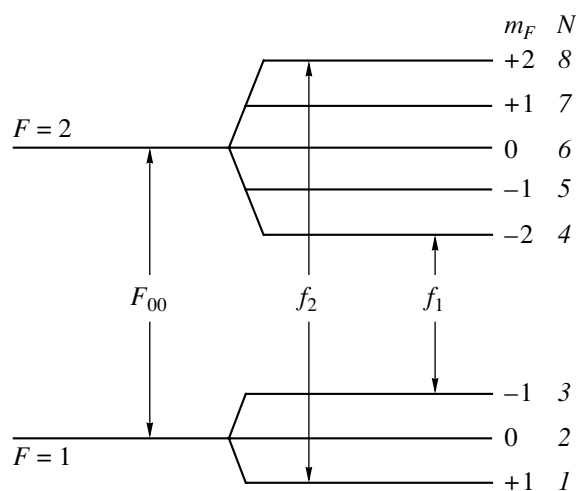

Fig. 1. Diagram of magnetic sublevels for an alkaline atom with a nuclear spin of 3/2: F_{00} is the frequency of hyperfine splitting; f_1 and f_2 are the frequencies of the magnetic transitions used in a conventional HFS magnetometer.

atom with a nuclear spin of 3/2 is presented). The two hyperfine $F = 1$ and $F = 2$ levels, which are separated by a hyperfine microwave interval in a zero field, are split, respectively, into three and five magnetic sublevels in a magnetic field. These sublevels are nearly equidistant in weak fields: the spectrum of the $\Delta F = 0$, $\Delta m_F = \pm 1$ transitions represents a tight group of lines whose frequencies are expressed by a series in powers of the magnetic induction H with the dominant linear term a identical for all lines (to within a few tenths of a percent) and equal to $\approx 7 \times 10^9$ Hz/T. The small nonlinear corrections to the frequencies of these transitions are proportional to the powers of the ratio $x = aH/f_{00}$, where f_{00} is the hyperfine splitting in a zero field. It is these radio lines that are used in conventional quantum magnetometers with optical pumping by circularly polarized light.

The $\Delta F = \pm 1$, $\Delta m_F = 0, \pm 1$ transitions between the magnetic sublevels belonging to different hyperfine states lie in the microwave range. Their frequencies in the linear approximation are $f_{00} + (m_F + m_F')aH$, forming nine lines with frequencies from $f_{00} + 3aH$ to $f_{00} - 3aH$; two of these lines with $f_{00} + aH$ and $f_{00} - aH$ are degenerate in pairs and differ in frequency only by small correction terms. Until recently, only the so-called 0–0-transition has been used in quantum electronics from this entire spectrum: $F = 2$, $m_F = 0 \longrightarrow F' = 1$, $m_F' = 0$. In the linear approximation, this transition does not depend on the magnetic field and, consequently, is of interest in frequency standardization: the existing atomic frequency standards and atomic clocks use such a transition in hydrogen, cesium, and rubidium atoms.

It is convenient to explain the essence of the design of the Λ-HFS magnetometer by using an idealized three-level scheme, in which levels 1 and 2 are associated with the $m_F = \pm 1$ sublevels of the lower ($F = 1$) hyperfine level, while level 3 is associated with the $m_F = 0$ sublevel of the upper ($F = 2$) hyperfine level. A sufficiently intense optical excitation of the upper hyperfine level is assumed, which provides a much faster relaxation of level 3 than the relaxation of levels 1 and 2. We therefore assume that, in the absence of microwave excitation, all atoms are concentrated at the lower hyperfine level and that the absorption of the pumping light is at a minimum. The separate excitation of the microwave 1 $\longrightarrow$ 3 or 2 $\longrightarrow$ 3 transitions is accompanied by the appearance of absorption in the pumping channel. In both cases, the change in the frequency of the applied field in the vicinity of the corresponding resonance allows us to obtain the resonance contour whose width is limited by the rate of the dominant optical relaxation caused by the optical excitation of level 3. However, the combined effect of two microwave fields changes radically the pattern, revealing the phenomenon of coherent population trapping or Λ resonance, where the difference of frequencies f_{13} and f_{23}
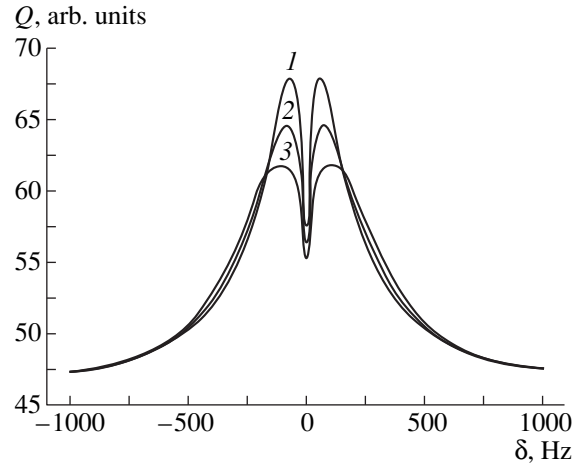

Fig. 2. Signal of absorption of the pumping light in an idealized Λ magnetometer for the scanning of the microwave-field frequency δ proportional to the measured magnetic field for three detunings of the microwave synthesizer relative to the frequency of hyperfine splitting F_{00}. The natural width of the lower levels γ is 20 Hz, and the width of the upper (common) level Γ is 200 Hz; Δ, Hz: (*1*) 0, (*2*) 100, (*3*) 150.

is equal to the frequency of the F_{12} transition between levels 1 and 2. Figure 2 shows the calculated behavior of the coefficient of absorption of the pumping light *versus* detuning δ from the exact Λ resonance, where $\delta = F_{12} - (f_{23} - f_{13})$, and three resonance curves that correspond to different detunings of the Δ microwave resonance, where Δ is equal to the difference between the frequency F_{13} of the 1 $\longrightarrow$ 3 transition and the frequency f_{13} of the applied field. All curves are characterized by a narrow dip centered at the frequency of the Δ resonance, i.e., at point δ = 0. The dip width is limited by the widths of levels 1 and 2. It is assumed to be equal for both levels and a fact of 10 smaller than the width of level 3 attributable to the optical excitation.

The curves in Fig. 2 were obtained from the results of an exact solution of the problem of two-frequency resonance in a three-level system. The analytical solution is too cumbersome to be reproduced. The expression for the coherence of sublevels 1 and 2 is given in [5]. We note only the most important features of the Λ resonance.

(1) Its width is abnormally small: the width of the resonance remains close to the natural width γ of levels 1 and 2, despite the microwave field which is strong enough to broaden the ordinary resonance 1 $\longrightarrow$ 3 or 2 $\longrightarrow$ 3 with a much larger width Γ (by a factor of 10 in Fig. 2).

(2) For a symmetric detuning δ of each of the microwave fields (i.e., when the condition of the Λ resonance holds), the resonance pattern changes only slightly as long as the detuning remains within the width of the upper (broad) level, i.e., for $\Delta \leq \Gamma$.
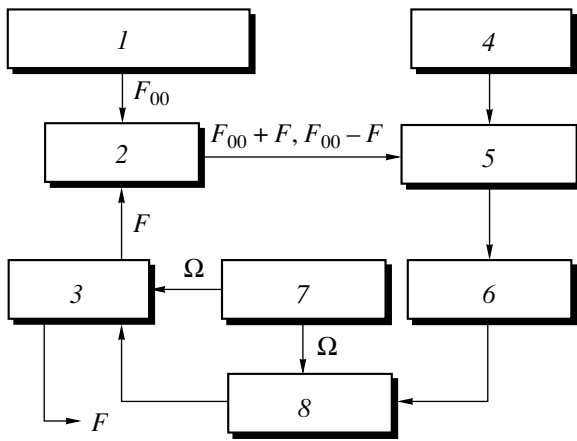
Fig. 3. Block diagram of the Λ-HFS magnetometer. The microwave field exciting the resonance in the sensor is formed at the output of the mixer from the signal of the microwave synthesizer with $F_{00} \approx 6834.6$ MHz and the signal of the radio-frequency oscillator with frequency F. The frequency F proportional to the measured magnetic field is fed to the frequency meter. (*1*) Microwave synthesizer, (*2*) mixer, (*3*) radio-frequency oscillator, (*4*) ⁸⁵Rb lamp, (*5*) ⁸⁷Rb sensor, (*6*) preamplifier, (*7*) low-frequency modulator, and (*8*) synchronous detector.

It is easy to see that the Λ resonance allows us to construct a Λ-HFS magnetometer possessing all the advantages of the prototype, with the requirements on the stability of the reference oscillator being relaxed significantly and with the design being simpler. A block diagram of the magnetometer is shown in Fig. 3. Appealing so far to an idealized scheme as before, we assume that the frequencies of the resonances 1 ⟶ 3 and 2 ⟶ 3 are linearly related to the magnetic induction, so the resonance frequencies are $f_{13} = f_{00} + aH$ and $f_{23} = f_{00} - aH$, respectively. In this case, the reference frequency f_0 is chosen to be f_{00}. The voltage of this frequency is subject to a balanced modulation by the signal of a controlled radio-frequency oscillator at frequency F, so two harmonics $f_{00} \pm F$ appear at the modulator output. By introducing an additional frequency modulation into the signal F at a low frequency Ω and simultaneously recording the intensity of the pumping light after the passage of the working cavity, it is easy to get the control signal of frequency F in order to hold it automatically at the center of the narrow dip. Measuring this frequency corresponds to measuring the magnetic induction.

By contrast to the well-known balanced HFS magnetometer, the modification we propose contains only one feedback loop and requires an appreciably lower stability of the reference oscillator, because the drifts of its frequency within the upper-level width do not affect significantly the characteristics of the Λ resonance. At the same time, the upper level can be artificially broadened through a high pumping rate, with this having no effect on the width of the narrow dip, i.e., on the magnetometer sensitivity.

MODELING AN ACTUAL SYSTEM

In contrast to an idealized three-level scheme, any actual alkaline atom introduces significant complications in the design associated with a more complex system of levels and with nonlinearity of their magnetic splitting. The former factor influences the intensity of calculated resonance signals, because the presence of the $F = 1$, $m_F = 0$ level unaffected by a microwave field in the ground state results in a considerable accumulation in this state of atoms that are not involved in the Λ resonance. Furthermore, allowance for additional levels shows that, apart from the Λ resonance $F' = 1$, $m'_F = -1 \rightleftarrows F = 2$, $m_F = 0 \rightleftarrows F' = 1$, $m'_F = 1$, we should reckon with the presence of the similar (in frequency) V resonance $F = 2$, $m_F = -1 \rightleftarrows F' = 1$, $m'_F = 0 \rightleftarrows F = 2$, $m_F = 1$.

