Electrostatic field of a thin, unclosed spherical shell and a torus

G. Ch. Shushkevich

Ya. Kupaly Grodno State University, 230023 Grodno, Belarus
(Submitted March 5, 1997)
Zh. Tekh. Fiz. **68**, 1–6 (July 1998)

The electrostatic problem for a thin, unclosed spherical shell and a torus is reduced to paired summation equations in the Legendre polynomials by means of formulas relating the spherical and toroidal harmonic functions. The paired equations are transformed to a Fredholm integral equation of the second kind. Formulas are obtained for computing the charges of the conductors in the form of a series in a small parameter. The capacitance is computed for certain geometrical parameters of the conductors. © *1997 American Institute of Physics.*
[S1063-7842(97)00107-4]

In designing various devices, it becomes necessary to calculate the electrostatic field of conductors of various configurations.[1–3] This paper discusses the electrostatic problem for a thin, unclosed spherical shell S and a torus T, where the spherical shell is on the surface of sphere S_1, which does not intersect the torus (Fig. 1).

The method of equivalent charges has been used previously to solve the electrostatic problem for a sphere and a torus with a common axis of rotation.[4] The method of paired equations has been used to solve the electrostatic problem for an unclosed spherical shell S and torus T when the surface S_1 intersects the torus.[5]

1. Let us consider the axisymmetric electrostatic problem for a thin, unclosed spherical shell S and a torus T with minor radius r and major radius R. Shell S lies on a sphere S_1 of radius d. An axial cross section of the conductors is shown in Fig. 1. To analytically describe conductors with point O as the center of sphere S_1, we connect the spherical coordinates $\{r,\Theta,\varphi\}$

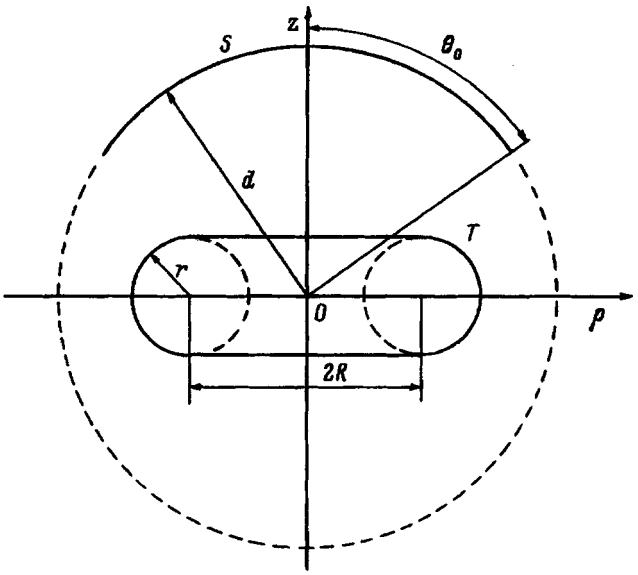


FIG. 1.

$$x=r\cos\varphi\sin\Theta,\quad y=r\sin\varphi\sin\Theta,\quad z=r\cos\Theta$$
$$(0\leqslant r<\infty,\quad 0\leqslant\Theta\leqslant\pi,\quad 0\leqslant\varphi\leqslant 2\pi)$$

with the toroidal coordinates $\{\alpha,\beta,\varphi\}$ (Ref. 6)

$$x=\frac{c\sinh\alpha\cos\varphi}{\cosh\alpha-\cos\beta},\quad y=\frac{c\sinh\alpha\sin\varphi}{\cosh\alpha-\cos\varphi},$$
$$z=\frac{c\sin\beta}{\cosh\alpha-\cos\beta}$$

$(0\leqslant\alpha<\infty,\ -\pi<\beta\leqslant\pi,\ 0\leqslant\varphi\leqslant 2\pi,\ c=\sqrt{R^2-r^2})$.

Then the conductors under consideration are described as follows

$$T=\left\{\alpha=\alpha_0=\ln\left(\frac{R}{r}+\sqrt{\left(\frac{R}{r}\right)^2-1}\right),\right.$$
$$\left.0\leqslant\beta\leqslant 2\pi,\quad 0\leqslant\varphi\leqslant 2\pi\right\},$$
$$S=\{r=d,\quad 0\leqslant\Theta\leqslant\Theta_0,\quad 0\leqslant\varphi\leqslant 2\pi\}.$$

To solve the problem, we arbitrarily divide all of space E_3 of sphere S_1 into two regions $W_1(r>d)$ and $W_2=E_3/(W_1UT)$. We denote the potentials of the electrostatic field in these regions as U_1 and U_2, respectively. The potentials U_i must satisfy Laplace's equation

$$\Delta U_i=0,\quad i=1,2, \tag{1}$$

(where Δ is the Laplacian operator) with the boundary conditions

$$U_2|_T=V_t \text{ const}, \tag{2}$$
$$U_2|_S=V_s \text{ const}, \tag{3}$$

and the condition at infinity

$$U_1(M)\to 0 \text{ as } M\to\infty, \tag{4}$$

where M is an arbitrary point of space.

Moreover, the potential must be continuous on sphere S_1, and the field on the part of sphere S_1 that is not a conductor must be continuous; i.e.,

$$U_1=U_2,\quad r=d,\quad 0\leqslant\Theta\leqslant\pi, \tag{5}$$
$$\frac{\partial U_1}{\partial r}=\frac{\partial U_2}{\partial r},\quad r=d,\quad \Theta_1<\Theta\leqslant\pi. \tag{6}$$

2. According to the method of separation of variables, we write potentials U_i, $i=1,2$ of the electrostatic field in the form[6,7]

$$U_1(r,\Theta)=\sum_{n=0}^{\infty} b_n\left(\frac{d}{r}\right)^{n+1} P_n(\cos\Theta) \quad \text{in} \quad W_1, \tag{7}$$

$$U_2=U_2^{(1)}(r,\Theta)+U_2^{(2)}(\alpha,\beta) \quad \text{in} \quad W_2, \tag{8}$$

where

$$U_2^{(1)}(r,\Theta)=\sum_{n=0}^{\infty} a_n\left(\frac{r}{d}\right)^{n} P_n(\cos\Theta), \tag{9}$$

$$U_2^{(2)}(\alpha,\beta)=\sqrt{2(\cosh\alpha-\cos\beta)}\sum_{n=0}^{\infty}(M_n\cos n\beta + N_n\sin n\beta)\frac{P_{n-\frac{1}{2}}(\cosh\alpha)}{P_{n-\frac{1}{2}}(\cosh\alpha_0)}, \tag{10}$$

$P_n(\cos\Theta)$ are the Legendre polynomials, and $P_{n-\frac{1}{2}}(\cosh\alpha)$ are Legendre functions of the first kind.[6–8]

Potential $U_1(r,\Theta)$ satisfies the condition at infinity given by Eq. (4). The unknown coefficients a_n, b_n, M_n, and N_n are determined by the conditions given by Eqs. (2), (3), (5), and (6).