The latter factor, the nonlinearity of magnetic splitting, leads to the fact that the reference frequency f_0, from which the frequencies f_1 and f_2 producing the Λ resonance are synthesized, is no longer equal to f_{00}, but slightly differs from this value by a relatively small correction that quadratically depends on the measured field induction. However, it is easy to cope with this difficulty. It will suffice to estimate the range of the half-sum of frequencies f_1 and f_2 in the working range of magnetic fields and to require that the light broadening of the upper hyperfine $F = 2$ level be larger than this parameter. In this case, the reference frequency can be chosen once and for all somewhere in the middle of the range.

At first glance, the effect of the V resonance is also insignificant, because it produces radio-frequency coherence of the sublevels of the broad upper hyperfine level intensively depopulated by light. However, since the frequency of this resonance is close to the Λ resonance, it, being broad, may result in an effective shift of the Λ-resonance center, i.e., in a systematic error. Therefore, this issue should be considered quantitatively.

Finally, because of the presence of the populated $F = 1$, $m_F = 0$ level, we should take into account the fact that the balanced modulator producing two harmonics, $f_0 \pm F$, from the reference frequency f_0 is not ideal. In fact, the f_0 harmonic, which is capable of causing a 0–0-transition will always be present at the modulator output to some extent. However, this may also turn out to be useful as a counterbalance to the above effect of accumulation of atoms at the $F = 1$, $m_F = 0$ sublevel.

We analyze the case of ⁸⁷Rb, for which the scheme of effective hyperfine pumping by the light of a ⁸⁵Rb lamp is known: as a result of the combination of isotopic and hyperfine shifts of resonance lines, the emission

of the ^{85}Rb lamp excites only the upper hyperfine level of the ^{87}Rb ground state.

The calculation was carried out primarily to find optimum conditions for obtaining a maximum steepness dS/df of the Λ resonance. Furthermore, it is necessary to find out how the frequency stability of the microwave synthesizer influences the position and steepness of the Λ resonance with allowance for the close proximity of the $2 \longrightarrow 7$ and $2 \longrightarrow 5$ resonances (see Fig. 1 for the level notation).

The problem is solved in the formalism of the density matrix for ^{87}Rb pumped by unpolarized light, which excites atoms from the $F = 2$ level. The calculation is carried out for a cell filled with a buffer gas, which suggests a complete mixing of the magnetic levels in the excited state and, accordingly, the same repopulation of all magnetic levels of the ground state in each pumping cycle. Optical pumping for the $F = 2$ ($k = 4, \dots, 8$) levels interacting with light is introduced through the additional relaxation of populations N_k caused by light,

$$\frac{d}{dt}N_k = -(\gamma + I_p)N_k + Q/8.$$

The term $Q = \gamma + \Sigma_k N_k I_p$ describes the level repopulation through the upper state and is proportional to the intensity of the pumping light I_p absorbed per unit time. The $F = 1$ ($i = 1, 2, 3$) levels do not interact with the light,

$$\frac{d}{dt}N_i = -\gamma N_i + Q/8.$$

Here, $2\pi\gamma$ is the inverse time of thermal population relaxation.

The interaction with the radio-frequency microwave fields V_{16} and V_{36}, the quasi-resonance transitions between the $1 \longrightarrow 6$, $2 \longrightarrow 7$, and $3 \longrightarrow 6$, $2 \longrightarrow 5$ levels, respectively, is specified in the rotating-field approximation

$$i\frac{d}{dt}\rho_{ik} = [\hat{H}, \hat{\rho}]_{ik}, \quad \hat{H} = \hat{H}_0 + \hat{V},$$

where $\hat{H}_0$ is the intrinsic Hamiltonian of the atom and $\hat{V}$ is the operator of interaction with the field.

The transition frequencies in the hyperfine structure depend on the induction H of a constant magnetic field and can be calculated using the Breit–Rabi formula. We use a expansion into a series to within the terms quadratic in H

$$F_{16} = F_{00} + aH + bH^2, \quad F_{36} = F_{00} - aH + bH^2,$$

$$F_{27} = F_{00} + (aH - a'H) + bH^2,$$

$$F_{25} = F_{00} - (aH - a'H) + bH^2,$$

$$F_{36} = F_{00} + b'H^2,$$

where $F_{00} = 6834.6$ MHz is the frequency of hyperfine splitting in a zero magnetic field [6], $a = 702.4 \times 10^7$ Hz/T, $a' = 2.8 \times 10^7$ Hz/T, $b = 502.5 \times 10^8$ Hz/T^2, and $b' = 574 \times 10^8$ Hz/T^2.

In the Earth's field ($H = 20$–80 μT), the transition frequencies F_{27} and F_{25} are close to the frequencies that form the Λ resonance and are shifted by $a'H \approx 28$ Hz/μT. The frequency of the 0–0 transition F_{36} in a nonzero constant magnetic field differs from the mean frequency $(F_{16} + F_{36})/2$ by $(b - b')H^2 \approx 71.5$ Hz/T^2.

We first assume that the microwave synthesizer generates a signal of the frequency $F_{HFS} = (F_{16} + F_{36})/2 = F_{00} + bH^2$ equal to the mean frequency of transitions F_{16} and F_{36} which form the Λ resonance. Using the method of balanced modulation by a signal of radio frequency $F = (F_{16} - F_{36})/2$, we obtain an output signal whose spectrum contains two frequencies, f_{16} and f_{36}. The detuning δ describes the scanning of the frequency $F \approx aH$ of the radio signal that modulates the synthesizer frequency F_{HFS} in an effort to detect the Λ-resonance signal. The frequency shifts of the microwave synthesizer with respect to the mean frequency of transitions F_{16} and F_{36} attributable to the frequency instability of the quartz oscillator are described by means of the detuning Δ. The appearance of detuning $\Delta \neq 0$ leads to the fact that the frequencies f_{16} and f_{36} turn out to be equally detuned with respect to their transitions.

The detunings obtained in the rotating-field approximation with respect to the frequencies of the corresponding transitions are the following: $\delta_{61} = \Delta + \delta$, $\delta_{63} = \Delta - \delta$, $\delta_{72} = \Delta + \delta + a'H$, $\delta_{52} = \Delta - \delta - a'H$. The complete equation for the density matrix includes the pumping, interaction with the microwave field, and relaxation

$$i\frac{d}{dt}\rho_{ik} = [\hat{H}, \rho]_{ik} - i(\hat{\Gamma}\rho)_{ik},$$

where the relaxation operator $(\hat{\Gamma}\rho)_{ik} = (1/2)(\Gamma_i + \Gamma_k)\rho_{ik}$ for $i \geq 4$, and $\Gamma_i = \gamma + I_p$ for $j = 1, 2, 3$.

We are interested in a steady-state solution, thereby reducing the problem to the solution of a set of 20 equations, seven of which describe the evolution of populations (levels 4 and 6 turn out to be equivalent), another 12 are responsible for six coherences, and one equation is for the quantity $Q(\delta)$, which is proportional to the signal observed in the light. The steady-state values of the density-matrix elements are determined for a given constant magnetic field H.

The most important parameter that determines the quality of the Λ resonance is the ratio $\Gamma/\gamma = (\gamma + I_p)/\gamma$, which relates the pumping rates and the rate of dark relaxation. In the calculations, we chose $I_p = 200$ Hz and $\gamma = 20$ Hz, which closely match the values attain-
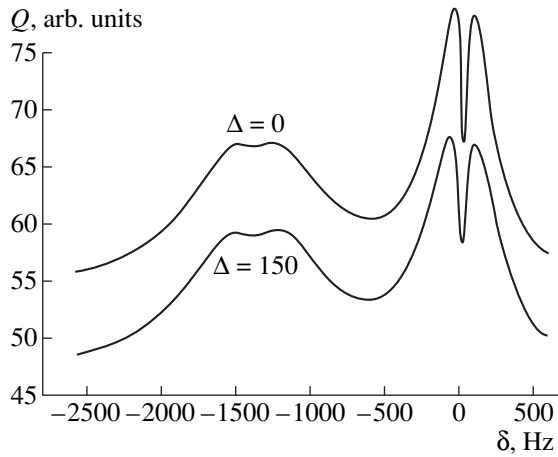
Fig. 4. Calculated signal of absorption of the light. The natural width Γ is 20 Hz, the magnetic sublevels of the $F = 2$ hyperfine state are additionally broadened by the pumping light with $I_p = 200$ Hz, and the Rabi frequency of the microwave field was chosen to be optimal in steepness, $V = 65$ Hz.

Fig. 5. Signal calculated for the case where an additional microwave field with the Rabi frequency $V_{26} = 65$ Hz, which destroys the V resonance, was introduced. The calculation was carried out with the same parameters as those for Fig. 4.

able in practice. The Rabi frequency of the radio-frequency field $V = V_{16} = V_{36}$ was varied in order to find a maximum of the steepness of the Λ resonance. We also studied the influence of detuning Δ, which describes the "quartz" drifts destroying the Λ resonance, on its steepness and shift. The calculation of the signal in a constant magnetic field of 50 μT showed that the optimum value is $V = V_{\text{opt}} \approx \sqrt{\Gamma\gamma}$. The dependence of the signal $Q(\delta)$ is shown in Fig. 4 both for the exact resonance $\Delta = 0$ and for the detuning $\Delta = 150$ Hz of the synthesizer frequency (which corresponds to the relative frequency instability $\Delta F_{HFS}/F_{HFS} = 3 \times 10^{-8}$ for ^{87}Rb). A shift in the peak of the Λ resonance by approximately 0.1 Hz caused by the proximity of the ordinary $2 \longrightarrow 5$ and $2 \longrightarrow 7$ transition is observed.

As the detuning Δ increases to 150 Hz, the shift increases to 0.12 Hz, while the steepness of the Λ resonance decreases by a factor of 1.65. Note that the steepness of an ordinary resonance (when one of the fields forming the Λ resonance is turned off) under optimum conditions is a factor of 12.5 smaller. The signal from two interfering spurious resonances is comparable in magnitude to the signal of the Λ resonance but has a considerably smaller steepness. However, the presence of a nearby superfluous resonance in automatic tracking and search systems may cause substantial inconveniences. When a detuning Δ comparable to the width Γ of the upper levels is introduced, the signals produced by the transitions f_{25} and f_{27} turn out to be permitted, which further complicates the overall picture. To suppress the signal from the transitions f_{25} and f_{27}, we propose to introduce an additional quasi-resonance microwave field with the frequency f_{26} corresponding to the transition between levels 2 and 6. As a first approximation, such a radio-frequency field depopulates level 2
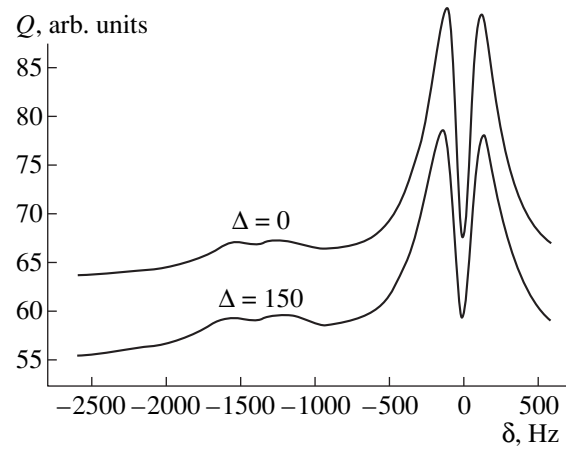
($F = 1$, $m_F = 0$), reducing the intensity of spurious resonances. In this case, the synthesis circuit does not become more complicated; we now simply do not require a complete balance from the device that generates the signals with frequencies F_{16} and F_{36}. By appropriately choosing the modulation index, we can vary the proportion between the spectral components F_{36}, F_{16}, and F_{26} over a wide range.

A more rigorous analysis shows that introducing an additional field V_{26} causes nonlinear resonances described by the coherences ρ_{12}, ρ_{23}, ρ_{56}, ρ_{67}, and even ρ_{15}, ρ_{17}, ρ_{35}, and ρ_{37}. To estimate the efficiency of this proposal, we carried out computer simulations by including the field $\hat{V}_{26}$ in the interaction operator. In this case, the total number of equations for the density matrix is 38, of which seven describe the populations, another 30 are responsible for the 15 coherences, and, finally, one equation gives the sought-for signal $Q(\delta)$. The results of our calculation of the dependence $Q(\delta)$ in a constant magnetic field of 50 μT for the same values of I_p, V, Δ, and γ as in Fig. 4 are shown in Fig. 5. Here, we again seek the maximum steepness of the Λ resonance, depending on the Rabi frequency of fields V_{16}, V_{36}, and V_{26}. It turns out that the maximum steepness is reached at $V_{26} \approx V_{16} \approx 65$ Hz. As the detuning Δ changes from 0 to 150 Hz, the shift in the peak of the Λ resonance changes from 0.138 to 1.142 Hz, while the steepness of the resonance decreases by a factor of 1.4.

In addition, we studied the form of the signal of the Λ resonance for constant magnetic fields of 20 and 80 μT, i.e., at the limits of the terrestrial range for similar variation scales of the detuning Δ. The following conclusions can be drawn from these calculations. In a weak terrestrial field of 20 μT, the noticeable proximity of the $2 \longrightarrow 5$ and $2 \longrightarrow 7$ transitions to the Λ reso-

nance results in a slightly larger shift and in its dependence on detuning Δ; the shift varies in the range 0.83–0.91 Hz. The steepness in this case essentially matches that for a constant field of 50 μT. In the strongest terrestrial magnetic field $H = 80$ μT, the variation in the shift is small, 0.02–0.023 Hz, while the steepness decreases by a factor of 1.75; this is explained by the elimination of the 0–0 resonance from the interaction because of the increase in its quadratic detuning.

CONCLUSION

We propose an essentially new version of a balanced magnetometer which uses hyperfine-state transitions (HFS magnetometer). It is based on a new principle with a nonlinear Λ resonance forming its basis. While retaining the unique properties inherent in this type of instrument, namely, the absence of dead zones (the sensor is nonorientable) and the suppression of light shifts, this version of a HFS magnetometer allows us to considerably simplify the electronic circuit of frequency synthesis by removing the requirement of active frequency stabilization of the reference quartz oscillator. To verify the fruitfulness of the idea, we performed computer simulations of the Λ resonance signal. The simulations revealed shifts in the resonance, depending on frequency drifts of the reference oscillator; however, these shifts are small, of the order of a few hundredths of a nanotesla. This type of instrument can be used both in magnetic search problems and in the metrology of the terrestrial-range magnetic field.

ACKNOWLEDGMENTS

REFERENCES

1. E. B. Aleksandrov and A. B. Mamyrin, Izmer. Tekh. **20** (7), 73 (1977).
2. E. B. Aleksandrov, A. B. Mamyrin, and N. N. Yakobson, Zh. Tekh. Fiz. **51**, 607 (1981).
3. W. E. Bell and H. L. Bloom, Phys. Rev. Lett. **29**, 1669 (1961).
4. E. Arimondo and E. Wolf, Progress in Optics **XXXV**, 257 (1996).
5. E. B. Aleksandrov, A. B. Mamyrin, and Yu. S. Chidson, Zh. Éksp. Teor. Fiz. **72**, 1569 (1977).
6. N. M. Pomerantsev, V. M. Ryzhkov, and G. V. Skrotskiĭ, *Physical Principles of Quantum Magnetometry* [in Russian] (Nauka, Moscow, 1972).

============ **EXPERIMENTAL INSTRUMENTS AND TECHNIQUES** ============

The Influence of a Cross-Linking Agent and C_{60} Fullerene on the Properties of a Solid Lubricant

B. M. Ginzburg and D. G. Tochil'nikov

Institute of Problems in Machine Science, Russian Academy of Sciences,
Vasil'evskiĭ Ostrov, Bol'shoĭ pr. 61, St. Petersburg, 199178 Russia
Received April 7, 1998

Abstract—The influence of a cross-linking agent, diamine, and C_{60} fullerene on the antifriction and wear properties of a solid lubricant made of trifluorochloroethylene–vinylidene fluoride copolymer was studied for steel-to-steel sliding friction. The wear characteristics are improved in the whole range of loads investigated, while the antifriction properties, only at small loads. A qualitative wear test method is proposed in which the test-period-averaged friction coefficient of a hybrid specimen (coating plus metal substrate) is measured with a standard friction machine. A model that considers the combined interaction of the substrate and the coating with the roller was used to calculate the linear wear rate of steel and the probabilistic parameter of wear. © 2000 MAIK "Nauka/Interperiodica".

Trifluorochloroethylene (TFCE)–vinylidene fluoride (VIF) copolymer is widely used in antifriction and wear-resistant coatings [1]. It is known that intermolecular cross-linking of polymer coatings enhances their wear resistance [2] and the addition of C_{60} fullerene to lubricating oils improves not only the wear resistance of friction couples but also the antifriction characteristics of triboassemblies [3–6]. In this work, we studied the combined effect of diamine as a cross-linking agent and C_{60} fullerene on the wear and antifriction properties of the TFCE–VIF copolymer used as a solid lubricant.

MATERIALS AND TEST TECHNIQUES

The materials used were F-32 LON industrial lacquer (NPO Okhtinskiĭ Khimkombinat, St. Petersburg), which is a 3% solution of the copolymer (with a molecular weight of 5×10^4 and a TFCE-to-VIF chain link ratio of 83 : 17) in ethyl acetate; reagent-grade 3,3'-diaminodipropylamine $NH(CH_2CH_2CH_2NH_2)_2$; and C_{60} fullerene obtained by extraction from fullerene soot [7] produced in an arc discharge [8]. In the final product, the fraction of C_{60} extracted by preparative chromatography was 96–98% [7].

Diamine was dissolved in the lacquer directly, and the fullerene was dissolved in toluene, which was then mixed with the lacquer. The resulting solution was applied on a 0.5-mm-thick St-30 steel sheet and dried in air at 190°C for 80 min with subsequent slow cooling (1°C/min). The amount of the solution was taken such that the coatings were ~70 μm thick. The coatings were prepared of the pure copolymer, a mixture of the copolymer with diamine (at a copolymer-to-diamine weight ratio of ~100 : 5), and a mixture of the copolymer with

diamine and C_{60} (copolymer : diamine : C_{60} ~ 100 : 5 : 1).

The studies were carried out with a 2070 SMT-1 standard roller friction machine. A test specimen was a cylindrical steel roller 50 mm in diameter tightly wrapped by a ~0.5-mm-thick St-30 steel sheet. The outer surface of the sheet was coated (or uncoated for comparison) by the lubricant (Fig. 1). Another roller, 46 mm in diameter and 16 mm wide, made of wear-resistant 18Kh2NChMA steel (State Standard 4543-71) rotated with an angular velocity $\omega = 400$ min^{-1}, which corresponds to the linear sliding velocity of 1 m/s. The tests were performed under dry friction conditions.
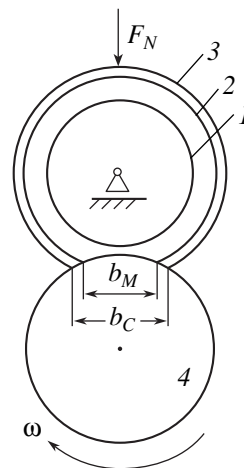
Fig. 1. Friction assembly employed in the tests: (*1*) fixed roller, (*2*) steel sheet, (*3*) coating, and (*4*) rotating roller; b_M and b_C are the widths of the wear grooves, used to determine the groove areas and the volumetric wear.

Normal loads applied to the specimens were $F_N = 10, 40, 80,$ and 120 N. At each of the loads, three or four tests for 300 s were performed. With this test duration, the coatings were completely worn away and a wear spot appeared on the steel substrate for all of the loads. The dependence of the moment of friction M_{fr} on time t was recorded in each test. The carriage of the friction machine lifted every 300 s, and the geometric parameters of the wear spot were determined with a measuring magnifier. A set of friction and wear characteristics of the coating and the substrate was calculated from these experimental data. In what follows, the subscripts C and M will refer to the coating and substrate parameters, respectively.