To solve the problem, we need the following formulas, which connect the spherical and toroidal harmonic functions:[9]

$$r^n P_n(\cos\Theta)=\frac{c^n}{2\pi}\sum_{s=-\infty}^{\infty} D_n^s\sqrt{2(\cosh\alpha-\cos\beta)}\times Q_{s-\frac{1}{2}}(\cosh\alpha)e^{is\beta}, \tag{11}$$

$$\sqrt{2(\cosh\alpha-\cos\beta)}P_{n-\frac{1}{2}}(\cosh\alpha)e^{in\beta}=\sum_{s=0}^{\infty}(-1)^s c^{s+1}D_s^n r^{-s-1}P_s(\cos\Theta), \tag{12}$$

where

$$D_n^s=2P_n(0)+\sum_{k=1}^{s} g_k^s(sP_{n-k}^k(0)-iknP_{n-k}^{k-1}(0)),$$

$$D_n^{-s}=\overline{D_n^s}, \tag{13}$$

$$g_k^s=\frac{2^{3k+1}k!(s+k+1)!}{((2k)!)^2(s-k)!},$$

$$P_n^m(0)=\frac{2^m}{\sqrt{\pi}}\cos\left[\frac{(n+m)\pi}{2}\right]\frac{\Gamma\left(\frac{n+m+1}{2}\right)}{\Gamma\left(\frac{n-m+1}{2}\right)},$$

$\Gamma(n)$ is the gamma function, $Q_{n-\frac{1}{2}}(\cosh\alpha)$ are the Legendre functions of the second kind, and $P_{n-k}^k(x)$ are the associated Legendre functions.[6–8]

To satisfy the boundary condition given by Eq. (2) on the surface of torus T, we write the potential $U_2^{(1)}(r,\Theta)$ in terms of the toroidal harmonic functions, using Eq. (11). Then

$$U_2^{(1)}(\alpha,\beta)=\frac{\sqrt{2(\cosh\alpha-\cos\beta)}}{2\pi}\times\sum_{n=-\infty}^{\infty}\left(\sum_{p=0}^{\infty}\left(\frac{c}{d}\right)^p a_p D_p^n\right)Q_{n-\frac{1}{2}}(\cosh\alpha)e^{in\beta}. \tag{14}$$

Using the equations

$$\cos n\beta=\frac{1}{2}(e^{in\beta}+e^{-in\beta}), \quad \sin\beta=\frac{1}{2i}(e^{in\beta}-e^{-in\beta})$$

and the following relationships for the Legendre functions:[6,8]

$$P_{n-\frac{1}{2}}(\cosh\alpha)=P_{-n-\frac{1}{2}}(\cosh\alpha),$$

$$Q_{n-\frac{1}{2}}(\cosh\alpha)=Q_{-n-\frac{1}{2}}(\cosh\alpha),$$

we can write the potential $U_2^{(2)}(\alpha,\beta)$ in complex form

$$U_2^{(2)}(\alpha,\beta)=\sqrt{2(\cosh\alpha-\cos\beta)}\times\sum_{n=-\infty}^{\infty} x_n\frac{P_{n-\frac{1}{2}}(\cosh\alpha)}{P_{n-\frac{1}{2}}(\cosh\alpha_0)}e^{in\beta}, \tag{15}$$

where

$$x_n=\frac{1}{2}(M_n-iN_n), \quad x_{-n}=\frac{1}{2}(M_n+iN_n),$$

$$\overline{x_n}=x_{-n}, \quad n=0,1,2,\ldots.$$

According to Eqs. (14) and (15), the boundary condition given by Eq. (2) on the surface of a torus takes the form

$$\sqrt{2(\cosh\alpha_0-\cos\beta)}\sum_{n=-\infty}^{\infty}\left(x_n+\frac{1}{2\pi}Q_{n-\frac{1}{2}}(\cosh\alpha_0)\times\sum_{p=0}^{\infty}\left(\frac{c}{d}\right)^p a_p D_p^n\right)e^{in\beta}=V_t. \tag{16}$$

Dividing both sides of Eq. (16) by $\sqrt{2(\cosh\alpha_0-\cos\beta)}$ and using the expansion[6]

$$\frac{1}{\sqrt{2(\cosh\alpha_0-\cos\beta)}}=\frac{1}{\pi}\sum_{n=-\infty}^{\infty} Q_{n-\frac{1}{2}}(\cosh\alpha_0)e^{in\beta},$$

we get from Eq. (16)

$$\sum_{n=-\infty}^{\infty}\left(x_n+\frac{1}{2\pi}Q_{n-\frac{1}{2}}(\cosh\alpha_0)\sum_{p=0}^{\infty}\left(\frac{c}{d}\right)^p a_p D_p^n\right)e^{in\beta}=\frac{V_t}{\pi}\sum_{n=-\infty}^{\infty} Q_{n-\frac{1}{2}}(\cosh\alpha_0)e^{in\beta} \tag{17}$$

or, because of the uniqueness of the expansion in Fourier series,

$$x_n+\frac{Q_{n-\frac{1}{2}}(\cosh\ \alpha_0)}{2\pi}\sum_{p=0}^{\infty}\left(\frac{c}{d}\right)^p a_p D_p^n=\frac{V_t}{\pi}Q_{n-\frac{1}{2}}(\cosh\ \alpha_0),$$

$$n=0,\pm 1,\pm 2,\ldots. \tag{18}$$

To satisfy the boundary condition given by Eq. (3) on the surface of spherical shell S and the continuity conditions given by Eqs. (5) and (6), we write the potential $U_2^{(2)}(\alpha,\beta)$ in terms of the spherical harmonic functions, using Eq. (12). As a result, we have

$$U_2^{(2)}(r,\Theta)=\sum_{n=0}^{\infty}(-1)^n\left(\sum_{s=-\infty}^{\infty}\frac{x_s D_n^s}{P_{s-\frac{1}{2}}(\cosh\ \alpha_0)}\right)\times\left(\frac{c}{r}\right)^{n+1}P_n(\cos\ \Theta). \tag{19}$$

According to Eqs. (7), (9), and (19) and the orthogonality condition of the Legendre polynomials $P_n(\cos\Theta)$, on the line segment $[0,\pi]$, the continuity condition given by Eq. (5) is equivalent to

$$b_n=a_n+f_n,\quad n=0,\,1,\,2,\ldots, \tag{20}$$

where

$$f_n=(-1)^n\sum_{s=-\infty}^{\infty}\frac{x_s D_n^s}{P_{s-\frac{1}{2}}(\cosh\ \alpha_0)}. \tag{21}$$

The boundary condition given by Eq. (5) takes the form

$$\sum_{n=0}^{\infty}(a_n+f_n)P_n(\cos\ \Theta)=V_s,\quad 0\leqslant\Theta<\Theta_0. \tag{22}$$

After computing the derivatives

$$\frac{\partial}{\partial r}U_1(r,\Theta),\quad \frac{\partial}{\partial r}U_2^{(i)}(r,\Theta),\quad i=1,2,$$

and having satisfied the continuity condition given by Eq. (6), and bearing in mind Eq. (20), we get

$$\sum_{n=0}^{\infty}\left(n+\frac{1}{2}\right)a_n P_n(\cos\ \Theta)=0,\quad \Theta_0<\Theta\leqslant\pi. \tag{23}$$

Thus, the conditions given by Eqs. (3), (5), and (6) result in the paired summation equations in the Legendre polynomials

$$\sum_{n=0}^{\infty}a_n P_n(\cos\ \Theta)=F(\Theta),\quad 0\leqslant\Theta<\Theta_0,$$

$$\sum_{n=0}^{\infty}\left(n+\frac{1}{2}\right)a_n P_n(\cos\ \Theta)=0,\quad \Theta_0<\Theta\leqslant\pi, \tag{24}$$

where

$$F(\Theta)=\sum_{n=0}^{\infty}(V_s\delta_{0n}-f_n)P_n(\cos\ \Theta), \tag{25}$$

and δ_{0n} is the Kronecker delta.