RESULTS AND DISCUSSION

An idealized dependence $M_{fr}(t)$ is shown in Fig. 2a. Initially, within fractions of a second (this time can be neglected), the friction moment sharply increases and then stabilizes at a level determined by the friction coefficient of the coating. It remains virtually constant, growing only slightly for a time t_C when the coating wears out. Next, coating wear gives way to substrate wear during some transition time t_{tr} when the friction moment noticeably increases. Finally, substrate wear takes place for a time t_M; the friction moment continues to increase but more slowly than in the transition regime. Over the period $(t_{tr} + t_M)$, the coating acts as a solid lubricant. Thus, in principle, the friction and wear characteristics of the coating and the metal substrate can be determined separately.

The initial value f_i of the friction coefficient was determined within the first 1–3 s of the tests. The coatings wore out in 10–130 s (depending on the coating type and the load applied); thus, f_i characterizes the frictional properties of the coatings (Fig. 3). As usual, f_i decreases at elevated loads but remains nearly unchanged as scoring is approached in the load range investigated. All of the coatings significantly reduce the steel-to-steel dry friction coefficient (curve *1*), but their antifrictional properties differ only slightly (curves *2–4*).

The wear properties of the coatings were difficult to determine, because the idealized dependence, shown in Fig. 2a, is observed only at small loads for wear-resistant coatings; actually, an $M_{fr}(t)$ dependence is similar to that presented in Fig. 2b.

The time instant when the metal surfaces come into contact was hard to reveal with certainty. First, the area S_M of the metal–metal contact is always much less than the total area of the friction contact (see table); second, once the metal surfaces have come in contact, the coating begins to act as a solid lubricant: its thin layer is continually applied to the contact surface by the rotating roller. Also, in most cases, the wear time of the coatings was very short (different for the different coatings) and could be measured with a great scatter.
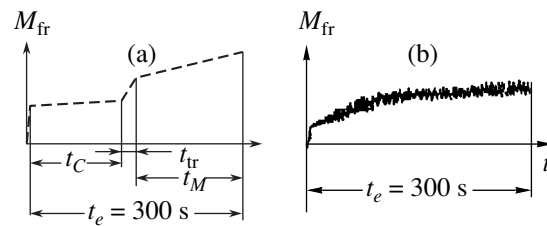
Fig. 2. $M_{fr}(t)$ dependences: (a) idealized curve and (b) curve typical of most of the cases studied.
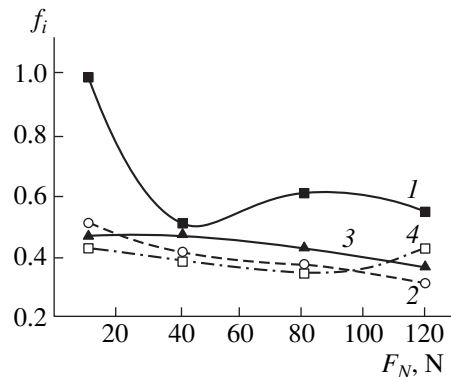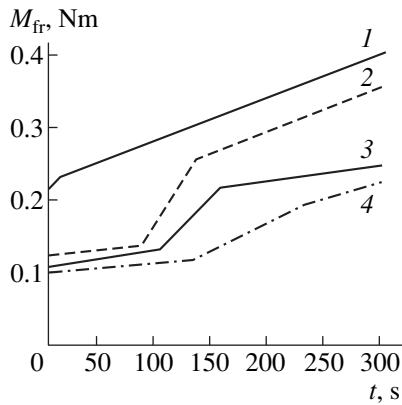

Fig. 3. Dependences of the initial friction coefficient f_i on the load applied to the friction assembly for the contact of steel with the (*1*) bare steel, (*2*) pure copolymer coating on the steel substrate, (*3*) copolymer + diamine coating, and (*4*) copolymer + diamine + C_{60} coating.

Therefore, the wear resistance of the coatings can be characterized by the wear time only for small loads; for high loads, such an approach fails. Consider first the results obtained at small loads.

The $M_{fr}(t)$ curves for different specimens at $F_N = 10$ N are given in Fig. 4. Each is obtained by averaging over the test results from three or four specimens and is also somewhat idealized, because noise is not shown. While the wear time for the coating from the pure copolymer equals 80 s, it increases to 100 and 130 s for the coating with diamine and that with diamine and C_{60}, respectively. This means that the wear resistance of the coatings rises in the same sequence. The value of M_{fr} at short times is nearly the same for all the three types of the coatings; i.e., their antifriction properties differ insignificantly (Fig. 3).

The transition time also increases in the same sequence (from 50 to 55 and 90 s, respectively), most noticeably for the coating with C_{60}. This coating is characterized by the smallest rate of increase M_{fr} and the least value of M_{fr} throughout the transition period. Hence, being used as a solid lubricant, this coating offers the best antifriction and wear properties.

On the completion of the transition stage, the values of M_{fr} remain the least for the coatings with C_{60}, and the coatings apparently continue to act as a solid lubricant.

Fig. 4. $M_{fr}(t)$ curves for friction of steel with the (*1*) bare steel, (*2*) pure copolymer coating on the steel substrate, (*3*) copolymer + diamine coating, and (*4*) copolymer + diamine + C_{60} coating.
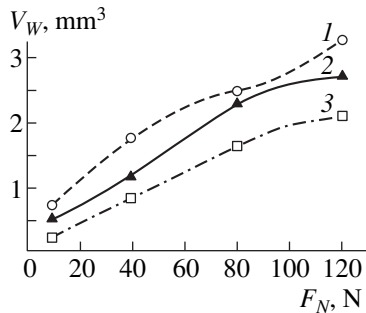
Fig. 5. Dependences of the volumetric wear of the "hybrid" (coating plus metal) specimens on the load applied to the friction assembly: (*1*) pure copolymer coating on the steel substrate, (*2*) copolymer + diamine coating, and (*3*) copolymer + diamine + C_{60} coating.

This follows from the fact that M_{fr} is still much lower than in the case of steel-to-steel dry friction (cf. curve *1* in Fig. 4).

Thus, at small loads, the coatings with diamine and C_{60} offer improved antifriction and wear properties. As determined by the wear time, the antiwear properties of these coatings are 30% higher than those of the coatings with diamine only. As solid lubricants, they have both better antifriction characteristics (the friction moment is lower by 30–50% in the transition regime) and better wear parameters (the transition time as increased by 65%).

In the range of high loads, a number of methodical difficulties appear; therefore, we examined the wear properties of "hybrid" (i.e., coating plus metal) specimens, supposing that the presence of the different coatings would markedly change these properties. From the load dependence of the total volumetric wear V_W of the hybrid specimens (Fig. 5), one can distinctly see the varying effect of the coatings: the addition of diamine and C_{60} leads to a reduction of V_W by 30–50%.

Although the volumetric wear of the substrate is much smaller than that of the coating, the measured values of V_W cannot characterize the coating wear alone, since it is greatly affected by the much higher wear resistance of the substrate.

To qualitatively characterize the wear resistance of the coating, we also used the following method. The friction coefficient f_{av} derived from the friction moment averaged over the test duration was plotted as a function of load (or pressure). The higher the wear resistance of the coating, the more its wear time, the longer acts its relatively low friction coefficient against the steel, and the less is f_{av}. As the load grows, the wear time of the coating decreases and f_{av} increases. This increase competes with the general trend of the friction coefficient to drop with load (Fig. 6, curves *1* and *2*). Thus, the $f_{av}(F_N)$ curves peak (Fig. 6, curves *3* and *4*). The existence of the maximum and its position depend on the test duration, range of applied loads, and wear resistance of the coating. Figure 6 shows that, under our test conditions, the curve obtained for the pure copolymer coating does not exhibit any maximum, the curve for the copolymer with diamine has a weak gently sloping maximum, and the curve for the copolymer with diamine and C_{60} has a distinct maximum. To obtain a maximum in the curve characterizing the pure copolymer coating, either very small loads had to be used (unachievable with our friction machine) or the test duration had to be cut to approximately 50–100 s.

Once the coating has been worn out and the metal surfaces have come in contact, the coating acts as a solid lubricant. From the behavior of the $f_{av}(F_N)$ curves at high loads (see Fig. 6), one can conclude that the antifriction properties of all of the coatings are very similar, but their wear characteristics again differ. It was of interest to see how the solid lubricant influences the wear characteristics of the steel, such as the linear wear rate I_h and the recently introduced parameter Q [9]. The latter is the probability that a contact spot of a friction couple will turn into a wear particle when the counterfaces are shifted by the mean diameter of the contact spot. This parameter characterizes the optimization of a friction surface in terms of wear reduction.

However, the determination of I_h is of no sense and that of Q is impossible if the pressure immediately at the metal–metal contact is unknown. Once a wear spot has appeared in the middle of the specimen, the pressure is gradually redistributed between the coating and the substrate as the spot grows. Using Hooke's law, one can write the following relationships for the substrate and the coating:

$$F_M = \varepsilon_0 E_M S_M, \tag{1}$$

$$F_C = \varepsilon_0 E_C S_C, \tag{2}$$

where F_M and F_C are the normal loads at the metal–metal and metal–coating frictional contacts, respectively; ε_0 is the total strain of the coating and the sub-
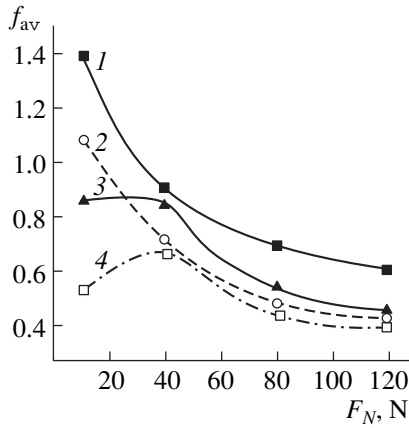
Fig. 6. Dependences of the averaged friction coefficient f_{av} on the load applied to the friction assembly for the contact of steel with the (*1*) bare steel, (*2*) pure copolymer coating on the steel substrate, (*3*) copolymer + diamine coating, and (*4*) copolymer + diamine + C_{60} coating.
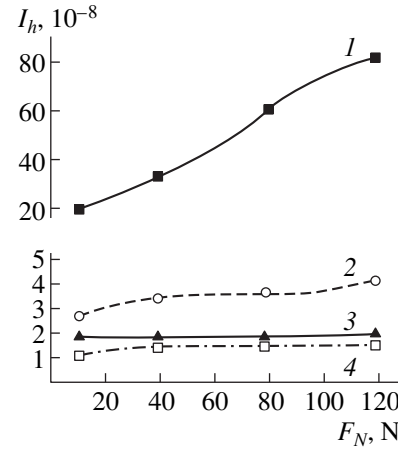
Fig. 7. Dependences of the linear wear rate I_h of the steel on the load applied to the (*1*) uncoated steel–steel contact and to the contact of the steel with (*2*) pure copolymer, (*3*) copolymer with diamine, and (*4*) copolymer with diamine and C_{60}.

strate; E_M and E_C are Young's moduli; and S_M and S_C are the wear spot areas on the metal and coating surfaces, respectively.