In place of the coefficients a_n, we introduce into the treatment the new function $\varphi(x)$ $[\varphi(x)\in C^{(1)}_{[0,\Theta_0]}]$, given by the formula

$$a_n=\int_0^{\Theta_0}\varphi(x)\cos\left(n+\frac{1}{2}\right)x\,dx. \tag{26}$$

Then the paired Eqs. (24) transform to (See Ref. 10, p. 166)

$$\varphi(x)=\frac{2}{\pi}\frac{d}{dx}\int_0^x\frac{F(\Theta)\sin\ \Theta\,d\Theta}{\sqrt{2(\cos\ \Theta-\cos\ x)}},\quad 0<x<\Theta_0. \tag{27}$$

Substituting function $F(\Theta)$ from Eqs. (25) into Eq. (27) and bearing in mind the integral (see Ref. 10, p. 198)

$$\frac{d}{dx}\int_0^x\frac{P_n(\cos\ \Theta)\sin\ \Theta\,d\Theta}{\sqrt{2(\cos\ \Theta-\cos\ x)}}=\cos\left(n+\frac{1}{2}\right)x,$$

we get that

$$\varphi(x)=\frac{2V_s}{\pi}\cos\frac{x}{2}-\frac{2}{\pi}\sum_{n=0}^{\infty}f_n\cos\left(n+\frac{1}{2}\right)x,$$

$$0<x<\Theta_0. \tag{28}$$

Substituting a_n from Eq. (26) into (18), we establish the connection between the coefficients x_s and the function $\varphi(x)$

$$x_s=\frac{Q_{s-\frac{1}{2}}(\cosh\ \alpha_0)}{2\pi}\left(2V_t-\int_0^{\Theta_0}\left(\sum_{p=0}^{\infty}\left(\frac{c}{d}\right)^p D_p^s\times\cos\left(p+\frac{1}{2}\right)u\right)\varphi(u)\,du\right),\quad s=0,\pm 1,\pm 2,\ldots. \tag{29}$$

Substituting x_s from Eq. (29) into Eq. (21), and f_n from Eq. (21) into the right-hand side of Eq. (28), after some transformations, we get an Fredholm integral equation of the second kind for function $\varphi(x)$

$$\varphi(x)-\int_0^{\Theta_0}K(x,u)\varphi(u)\,du=H(x),\quad 0<x<\Theta_0, \tag{30}$$

where the kernel of the integral equation has the form

$$K(x,u)=\frac{1}{\pi^2}\sum_{n=0}^{\infty}\sum_{p=0}^{\infty}(-1)^n\left(\frac{c}{d}\right)^{n+p+1}\times\cos\left(n+\frac{1}{2}\right)x\ \cos\left(p+\frac{1}{2}\right)uR_{np}, \tag{31}$$

$$R_{np}=\sum_{s=-\infty}^{\infty}\frac{Q_{s-\frac{1}{2}}(\cosh\ \alpha_0)}{P_{s-\frac{1}{2}}(\cosh\ \alpha_0)}D_n^s D_p^s, \tag{32}$$

while the right-hand side is

$$H(x)=\frac{2V_s}{\pi}\cos\frac{x}{2}-\frac{2V_t}{\pi^2}\sum_{n=0}^{\infty}(-1)^n\left(\frac{c}{d}\right)^{n+1}\times\cos\left(n+\frac{1}{2}\right)xT_n, \tag{33}$$

$$T_n=\sum_{s=-\infty}^{\infty}\frac{Q_{s-\frac{1}{2}}(\cosh\ \alpha_0)}{P_{s-\frac{1}{2}}(\cosh\ \alpha_0)}D_n^s. \tag{34}$$

3. The total charge Q_s of the thin, unclosed spherical shell S is computed from

$$Q_s=4\pi\varepsilon d\int_0^{\Theta_0}\varphi(x)\cos\frac{x}{2}dx, \tag{35}$$

while the total charge Q_t of torus T is computed from

$$Q_t=8\pi\varepsilon c\sum_{n=-\infty ty}^{\infty}\frac{x_n}{P_{n-\frac{1}{2}}(\cosh\alpha_0)}=4\varepsilon c\Bigg(2V_tS_0-\int_0^{\Theta_0}\Bigg(\sum_{p=0}^{\infty}\left(\frac{c}{d}\right)^p T_p\cos\left(p+\frac{1}{2}\right)t\Bigg)\varphi(t)dt\Bigg), \tag{36}$$

where

$$S_0=\sum_{s=0}^{\infty}\delta_s\frac{Q_{s-\frac{1}{2}}(\cosh\alpha_0)}{P_{s-\frac{1}{2}}(\cosh\alpha_0)},\quad \delta_s=\begin{cases}1, & s=0,\\ 2, & s\geqslant 1,\end{cases}$$

and ε is the permittivity of the medium.

Knowing the charges of the conductors, it is possible to compute the capacitance from[1]

$$C=\frac{C_{11}C_{12}+C_{11}C_{22}+C_{12}C_{22}}{C_{11}+C_{22}}, \tag{37}$$

where $C_{11}=Q_s$ when $V_s=V_t=1$, $C_{22}=Q_t$ when $V_s=V_t=1$, $C_{12}=Q_s$ when $V_s=0$, $V_t=-1$, $C_{21}=Q_t$ when $V_s=-1$, $V_t=0$, and $C_{12}=C_{21}$.

In the general case, the Fredholm integral equation of the second kind, Eq. (30), can be solved by numerical methods.[11]

We assume that $\mu=c/d\ll 1$, and, in what follows, we assume that $\mu^n\approx 0$ when $n>4$. In this case, the solution of the integral Eq. (30) can be represented in the form of a series in the small parameter μ

$$\varphi(x)=\varphi_0(x)+\varphi_1(x)\mu+\varphi_2(x)\mu^2+\varphi_3(x)\mu^3+\ldots, \tag{38}$$

where $\varphi_i(x)(i=1,2,3)$ are unknown functions.

Let us also expand the kernel of the integral equation, Eq. (31), and the right-hand side, given by Eq. (33), in series in the small parameter μ:

$$K(x,u)=\frac{1}{\pi^2}\Bigg[\cos\frac{x}{2}\cos\frac{u}{2}R_{00}\mu+\Bigg(\cos\frac{x}{2}\cos\frac{3u}{2}-\cos\frac{3x}{2}\cos\frac{u}{2}\Bigg)R_{10}\mu^2+\Bigg(\cos\frac{x}{2}\cos\frac{5u}{2}R_{20}-\cos\frac{3x}{2}\cos\frac{3u}{2}R_{11}+\cos\frac{5x}{2}\cos\frac{u}{2}R_{02}\Bigg)\times\mu^3+\ldots\Bigg],$$

$$H(x)=\frac{2}{\pi}\Bigg[V_s\cos\frac{x}{2}-\frac{V_t}{\pi}\cos\frac{x}{2}T_0\mu+\frac{V_t}{\pi}\cos\frac{3x}{2}T_1\mu^2-\frac{V_t}{\pi}\cos\frac{5x}{2}T_2\mu^3+\ldots\Bigg].$$

Substituting the expansions for $K(x,u)$, $H(x)$, and $\varphi(x)$ into Eq. (30) and setting the coefficients for identical powers of μ equal, we find the functions $\varphi_i(x)(i=1,2,3)$:

$$\varphi_0(x)=\frac{2V_s}{\pi}\cos\frac{x}{2},$$

$$\varphi_1(x)=\frac{1}{\pi^2}(V_sR_{00}\alpha_{00}-2V_tT_0)\cos\frac{x}{2},$$

$$\varphi_2(x)=\frac{V_s}{\pi^2}\Bigg(\alpha_{10}\cos\frac{x}{2}-\alpha_{00}\cos\frac{3x}{2}\Bigg)R_{10}+\frac{\alpha_{00}}{2\pi^3}(V_sR_{00}\alpha_{00}-2V_tT_0)R_{00}\cos\frac{x}{2}+\frac{2V_t}{\pi^2}T_1\cos\frac{3x}{2},$$

$$\varphi_3(x)=\frac{V_s}{\pi^2}\Bigg(R_{20}\alpha_{20}\cos\frac{x}{2}-R_{11}\alpha_{10}\cos\frac{3x}{2}+R_{02}\alpha_{00}\cos\frac{5x}{2}\Bigg)+\frac{R_{10}}{2\pi^3}(V_sR_{00}\alpha_{00}-2V_tT_0)\times\Bigg(\alpha_{10}\cos\frac{x}{2}-\alpha_{00}\cos\frac{4x}{2}\Bigg)-\frac{2V_t}{\pi^2}T_2\cos\frac{5x}{2}+R_{00}\cos\frac{x}{2}\Bigg(\frac{1}{4\pi^4}(V_sR_{00}\alpha_{00}-2V_tT_0)R_{00}\alpha_{00}^2+\frac{V_t}{\pi^3}T_1\alpha_{10}\Bigg),$$

where

$$\alpha_{kn}=\frac{1}{\pi}\Bigg(\frac{\sin(k-n)\Theta_0}{k-n}+\frac{\sin(k+n+1)\Theta_0}{k+n+1}\Bigg),\qquad \frac{\sin(k-n)\Theta_0}{k-n}\Bigg|_{k=n}=\Theta_0.$$

It follows from Eq. (13) that $D_0^s=2$, $D_1^s=-4is$, $D_2^s=-1-4s^2$, $D_n^{-s}=\overline{D}_n^s$, $s\geqslant 0$, $n=0,1,2$. Then, from Eqs. (32) and (34), we get

$$R_{00}=4S_0,\quad R_{11}=8S_1,\quad R_{20}=R_{02}=-2S_2,$$

$$T_0=2S_0,\quad T_1=0,\quad T_2=-S_2,\quad S_1=S_2S_0,$$

where

TABLE I.

	Parameter p								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
S_0	1.139	1.393	1.633	1.896	2.205	2.598	3.143	4.016	5.903
S_2	1.189	1.601	2.140	2.909	4.086	6.052	9.811	18.87	55.68

$$S_2=\sum_{s=0}^{\infty}\delta_s(1+4s^2)\frac{Q_{s-\frac{1}{2}}(\cosh\alpha_0)}{P_{s-\frac{1}{2}}(\cosh\alpha_0)}.$$

Substituting the expansions given by Eqs. (38) into Eqs. (35) and (36), we get formulas for computing the charges of the conductors in the form of a series in the small parameter μ:

$$Q_s=4\pi\varepsilon d\left\{V_s\alpha_{00}+\frac{2\alpha_{00}}{\pi}(V_s\alpha_{00}-V_t)S_0\mu+\frac{4\alpha_{00}^2}{\pi^2}(V_s\alpha_{00}-V_t)S_0^2\mu^2+\left[\frac{2V_s}{\pi}(2\alpha_{10}^2S_1-\alpha_{00}\alpha_{20}S_2)+\frac{8\alpha_{00}^3}{\pi^3}(V_s\alpha_{00}-V_t)S_0^3+\frac{2V_t}{\pi}\alpha_{20}S_2\right]\mu^3+\ldots\right\},$$

$$Q_t=8c\varepsilon\left\{V_tS_0-V_s\alpha_{00}S_0-\frac{2\alpha_{00}}{\pi}(V_s\alpha_{00}-V_t)S_0^2\mu-\left[\frac{4\alpha_{00}^2}{\pi^2}(V_s\alpha_{00}-V_t)S_0^3-V_s\alpha_{20}S_2\right]\mu^2+\ldots\right\}.$$

The charge of a torus inside a grounded sphere ($V_s=0$, $\Theta_0=\pi$) is computed from

$$Q_t=8c\varepsilon V_tS_0\left(1+\frac{2}{\pi}S_0\mu+\frac{4}{\pi^2}S_0^2\mu^2+\ldots\right).$$

Capacitance C has been computed for certain geometrical parameters of the conductors. For a numerical solution, the Fredholm integral of the second kind, Eq. (30), was transformed into a finite system of linear algebraic equations (SLAE) by means of Simpson's quadrature formula:[11]

TABLE II.

Θ_0, deg	Normalized capacitance $\tilde{C}$			
	$\frac{R}{d}=\frac{1}{2}$	$\frac{R}{d}=\frac{2}{5}$	$\frac{R}{d}=\frac{1}{5}$	$\frac{R}{d}=\frac{1}{10}$
10	0.113	0.106	0.089	0.066
30	0.347	0.308	0.192	0.108
60	0.754	0.579	0.266	0.127
90	1.289	0.813	0.301	0.134
120	1.585	0.929	0.316	0.137
150	1.632	0.956	0.321	0.138
170	1.634	0.958	0.322	0.138

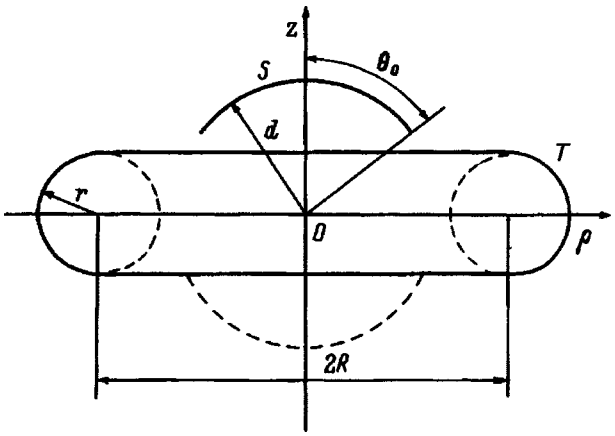


FIG. 2.

$$\int_a^b f(x)dx\approx\frac{h}{3}[f_0+f_{2n}+4(f_1+f_3+\ldots+f_{2n-1})+2(f_2+f_4+\ldots+f_{2n-2})],\qquad(39)$$

where

$$f_i=f(a+ih),\quad h=\frac{b-a}{2n},\quad i=0,1,\ldots,2n.$$

The finite SLAE was solved by Gauss's method with the choice of the principal element over the entire matrix. All the infinite sums were computed to within 10^{-5}. As was explained, to obtain a numerical solution of the integral Eq. (30) to within 0.001, it is sufficient to take $n=20$ in Eq. (39) for the geometrical parameters of the conductors under consideration.