Normalizing the sum of loads and the sum of wear spot areas to unity, one can write equations (1) and (2) in the form

$$\varphi_M = \varepsilon_0 E_M \chi_M, \qquad (3)$$

$$\varphi_C = \varepsilon_0 E_C \chi_C, \qquad (4)$$

where φ_M and φ_C are the shares of the total force and χ_M and χ_C are the shares of the total area of the wear spots. Then, equation (4) can be written as

$$(1 - \varphi_M) = \varepsilon_0 E_C (1 - \chi_M). \qquad (5)$$

Dividing (5) by (3), one obtains

$$(1 - \varphi_M)/\varphi_M = (E_C/E_M)(1/\chi_M - 1). \qquad (6)$$

Putting $E_C \approx 600$ MPa [10] and $E_M \approx 2 \times 10^5$ MPa [11], one can easily derive a relationship between the forces F_M and F_C at any time instant of the test, having experimentally determined associated wear spot areas. For an ideal contact between the substrate and the roller and an ideally uniform coating thickness, the wear spot areas on the substrate and the coating can be uniquely related from purely geometric considerations (Fig. 1). However, because of the contact nonideality and a specimen-to-specimen scatter in coating thickness, we used the experimentally found values, which are listed in the table. At the final stage of the tests, most of the load (90–98%) always appears to be applied to the substrate.

The linear wear rate for a linear initial contact was defined as $I_h = 2h/3L$ [12], where h is the depth of a wear groove and L is the friction path over the test duration. The latter was determined as $L = \pi D \omega t$, where D

is the diameter of the rotating roller and t was taken to be equal to $(t_{tr} + t_M)$ when it could be measured or 300 s otherwise.

The load dependence of I_h for the metal–metal contact is given in Fig. 7. The linear wear rate for steel-to-steel dry friction is shown for comparison. The addition of diamine and fullerene into the coating significantly decreases I_h compared to the pure copolymer coating. As for the coatings with diamine only and with diamine and fullerene, one can see that C_{60} lowers I_h by 35% at very small loads and by ~15% for the others.

The value of Q was calculated from the relationship [9]

$$I_h = 3(P_N Q/H_B)(g/d), \qquad (7)$$

where $P_N = F_M/S_M$ is the contact pressure, H_B is the Brinell hardness, and g/d is the ratio of the mean height of a wear particle either to the mean diameter of a contact spot or (if the spot turned into the wear particle) to the diameter of the particle. It was assumed that $g/d \approx 1$, i.e., that wear particles are lumpy, which is often observed in the case of metals. The value of H_B for St-30 steel equals 1790 MPa [11].

Q vs. load curves for the metal–metal contact run similarly to the $I_h(F_N)$ ones. Here, too, the addition of fullerene C_{60} into the lubricant improves the wear properties by ~20%. The pressure dependences of I_h and Q_M for the metal–metal contact that were calculated within the combined interaction model are presented in the table. These data also indicate that the wear properties of the coatings are improved with the addition of diamine and C_{60}. The reduction of Q can be viewed as the optimization of the wear properties of the surface. The values of Q decrease nearly 1000 times as compared to those typical of steel-to-steel dry friction. For

The influence of load and coating type on some parameters

Friction couple	F_N, N	φ_M	F_M, N	S_M, mm^2	S_C, mm^2	χ_M	P_M, MPa	I_h, 10^{-8}	Q_M, 10^{-7}
Steel–steel	10	1	10	18.5	–	1	0.6	19	1900
	40	1	40	23.8	–	1	1.9	33	1000
	80	1	80	32.6	–	1	3.1	62	1200
	120	1	120	37.8	–	1	4.0	81	1210
(F-32 LON)–steel	10	0.95	9.5	0.7	11.5	0.054	13.6	2.5	11
	40	0.97	38.8	2.1	23.7	0.081	18.5	3.3	11
	80	0.98	78.4	3.1	26.2	0.106	25.2	3.6	8
	120	0.98	117.6	5.4	31.2	0.148	21.8	4.1	11
(F-32 LON with amine)–steel	10	0.95	9.5	0.6	9.7	0.058	15.8	1.8	6.7
	40	0.97	38.8	1.4	16.7	0.077	27.7	1.8	3.8
	80	0.97	77.6	2.4	26.3	0.084	32.3	1.8	3.3
	120	0.98	117.6	3.7	27.7	0.118	31.8	1.9	3.6
(F-32 LON with amine and C_{60})–steel	10	0.89	8.9	0.3	12.5	0.024	29.7	1.1	2.1
	40	0.96	38.4	1.5	18.6	0.072	26.5	1.5	3.3
	80	0.97	77.6	2.1	24.0	0.080	36.7	1.6	2.5
	120	0.98	117.6	3.3	25.9	0.113	35.6	1.7	2.9

Note: φ_M is the total load share on the metal contact, F_M is the load applied to the metal contact, S_M is the area of the wear spot on the substrate, S_C is the area of the wear spot on the coating, χ_M is the substrate share of the wear spot area, P_M is the pressure acting on the metal contact, I_h is the linear wear rate of the substrate at the metal contact, and Q_M is the stochastic substrate parameter as applied to the metal contact.

the coatings with diamine and diamine plus C_{60}, the values of Q, $\sim 2 \times 10^{-7}$ and $\sim 1.6 \times 10^{-7}$, respectively, are even slightly better (all other things being the same) than for steel-to-steel friction lubricated by a liquid industrial oil, $\sim 2.5 \times 10^{-7}$ [13]. However, they are still 1000 times higher than the reference value of $Q \sim 2 \times 10^{-10}$, achieved for lubricated babbitt-to-steel friction [9].

Considering the high price of C_{60} fullerene, its addition to the diamine-containing coatings appears to be efficient only in steel-to-steel dry friction units working at low pressures. The use of much cheaper fullerene soot may result in the same results, because its effect on liquid lubricants is nearly identical to that of pure C_{60} fullerene [4].

REFERENCES

1. Yu. A. Panshin, S. G. Malkevich, and Ts. S. Dunaevskaya, *Fluoroplastics* [in Russian] (Khimiya, Leningrad, 1978).
2. V. K. Kryzhanovskiĭ and O. V. Konova, Trenie i Iznos **14**, 322 (1993).
3. B. M. Ginzburg *et al.*, Pis'ma Zh. Tekh. Fiz. **21** (22), 62 (1995).
4. B. M. Ginzburg *et al.*, Pis'ma Zh. Tekh. Fiz. **21** (23), 35 (1995).
5. B. M. Ginzburg, D. G. Tochil'nikov, and V. P. Bulatov, Trenie i Iznos **18** (2), 235 (1997).
6. B. M. Ginzburg *et al.*, Trenie i Iznos **18** (4), 523 (1997).
7. *Russian Research and Development Program "Fullerenes and Atomic Clusters:" Project Index* (Fond Intellektual'nogo Sotrudnichestva, St. Petersburg, 1994), issue 1, p. 13.
8. D. V. Afanas'yev *et al.*, in *The 2nd Int. Workshop in Russia "Fullerenes and Atomic Clusters:" Abstracts of Invited Lectures and Contributed Papers*, St. Petersburg, Russia (1995), p. 49.
9. Yu. P. Kozyrev and B. M. Ginzburg, Zh. Tekh. Fiz. **68** (4), 48 (1998) [Tech. Phys. **43** (4), 392 (1998)].
10. *Use of Polyolephynes, Polystyrenes, Fluoroplastics, and Polyvinylacetate Plastics: A Catalogue* (NIITÉKhIM, Cherkassy, 1981).
11. *Mechanical Engineering: Reference Encyclopedia* (Mashgiz, Moscow, 1948).
12. Yu. P. Kozyrev *et al.*, Wear **171**, 71 (1994).
13. D. G. Tochil'nikov and B. M. Ginzburg, Zh. Tekh. Fiz. **68** (6), 102 (1999) [Tech. Phys. **44** (6), 700 (1999)].

Translated by M.L. Skorikov

EXPERIMENTAL INSTRUMENTS AND TECHNIQUES

Tunnel Spectroscopic Measurements in Air with a Scanning Tunnel Microscope

S. Yu. Vasil'ev and A. V. Denisov
Moscow State University, Vorob'evy gory, Moscow, 119899 Russia
Received July 1, 1998; in final form, July 26, 1999

Abstract—Central methodical and hardware problems in taking spectra of different types with a scanning tunneling microscope in air are briefly discussed. Ways to overcome these problems are considered. A procedure for recording a new type of tunneling spectra, namely, voltage–height (U–H) curves, which allow the estimation of the probe–specimen gap, is offered. It is suggested that the current flowing in the course of measurements in air is of electrochemical nature. Different tunnel spectra for metal, oxide, carbon, and polymer materials, including heterogeneous substances, are presented to illustrate the experiments. © 2000 MAIK "Nauka/Interperiodica".

INTRODUCTION

In the last fifteen years, scanning tunnel microscopes (STMs) and related probe nanoscopes have become available laboratory instruments and are applied for the most part for studying surface topography with a high resolution [1]. STM-based tunnel spectroscopic instruments [2–4] with a more complex hardware have found limited application. Their potentialities for fundamental physicochemical surface investigations are not fully exploited because of data processing problems and methodical difficulties in obtaining reliable and reproducible results.

This article considers hardware and software implementations of several kinds of tunnel spectroscopic measurements, as well as concerns the optimum modes of topographical investigations and image reliability control. These issues are extremely essential, e.g., for STM investigation of solid surface morphology in air at room temperature, when condensed films and/or adsorbed monolayers (contamination, moisture, etc.) cover the surface in the tunnel gap. For brevity, these films will be referred to as condensates.