The Legendre functions $P_{n-\frac{1}{2}}(\cosh\alpha_0)$ and $Q_{n-\frac{1}{2}}(\cosh\alpha_0)$ were computed from[6]

$$P_{n-\frac{1}{2}}(\cosh\alpha_0)=\frac{1}{\pi}\int_0^{\pi}\frac{d\varphi}{(\cosh\alpha_0+\sinh\alpha_0\cos\varphi)^{n+0.5}},$$

$$Q_{n-\frac{1}{2}}(\cosh\alpha_0)=\int_0^{\pi}\frac{\cos n\varphi d\varphi}{\sqrt{2(\cosh\alpha_0-\cos\varphi)}},$$

using the quadrature formula, Eq. (39) ($n=20$).

Table I shows the values of S_0 and S_2 for certain values of $p=r/R$, while Table II shows the values of the normalized capacitance $\tilde{C}=C/(4\pi\varepsilon d)$ for $R/r=2$ and certain values of R/d and the vertex angle Θ_0 of the thin, unclosed spherical shell S.

The electrostatic problem for the conductors shown in Fig. 2 is solved in a similar way. However, the following formula is used in this case to connect the spherical and toroidal harmonic functions:

$$r^{-n-1}P_n(\cos\Theta)=\frac{1}{2\pi c^{n+1}}\sum_{s=-\infty}^{\infty}\tilde{D}_n^s\sqrt{2(\cosh\alpha-\cos\beta)}\times Q_{s-\frac{1}{2}}(\cosh\alpha)e^{is\beta},$$

$$\sqrt{2(\cosh\alpha-\cos\beta)}P_{n-\frac{1}{2}}(\cosh\alpha)e^{in\beta}$$

$$=\sum_{s=0}^{\infty}(-1)^s c^{-s}\tilde{D}_s^n r^{-s-1}P_s(\cos\Theta),$$

where

$$\tilde{D}_n^s=2P_n(0)+\sum_{k=1}^{s}g_k^s(sP_{n+k}^k(0)-ik(n+1)P_{n+k}^{k-1}(0)).$$

[1] Yu. Ya. Iossel', É. S. Kochanov, and M. L. Strunskiĭ, *Calculating Electrical Capacitance* (Énergoizdat, Leningrad, 1981), 288 pp.
[2] V. P. Il'in, *Numerical Methods of Solving the Problems of Electrical Physics* (Nauka, Moscow, 1985), 336 pp.
[3] E. Boridy, IEEE Trans. Electromagn. Compat. **EMC-29**, 131 (1987).
[4] J. P. Chevailler, J. Electrost. **9**, 307 (1981).
[5] Ya. S. Uflyand, Zh. Tekh. Fiz. **48**, 1741 (1978) [Sov. Phys. Tech. Phys. **23**, 987 (1978)].
[6] N. N. Lebedev, *Special Functions and Their Applications* (GITTL, Moscow, 1953), 380 pp.
[7] V. Ya. Arsenin, *Methods of Mathematical Physics and Special Functions* (Nauka, Moscow, 1984), 384 pp.
[8] *Handbook of Special Functions, with Formulas, Graphs, and Mathematical Tables* (Nauka, Moscow, 1984), 832 pp.
[9] V. T. Erofeenko, Dif. Uravneniya **19**, 1416 (1983).
[10] Ya. S. Uflyand, *Methods of Paired Equations in Problems of Mathematical Physics* (Nauka, Leningrad, 1977), 220 pp.
[11] A. F. Verlan' and V. S. Sizikov, *Integral Equations: Methods, Algorithms, and Programs* (Naukova Dumka, Kiev, 1986), 732 pp.

Translated by W. J. Manthey

Features of the hydrodynamic pressure field of an electric explosion in an equilibrium gas–liquid medium

V. G. Kovalev

Institute of Pulsed Processes and Technologies, Ukrainian Academy of Sciences, 327018 Nikolaev, Ukraine
(Submitted January 8, 1997)
Zh. Tekh. Fiz. **68**, 7–12 (July 1998)

The hydrodynamic features of an electric explosion in a bubble gas–liquid mixture are studied in the equilibrium approximation of the medium for the case of fine gas bubbles when the initial size of the latter ranges from units to tens of micrometers, as is observed when actual liquids contain natural gas. In mathematically modeling the electric-explosion processes, the characteristics of the hydrodynamic field were calculated, taking into account the finite size of the plasma piston for which the quasi-wave equation with the nonlinear barotropic equation of state of the mixture was numerically integrated, using an explicit finite-difference scheme in an ellipsoidal coordinate system. It is established that the presence of gaseous inclusions manifests itself when the gas concentration is $\varepsilon_0 \gtrsim 10^{-4}$, whereas appreciable nonlinear effects appear when $\varepsilon_0 \gtrsim 5\times 10^{-3}$. © *1998 American Institute of Physics.*
[S1063-7842(98)00207-4]

As is well known, during electric-explosion energy conversion,[1,2] the dynamics of the plasma channel and the generated hydrodynamic perturbations are essentially determined by the hydrodynamic properties of the working liquid. Under actual conditions, the presence of gas bubbles can change the indicated properties and consequently can alter the conditions under which the channel develops and the emission and propagation processes of the pressure waves. As shown in Refs. 3 and 4, an electric explosion in a bubble liquid in which the gas bubbles have characteristic dimensions from units to tens of micrometers can be mathematically modeled in terms of a model of an equilibrium medium.

The dynamic properties of the plasma channel of an electric explosion in an equilibrium medium were studied in Refs. 3, 5, and 6, where it was shown that, for gas concentrations of $\varepsilon_0 \sim 0.01$, the rate at which the channel expands is a factor of 2 greater than the expansion rate in a pure liquid, with the pressure levels being approximately identical because of a negative feedback mechanism. It should be pointed out that, to solve the external hydrodynamic problem for the expanding channel, Refs. 3, 5, and 6 use the analytical dependence between the pressure and the expansion rate of the latter, obtained by solving the self-similar problem concerning the expansion of a cylindrical piston in an equilibrium gas-liquid medium.[7,8] The indicated quasi-self-similar approximation, as for the case of a pure liquid,[9,10] is valid only in the immediate vicinity of the piston and is unsuitable for calculating the field characteristics.