METHODS OF RECORDING TUNNEL SPECTRA

A tunnel spectrum, uniquely characterizing the properties of a particular sample–gap–probe system, represents, generally speaking, a three-dimensional current–voltage–height (I–U–H) diagram. As a rule, two-dimensional (I–U and I–H) sections of these spectra, i.e., tunnel voltage–current curves (VCCs) and tunnel current–height (or rise) curves, are taken in experiments. The determination of the absolute value of a tunnel gap is a challenging task and calls for special techniques. It is obvious that the equilibrium position of the probe for given I and U is defined by the electronic state of the sample surface and thus is material-dependent. Moreover, the position of the sample–air interface is a matter of convention because of a complicated spatial distribution of the electronic density. Thus, in comparing the height dependences, one should take into account uncertainties in experimental values of H.

Voltage–Current Curves [I = f(U), H = const]

The control circuit of a conventional tunnel microscope keeps the gap current constant by varying the probe–sample distance [4]. In taking the VCCs, when the current changes, the electronic system of the microscope will inevitably alter the value of H, which distorts the results of measurements. We can fix H by disconnecting the negative feedback loop of the microscope or substantially narrowing its bandwidth, thus lowering the speed of system response to external disturbances. The complete disconnection of the negative feedback will bring the system into the state of unstable equilibrium, and the probe height may uncontrollably vary. Transients in the microscope's electronic circuits arising at the instant of feedback disconnection have an additional destabilizing effect.

Thus, the narrowing of the feedback bandwidth is the only reasonable way to provide the H = const mode. The maximum achievable operating time of the feedback circuit is several seconds. One should take into account that the displacement of the probe begins much earlier than the circuit begins to respond; thus, the duration of the nonequilibrium state must be within a fraction of a second. On the other hand, at fast measurements, the contribution of the capacitive component to the current being measured increases. This component depends on both the capacitance of the probe–sample system and the uncompensated input capacitance of the measuring system. In the case of 1-pF capacitance and
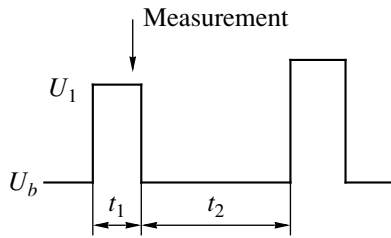
Fig. 1. Variation of the probe potential during VCC measurements.

10-pA current, the charging time for the effective capacitor is estimated at 0.1 s.

Measurements in the H = const mode are thus possible only in a very narrow time interval. This makes accurate and detailed measurements unfeasible and calls for the significant redesign of the microscope.

To meet these mutually contradictory requirements, we used pulsed measurements of the VCCs. The equilibrium state of the system (U_b, I_b, and H_b) for which the spectral properties are measured will be called basic. It is necessary to emphasize that any correlations should be made for the same basic state. After switching the feedback circuit to a narrow bandwidth, the system is reset (for 2–10 s) to the equilibrium state. Then, Π-shaped pulses of variable height with a duration t_1 (Fig. 1) are applied to the system. The time between the pulses (when the tunnel voltage is reset to the basic state) is t_2. The pulse duration is selected in such a way as to ensure a low value of the capacitive component of the tunnel current and, at the same time, to prevent the action of the feedback loop. The tunnel current is measured during the last quarter of a pulse, and the associated pair of values is considered as a point in the voltage–current curve for a given basic state. The time between pulses (holdup time) for a basic voltage ($t_2 \gg t_1$) is experimentally selected such that the system is brought back to its equilibrium state. For our system, the values of t_1 = 0.5–1 ms and t_2 = 250–500 ms were found to be optimum.

A comparison between the VCCs obtained for the pulse mode and for a linear variation of the tunnel voltage has proved the validity of the pulse measurements. For our tunnel microscope (Litscan II [5]), the typical ramp rates of the tunnel voltage that exclude the drift of the feedback loop were 20–40 V/s for $\Delta U_{\text{tun}} = \pm 0.5$ V in reference to the basic state and 500–700 V/s for $\Delta U_{\text{tun}} = \pm 1.0$ V. In the latter case, essential distortions of the curves due to high capacitive charge currents were observed. The accuracy of measurements was also reduced because of large conversion times of the setting and detecting units of the instrument (for a ramp rate of the tunnel voltage of ~500 V/s, no more than 30 or 40 points in a curve could be detected). For $\Delta U_{\text{tun}} > \pm 1$ V, only the pulse-mode measurements were possible.

The STM operation is inevitably associated with noise of a different nature. The noise becomes significant at high values of the tunnel current; therefore, the obtained data must be averaged and smoothed, which adversely affects the measurement accuracy. In particular, this makes the differentiation of the curves, which is routinely used in tunnel VCC processing, difficult.

Obviously, for any system studied at fixed basic values of U_b and I_b, the resultant VCC will pass through two singular points. The first one corresponds to the zero current at the zero potential ($U = 0$, $I = 0$), and the second, to the basic current and potential values ($U = U_b$, $I = I_b$). For the systems with metal conductivity, the point with the coordinates $U = -U_b$, $I = -I_b$ is singular by virtue of the VCC symmetry. Thus, the VCC shape and slope can be varied in a very wide range by varying basic current and potential values (Fig. 2). For the majority of metal-conductivity systems, the VCCs at the identical basic conditions essentially coincide within a scatter of data points. Hence, these data cannot be used to determine the composition and the state of a sample. The probability of distinguishing the VCC curves grows when their singular points are as close to each other as possible, that is, at U_b and I_b approaching zero. In this case, however, a special low-current (several picoamperes) head with an extremely wide dynamic range (the upper limit to several nanoamperes) is required. At typical values of U_b (several hundred millivolts) and I_b (several hundred picoamperes), the VCCs are practically insensitive to the surface conditions, for instance, to a condensed contamination film formed in air.

The VCC differential form is undoubtedly more informative. However, a significant scatter of data points, which is inherent in conventional STMs, makes exact and reproducible differentiation impossible.

Actually, tunnel VCCs carry information only upon comparing the properties of objects differing in electrophysical nature, such as metals and semiconductors (Figs. 2 and 3). In topographical investigations of semiconductors with an STM, VCC measurements can be useful as a simple technique for choosing optimum U_b and I_b. The diode properties of a number of semiconducting substrates (Si, SnO_2, and TiO_2) make correct measurements impossible even at low values of I_b with the cutoff voltage across a $p–n$ junction.
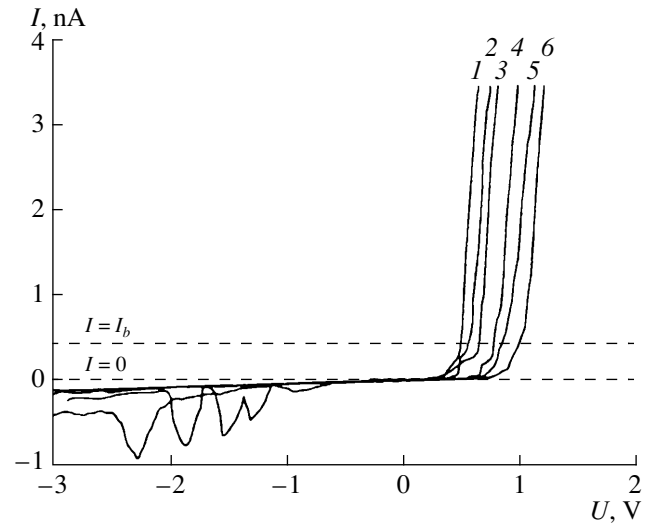
Current–Height Curves [$I = f(H)$, $U = const$]

Here, the temporary restrictions are identical to those for the VCC measurements.

For a precise variation of the tunnel gap (for microscopes operating in air, this gap usually ranges from 1 and 10 nm), a separate piezoceramic driver is used. It is either connected to the main driver, which is responsible for the equilibrium state of the system (typical distances are 500–1000 nm), or has a separate channel for
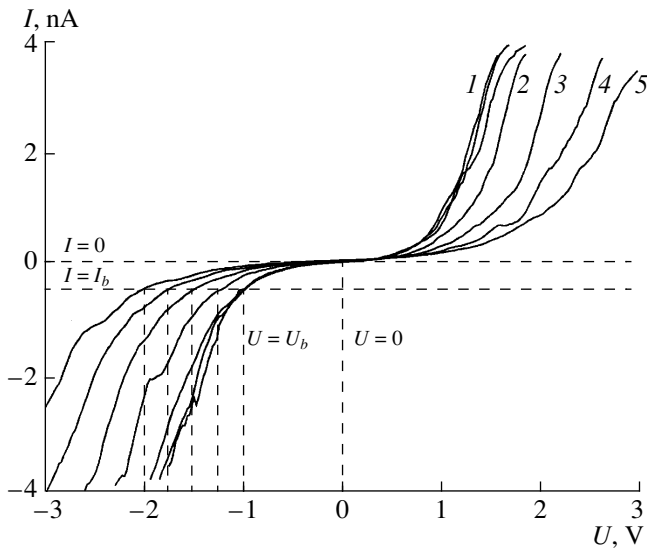
Fig. 2. Doubly smoothed VCCs of the Pt–polyaniline–air gap–Pt system at a basic tunnel current of 400 pA and basic tunnel voltages of (*1*) –1, (*2*) –1.25, (*3*) –1.5, (*4*) –1.75, and (*5*) –2 V. Dashed lines, singular point coordinates.

Fig. 3. VCCs of the semiconducting sample (SnO_2 on the conducting In–Sn oxide substrate) at a basic tunnel current of 400 pA and basic voltages of (*1*) 0.5, (*2*) 0.6, (*3*) 0.7, (*4*) 0.8, (*5*) 0.9, and (*6*) 1.0 V.
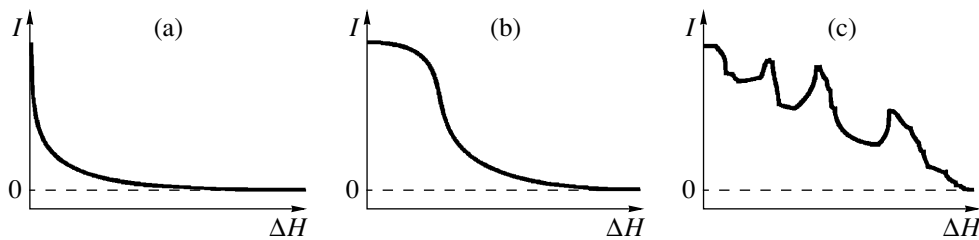
Fig. 4. Current–height curves in the (a) absence and (b, c) presence of the condensate. In the case (c), pulse-mode measurements are performed for the highly mobile condensate.

setting the potential of the piezoceramic. When the current through the system changes, the negative feedback loop causes the main (topographic) driver to reverse. This results in a distortion of the curves obtained. As in the previous case, the way out of this situation is to increase the response time of the feedback, which, in turn, imposes limitations on the duration of stay of the system in the nonequilibrium state.