We shall consider the problem of determining the hydrodynamic pressure field of an electric explosion in an equilibrium gas–liquid medium. Taking into account the decisive role of bubble nonlinearity,[11–13] the system of relationships consisting of the equations of motion, and the continuity of the continuous medium in terms of which the given bubble liquid is considered, it is easy to write the equation that relates density ρ and pressure P,

$$\frac{\partial^2 \rho}{\partial t^2} = \Delta P, \tag{1}$$

which, along with the equilibrium equation of state,[14,15]

$$\frac{\rho}{\rho_0} = \left[(1-\varepsilon_0)\left(\frac{P_0+A}{P+A}\right)^{1/n} + \varepsilon_0 \left(\frac{P_0}{P}\right)^{1/\gamma} \right]^{-1}, \tag{2}$$

form a closed system. Here ρ_0 and P_0 are the unperturbed density of the mixture and the pressure, $A = 304.5$ MPa and $n = 7.15$ (for water) are the constants of the Tait equation of state, γ is the exponent of the adiabatic curve of the gas, and ε_0 is the initial volume concentration of gas.

The indicated equations can lead to one equation of hyperbolic type. The wave equation obtained when the equation of state is linear can be solved analytically in a number of cases; numerical methods are ordinarily used to solve nonlinear equations.

However, when the equation of state is expressed by a complex function or is given in tabular form, it is more convenient to directly integrate the system of Eqs. (1) and (2). In this case, the one-dimensional problem in the Euler variables can be written as

$$\rho_i^{k+1} = 2\rho_i^k - \rho_i^{k-1} + \frac{\tau^2}{h^2} \times \left[P_{i+1}^k + P_{i-1}^k - 2P_i^k + \frac{\nu}{2i}(P_{i+1}^k - P_{i-1}^k) \right], \tag{3}$$

where $\rho_i^k = \rho(ih, k\tau)$, and $P_i^k = P(ih, k\tau)$, while the values of the symmetry index $\nu = 0,1,2$ correspond to flat, cylindrical, and spherical symmetry.

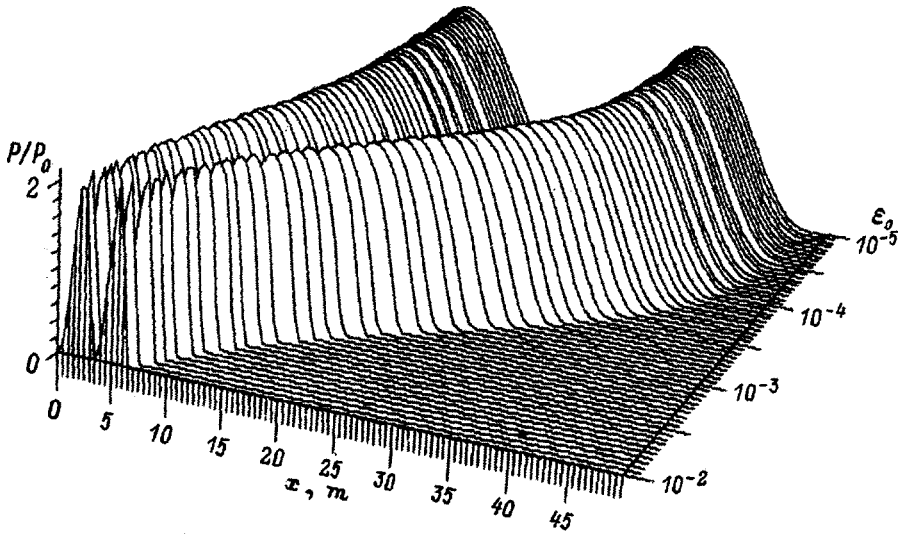


FIG. 1.

Let us consider the question of determining the value of τ. To do this, we write Eq. (1) for the case of planar symmetry and we reduce it to dimensionless form.

We introduce the dimensionless density $\overline{\rho}=\rho/\rho_0$ and pressure $\overline{P}=P/P_A$, where P_A is a pressure that is known to be greater than the maximum possible in the process under consideration. We introduce the dimensionless coordinate $\overline{x}=x/h$ and time $\overline{t}=t/\tau$, and we write Eq. (1) in the form

$$\frac{\partial^2\overline{\rho}}{\partial\overline{t}^2}=\frac{P_A\tau^2}{\rho_0 h^2}\frac{\partial^2\overline{P}}{\partial\overline{x}^2}.$$

Since $\partial^2\overline{\rho}/\partial\overline{t}^2$ and $\partial^2\overline{P}/\partial\overline{x}^2$ are of the same order of magnitude in the dimensionless equation, the stability condition for numerically integrating it has the form $P_A\tau^2/(\rho_0 h^2)\leqslant 1$. Thus, $\tau\leqslant(\rho_0/P_A)^{1/2}h$.

As is well known, when hyperbolic equations are numerically integrated, oscillations arise on the perturbation front that propagate with damping into the depth of the region encompassed by the perturbation. Various methods are used to suppress these oscillations, which reduce to introducing additional dissipative terms into the equation or to artificially smoothing the results, which is also equivalent to introducing additional dissipation. The enumerated measures, besides suppressing the oscillations of the calculation, cause blurring of the perturbation front, which is especially appreciable in the case of planar symmetry. In this case, it is suitable to use centrally symmetric smoothing with weight W to suppress the oscillations:

$$\rho_i^{k+1}=(\rho_{i-1}^{k+1}+W\rho_i^{k+1}+\rho_{i+1}^{k+1})/(W+2). \tag{4}$$

As shown by solving a test problem concerning the propagation of a planar step wave in an acoustic medium, satisfactory suppression of the oscillations is achieved even when $W\sim 10$, with smearing of the wave front being insignificant.

Let us consider the features of the propagation of a sinusoidal perturbation in an equilibrium medium. Even though this process is somewhat outside the main topic of the study, it is of interest for better understanding of the regularities of the main process and can also be regarded as an additional test of the algorithm. Let us trace the propagation of a plane acoustic wave with a frequency of 30 Hz for various amplitudes and gas concentrations in the medium, measuring the spatial profile of the wave at time $t=0.33$ sec. The choice of the perturbation frequency in this case is predicated on the need for the presence of a regime involving an equilibrium medium with wide variation of the perturbation amplitude, and also for reasons of clarity. Here and below, the calculations are carried out for water with air bubbles.

At a pressure of 0.2 MPa (Fig. 1), it can be seen that the wave velocity increases as the gas concentration decreases from 10^{-2} to 10^{-4} and is virtually constant at lower concentrations. Thus, the equilibrium medium with $\varepsilon_0<10^{-4}$ can be assumed to be a pure liquid.

The wave amplitude is virtually constant, which indicates that dissipation is absent. Since the adiabatic approximation is used in the equation of state both for the liquid and for the gas, dissipation is intrinsically absent from the model of the medium. The averaging algorithm given by Eq. (4) must introduce some dissipation in the neighborhood of the perturbation front, but, since the weighting factor $W=10$ is rather large in all the calculations, the dissipation introduced by averaging is virtually imperceptible. Finally, the wave profile changes somewhat at the maximum concentrations of gas (considered here) of $\varepsilon_0=10^{-2}$, revealing a certain twisting of the wave front, caused by the nonlinearity of the gas component.

Furthermore, calculation shows that the nonlinearity of the medium shows up much more strongly when the pressure has an amplitude of 1 MPa. When $\varepsilon_0=10^{-2}$, a discontinuity appears at the perturbation front, but the influence of nonlinearity becomes negligible even at $\varepsilon_0=10^{-3}$. As in the preceding case, the wave velocity stops increasing at $\varepsilon_0<10^{-4}$.