On the other hand, the mechanical system of a scanning head has its own specific time—the time the head takes to set in a given position (tens or hundreds of microseconds, resonant frequency 10–100 kHz).

To correctly measure the current–height characteristics, we developed a pulse procedure similar to the one used for the VCC measurements. We recorded rise curves, when the tunnel gap increased. In the case of noncomplicated tunneling, the tunnel gaps usually measure shares of a nanometer; therefore, the displacement of the probe toward the sample may result in mechanical contact and hence damage the probe tip.

When a condensate is absent, typical current–height curves exhibit an abrupt exponential drop in current (Fig. 4a) within a distance of hundredths of a nanometer. If an insignificant amount of a condensate is present, the curves are of the same shape but with a lesser slope. In the presence of a mobile condensate, it will be entrapped by the contacting tip, as the tunnel gap increases, and detached from the tip at considerably larger gaps. In this case, the curves show the initial plateau with a subsequent abrupt drop (Fig. 4b). For the pulse-mode measurements, noisy unsmooth curves with sporadic extrema (Fig. 4c) due to the stochastic character of condensate separation from the tip may be observed.

Experimental data obtained for different samples of highly oriented pyrographite are presented in Fig. 5. Most of the curves show the presence of a condensate on the surface (Figs. 3 and 4). In the case of the sample surface carefully prepared by cleavage, the curve is close to exponential (curves *1* and *2*). It is in this case that acceptable atomic-resolution topographic results
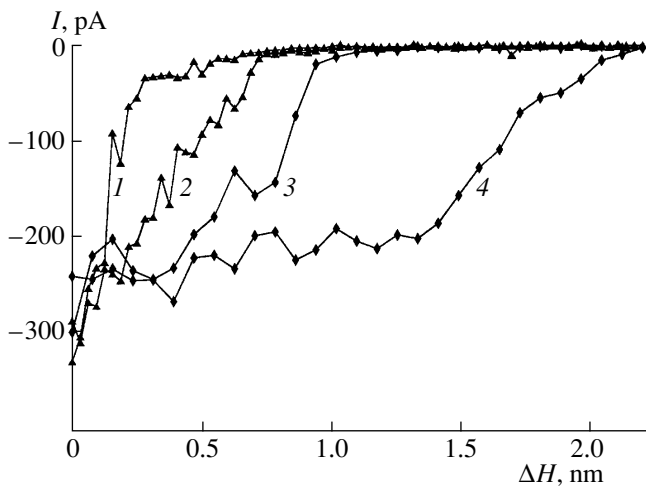
Fig. 5. Current–height curves for highly oriented pyrographite with various amounts of the condensate on the surface.
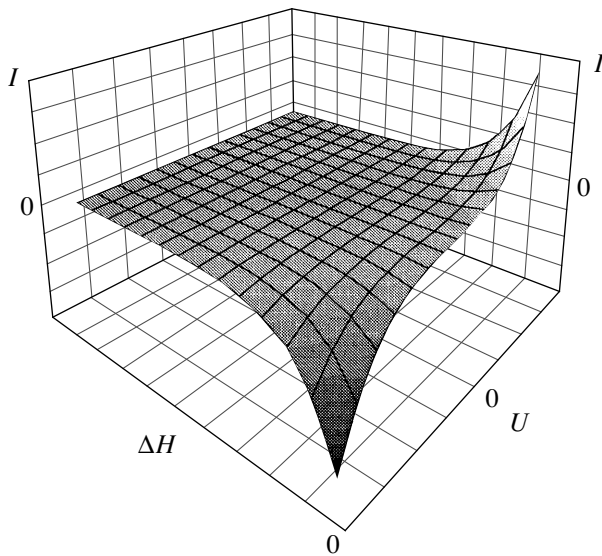
Fig. 6. Three-dimensional current–voltage–height diagram for a bipolar variation of the tunnel current. In the *UH* plane, $I = const$ sections are shown.

are obtainable. Being exposed to air or other contaminating media, the surface accumulates a condensate, and the plateau in the current–height curves is gradually extended (Fig. 5, curves *3* and *4*), causing the STM images to degrade. Under the conditions corresponding to curve *4* in Fig. 5, topographical measurements become impossible.

As a rule, the height of an effective tunnel barrier is estimated from the slope of the initial portion of the *I*–*H* curves by the equation [6]

$$V_{eff} = \frac{\hbar^2}{8m_e}\left(\frac{d\ln I}{dH}\right)^2, \qquad (1)$$

where $\hbar$ is Planck's constant and m_e is the mass of an electron.

Even for the as-prepared surface of pyrographite, when the greatest values of $\frac{d\ln I}{dH}$ are used, one fails to obtain the value of V_{eff} in excess of several tenths of electronvolt; usually, V_{eff} is under 0.1 eV. These values are much smaller than those obtained in a vacuum and/or in low-temperature STM experiments for the same substrate materials and tip [7]. They are close to V_{eff}'s obtained *in situ* in electrochemical STMs, having a polar liquid in the tunnel gap [8]. A condensate in an STM operating *ex situ* in air seems to be an analog of such a medium, because it is believed, starting from the literature data, that, in this case, the tunnel barrier lowers due to intermediate-state, or step, tunneling [9]. From our standpoint, there is also another reason, which is discussed in the next section.

The standard practice used for removing a condensate (heating of samples and purging of the chamber) cannot fully desorb contaminants but, probably, makes condensate films thinner and changes its properties, in particular, the effective viscosity. It is essential that, as a rule, the presence of relatively thin condensate films does not reduce the topographic potentialities of an STM when the resolution is not too high. In the case of the as-prepared surface of graphite, one can obtain atomic-resolution images even within 15–20 min after cleavage, although the values of V_{eff} are lower than the "vacuum" ones and further decline with time. However, the quality and the resolution of the images degrade as the film thickens and extended current plateaus in the *H*–*I* curves appear. It should be noted that the quality of STM images considerably depends on the composition and physicochemical properties of the condensate on the sample surface. For many samples (especially for those prepared from solutions), the portion of the curve where the current smoothly decays covers 2–10 nm. However, in the absence of the initial plateau, good topographic data with a fairly high resolution are obtainable even from such samples [5]. Here, the degradation of STM images also correlates with the length of the initial plateau in the *H*–*I* curves. This correlation is well explained in terms of the above assumption that the tip is in direct contact with a condensate and the meniscus is pulled off.

Voltage–Height Curves [H = f(U), I = const]

So far as we know, the third section of the three-dimensional tunnel spectrum has not been discussed in the literature; moreover, the associated curves have not been directly taken. A three-dimensional dependence of the tunnel current on the probe voltage and the tunnel gap is schematically shown in Fig. 6. The *H*–*U* curves must asymptotically tend to the straight line $U = 0$ and gradually rise with increasing *U*. The fundamental physical interpretation of *H*–*U* curves in terms of avail-
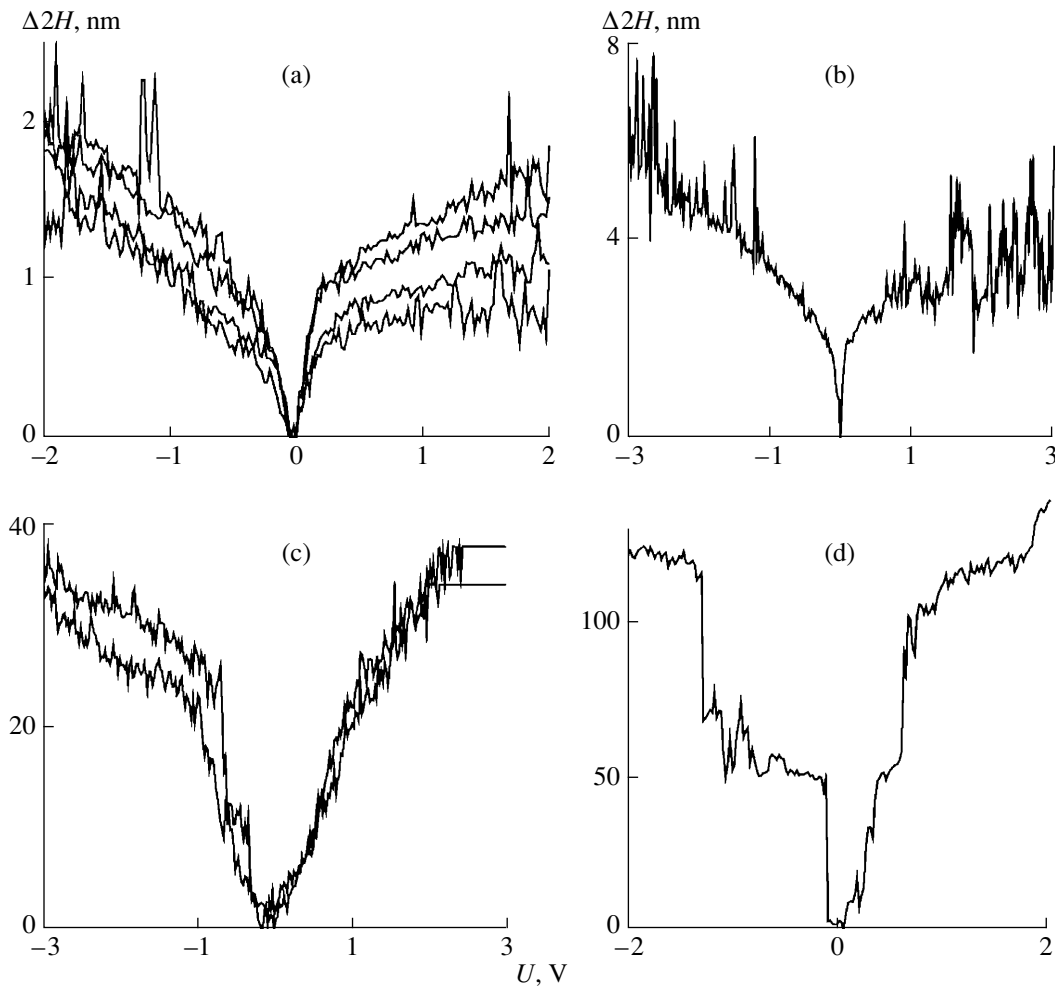
Fig. 7. *H–U* curves obtained for (a) highly oriented pyrographite, (b) platinum foil, (c) conducting polymer (polyaniline) film on platinum, and (d) SnO$_2$ film on conducting In–Sn oxide substrate.

able theoretical concepts is still in the future. In the literature, the associated data are lacking. The reason is likely to be that in a vacuum, where spectroscopic measurements are usually conducted, the condensate on the sample surface is absent. In this case, the tunnel gaps are usually tenths of a nanometer and the corresponding height differences in the *H–U* curves do not exceed hundredths of a nanometer. It is difficult to detect such distances by conventional topographic systems. At the same time, our experiments showed that, in air, the height differences in the *H–U* curves are 2–100 nm and can be easily detected with a standard STM topography-recording channel.