When the pressure has an amplitude of 10 MPa (Fig. 2), an unperturbed velocity is established even for $\varepsilon_0=10^{-3}$. The same features are valid for a pressure of 100 MPa. In both cases, the nonlinearity of the medium is caused by the

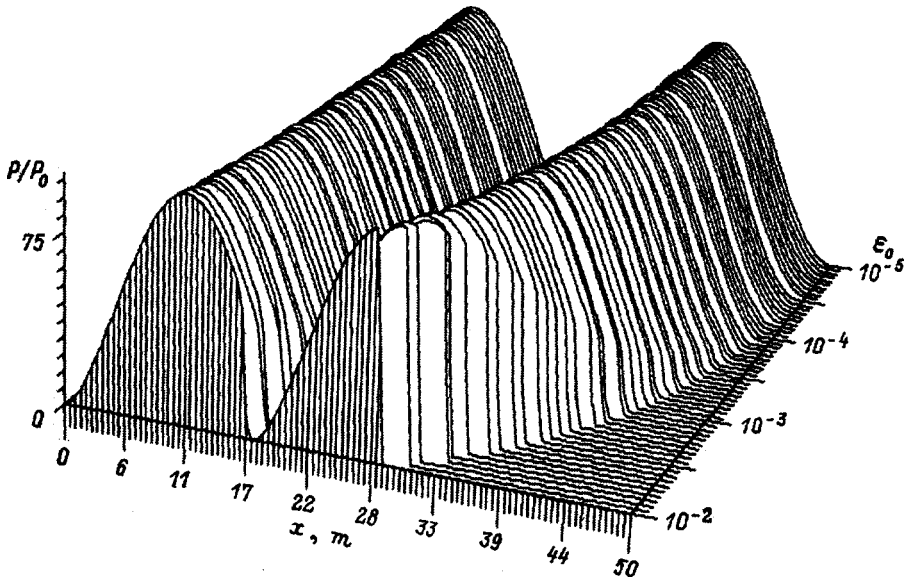


FIG. 2.

nonlinearity of its gas component and is appreciable only when the gas concentration is rather large.

There is interest in the wave profile when the pressure has an amplitude of 1000 MPa (Fig. 3). Here the effect of the nonlinearity becomes significantly stronger than at lower pressures, and this is explained by the combined effect of the nonlinearity of the gas and liquid components. When $\varepsilon_0 = 5\times10^{-3}$, the indicated combined effect causes the formation of an exponentially decaying shock wave whose amplitude is somewhat less than that of a sinusoidal perturbation. It is interesting to note that the combined nonlinearity effect does not reduce to a summation of the two components, since the nonlinearity of the pure liquid, as is seen for small concentrations, is negligible.

We now turn to a description of the wave generated in an electric explosion. As is well known, the source of hydrodynamic perturbation in most cases is a long thin finite cylinder. It is of interest to describe the wave field, taking into account such a source geometry. This can be done in an ellipsoidal coordinate system.

We introduce ellipsoidal coordinates η, ζ, ψ of a prolate ellipsoid of revolution, formed by a system of confocal ellipsoids whose semimajor axes coincide with the axis of symmetry, two-focus hyperboloids of revolution, and the half-planes that passes through the axis of symmetry. In this coordinate system, taking into account the axial symmetry, Eq. (1) has the form

$$\frac{\partial^2\rho}{\partial t^2} = \frac{1}{a^2(\eta^2-\zeta^2)}\left\{\frac{\partial}{\partial\eta}\left[(\eta^2-1)\frac{\partial P}{\partial\eta}\right]+\frac{\partial}{\partial\zeta}\left[(1-\zeta^2)\frac{\partial P}{\partial\zeta}\right]\right\}, \tag{5}$$

where a is the semimajor axis of the system of coordinate ellipsoids.

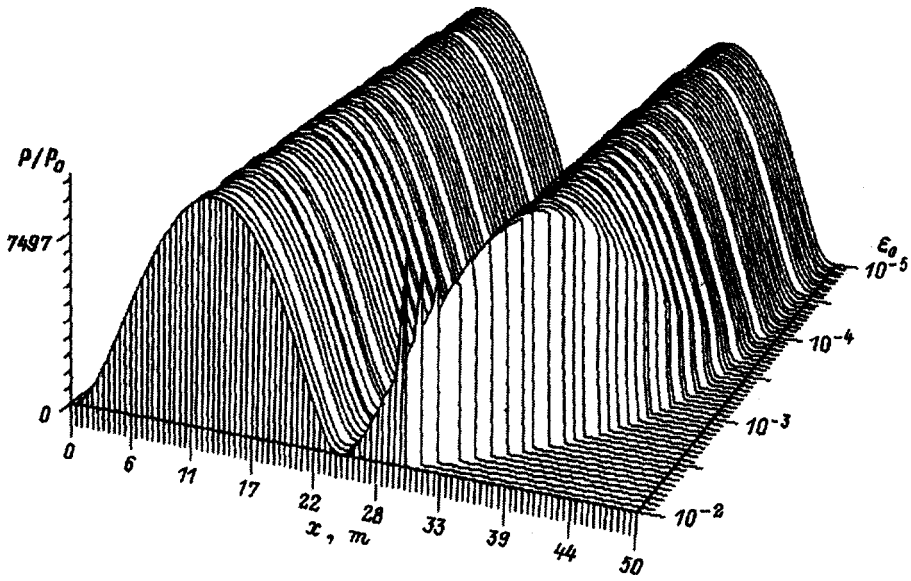


FIG. 3.

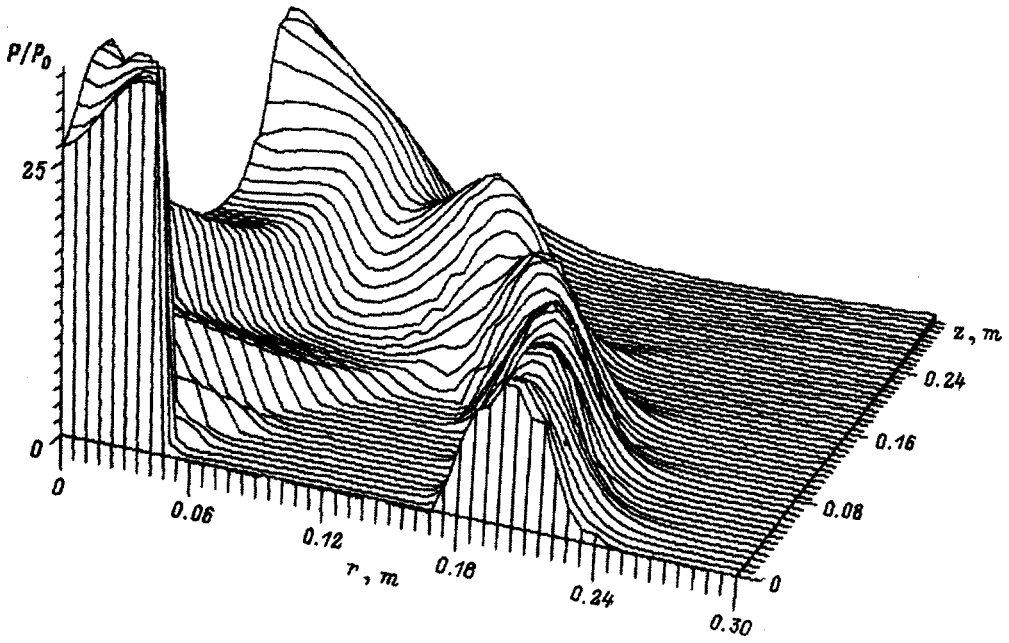


FIG. 4.