The major advantage of the new mode for measuring the tunnel spectra is that the maintenance of the tunnel current at a constant level by varying the probe position is routine in tunnel microscope operation. Accordingly, neither special hardware requirements nor stringent time restrictions are imposed: the measurements can be carried out with any tunnel microscope

operating in air. However, the physical interpretation of the obtained curves is very difficult.

It should be noted that, even in bipolar measurements, very low values of the potential difference between the probe and a sample are impermissible to avoid short circuit conditions. In our experiments, the minimum difference between the probe and sample potentials was 50 mV; that is, the measurements were performed for $|U| > 50$ mV. The measured values of ΔH are relative; to compare the obtained curves, they must be normalized by either of two methods. It is best if the tip position before the measurement is taken as zero (when $U = U_b$ and $I = I_b$). We chose another, simplified, way—normalization to the minimum height on the curve.

The shape of *H–U* curves and their extent heavily depend on the nature of a sample and surface preparation (Fig. 7). For as-cleaved pyrographite (Fig. 7a), the shape of the curve is consistent with theoretical assumptions and its specific height is about 2 nm. When the sample is exposed to air (accumulation of a conden-
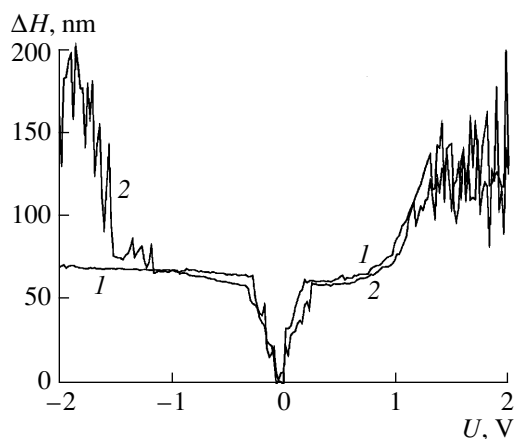
Fig. 8. *H–U* curves obtained with the copper-cluster-coated probe for the Pt + polyaniline sample. (*1*) As-deposited polyaniline film and (*2*) after several measurements.
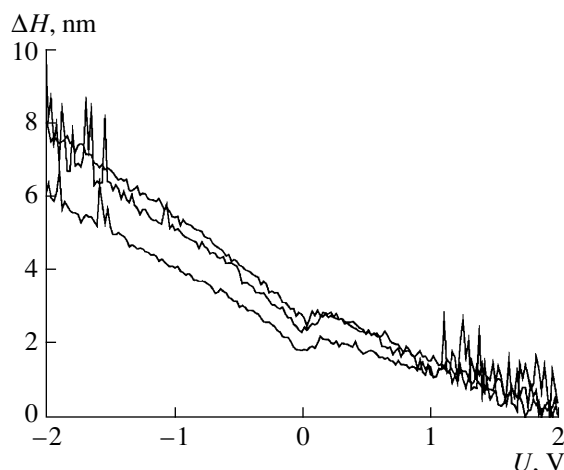
Fig. 9. *H–U* curves for highly oriented pyrographite that are distorted by thermal drift immediately after the positioning of the sample.

sate), the height of the curve gradually grows and the quality of STM images is deteriorated. In the case of a platinum foil, the characteristic heights approach 10 nm (Fig. 7b). For chemically active samples or those synthesized in a water medium, the curves are much higher: for a polyaniline film electrodeposited on the surface of platinum, their heights are 30–70 nm (Fig. 7c). For samples chemically active in air, the characteristic heights grow to 100–500 nm and the curves often show sharp changes in the height (Fig. 7d). It should be emphasized that the chemical activity of samples can be defined not only by their basic component but also by impurities.

As was indicated, the measured heights are relative values; however, the curves suggest that, for the standard operating mode of a tunnel microscope, the tunnel gap is no less than the height difference in an *H–U* curve. Thus, even for highly oriented pyrographite, the tip–sample distances are such that direct electron tunneling is unlikely, but some intermediate-state mechanisms are quite possible. In the case of vacuum microscopes, where direct electron tunneling takes place, the associated distances are one order of magnitude less [4, 7].

In the system discussed, tunneling through a set of intermediate centers [9] cannot be ruled out. These centers may be condensate molecules. The probability of step tunneling for an electron is the product of the probabilities of tunneling through each of the centers. Thus, the probability drastically decreased with increasing number of the centers; therefore, this model cannot explain the effect of long-range tunneling (tens or hundreds of nanometers), for which hundreds of the intermediate centers are necessary. It seems that the processes taking place in air tunnel microscopes are similar to those occurring in electrochemical scanning microscopes [10]. The operation of the latter is based

on keeping the Faraday current through probe ultramicroelectrodes fixed.

A condensate film on the sample surface is, in essence, an electrolyte of unknown composition. The STM probe dipped into the condensate and the surface form a two-electrode electrochemical cell. When the potential difference between the probe and the sample is changed, redox processes giving rise to the Faraday current in the system occur at the electrodes. In this case, the equilibrium probe–sample distance is determined by the ohmic resistance of the condensate and the nature of the electrode processes. Such an approach allows us to explain the abrupt height fluctuations in the *H–U* curves observed in the experiment (Fig. 7d). As the interelectrode voltage increases, so do the electrode potentials; therefore, at some moment, additional redox processes begin to proceed. This results in a drastic increase in the current (the total resistance decreases), and the tip abruptly bounces back from the surface.

By way of illustration, let us consider the following experimental result. If an electrochemically active (Cu) cluster is deposited on the surface of a Pt–Ir probe, the second rise appears in the *H–U* curves at the positive potentials of the probe (Fig. 8, curve *1*). This region, absent in the curves obtained with a usual probe (Fig. 7c), can be related to the anodic dissolution of copper. After several such measurements, the copper ions are accumulated in the condensate film and a spike also appears in the negative branch of the *H–U* curve (Fig. 8, curve *2*). It apparently corresponds to the cathodic reduction of the copper ions. Once the tip has been shifted by 1–2 μm along the surface, the spike on the negative branch dies away and arises again only after the accumulation of products of anodic dissolution.

Sometimes, the *H–U* curves are distorted by thermal drift of a sample. If, during the measurements, the ref-

erence voltage is varied linearly with time, the curves show a systematic approach of the probe to the surface (Fig. 9). After a lapse of 15–20 min taken for thermal equilibrium to establish, these distortions disappear. It follows that the *H–U* dependences can also be applied to reveal thermal drift of a sample and prevent distortions in the geometrical sizes of objects under study.

Thus, the *H–U* curves can be useful in finding optimum conditions for the topographic measurements and in controlling the surface condition. It is necessary to stress that the measurements should not be performed at probe voltages when the height abruptly increases in the *H–U* curves, because, under these conditions, the system is in unstable equilibrium. As a result, the topographic data will be heavily distorted by sharp spikes. For a number of samples, two equilibrium probe positions are observed in the near-spike region (at the peak and the bottom of a spike). The distance between these positions may vary from 300 to 400 nm. Under these conditions, the topographic measurements are nearly impossible. The optimum parameters correspond to the smooth extended portions of the curves with a moderate inclination.

CONCLUSION

With a conducting film of a condensate present on the sample surface, in an STM operating *ex situ*, the scanning electrochemical microscopy mode is realized. This mode, whose application was pioneered by Bard *et al.* [10], is based on the detection of Faraday currents through ultramicroelectrodes. Topographic images obtained under this mode offer a fairly high resolution; here, the film-coated solid surface, rather than a conducting condensate exposed to air, is visualized. In the strict sense, this is the limiting case of the step-tunnel-ing mechanism, which occurs when the separation between neighboring centers of intermediate electron localization is small.

REFERENCES

1. A. I. Danilov, Usp. Khim. **64,** 818 (1995).
2. *Scanning Tunneling Microscopy*, Ed. by W. J. Guntherodt and R. Wiesendasnger (Springer, Berlin, 1991).
3. R. Jansen, H. Kempen, and R. M. Wolf, J. Vac. Sci. Technol., B. **14,** 1173 (1996).
4. *Handbook of Microscopy: Methods II,* Ed. by S. Amelinckx, D. van Dyck, J. van Landuyt, *et al.* (VCH, Berlin, 1997), pp. 807–825.
5. G. A. Tsirlina, O. A. Petriĭ, and S. Yu. Vassiliev, J. Electroanal. Chem. Interfacial Electrochem. **414,** 41 (1996).
6. A. Vaught, T. W. Jing, and S. M. Lindsay, Chem. Phys. Lett. **236,** 306 (1995).
7. O. Pecina, W. Schmickler, K. Y. Chan, *et al.*, J. Electroanal. Chem. Interfacial Electrochem. **396,** 303 (1995).
8. J. Pan, T. W. Jing, and S. M. Lindsay, J. Phys. Chem. **98,** 4205 (1994).
9. J. Halbritter, G. Repphun, S. Vinzelberg, *et al.*, Electrochim. Acta **40,** 1385 (1995).
10. A. J. Bard, D. E. Cliffel, C. Demaille, *et al.*, Ann. Chim. **87,** 15 (1997).

Translated by B.A. Malyukov