Let us consider Eq. (5) for two limiting cases: with the plane of symmetry passing through the middle of the major axis and orthogonal to it (the equatorial plane) and in the axial direction. Both directions are determined by the value of ζ, and the variable coordinate in both cases is η. In the equatorial plane, $\zeta=0$, and the Laplacian has the form

$$\Delta_1=\frac{1}{a^2\eta^2}\left[(\eta^2-1)\frac{\partial^2}{\partial\eta^2}+2\eta\frac{\partial}{\partial\eta}\right]. \tag{6}$$

On the axis of symmetry, $\zeta=1$, and the Laplacian in this case is

$$\Delta_2=\frac{1}{a^2}\frac{\partial^2}{\partial\eta^2}+\frac{2\eta}{a^2(\eta^2-1)}\frac{\partial}{\partial\eta}. \tag{7}$$

It is easy to see that both operators, as expected, transform to spherically symmetric Laplacians far from the axis, i.e., when $\eta\gg1$. It is interesting to note that Eq. (6) does not tend to the cylindrical Laplacian close to the axis, i.e., when $\eta\sim1$, since, even from a point located very close to the axis, the segment of it limited by the foci qualitatively differs from the infinite axis of the cylindrical coordinates.

Let us consider Eq. (5) in the equatorial plane:

$$a^2\frac{\partial^2\rho}{\partial t^2}=\frac{\eta^2-1}{\eta^2}\frac{\partial^2P}{\partial\eta^2}+\frac{2}{\eta}\frac{\partial P}{\partial\eta}. \tag{8}$$

As is well known, the wave equation can be factorized in a plane of symmetry. In the case of spherical symmetry, the equation factorizes for the quantity rP. Thus, in the plane of

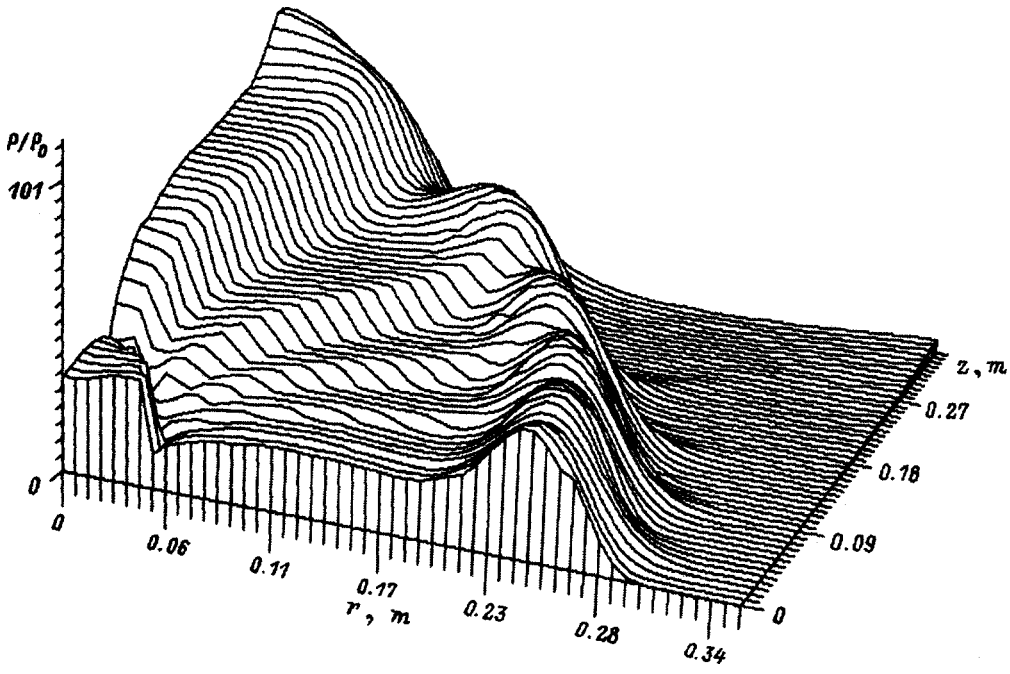


FIG. 5.

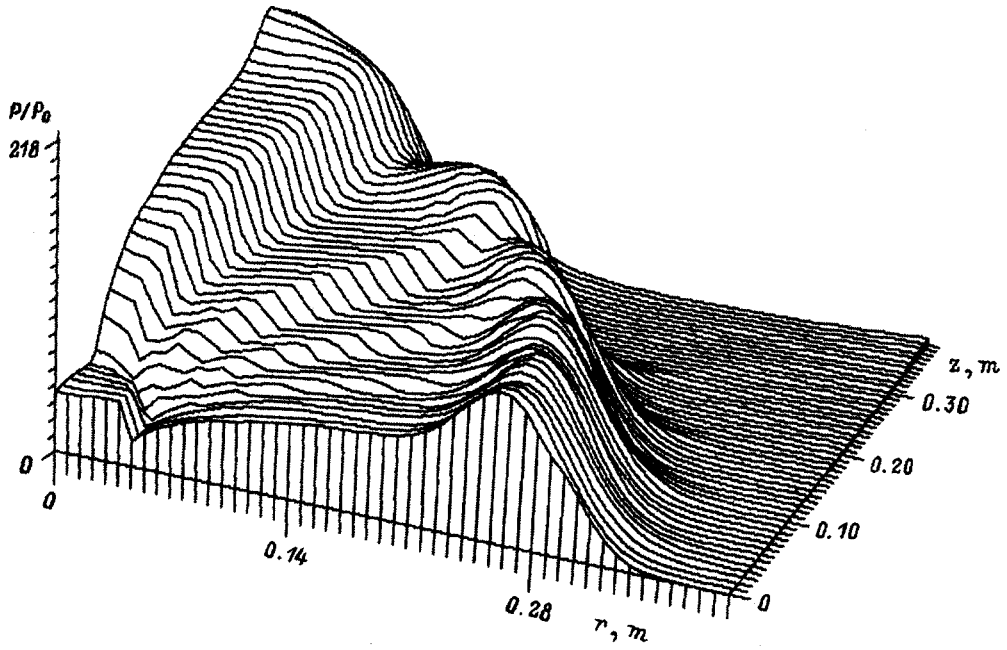


FIG. 6.

symmetry, the solution of the wave equation breaks down into outgoing and incoming waves. A similar situation, with allowance for a scale factor $1/r$, is also valid for spherical symmetry. However, Eq. (8), like the equation for cylindrical symmetry, does not factorize; consequently, even in a linear medium with a constant sound velocity, there is no wave as such, i.e. no perturbation that maintains its shape as it propagates.

Thus, Eqs. (6) and (7) can be used to solve the pressure-distribution problem in two important asymptotic cases: in the axial and radial directions of a finite cylindrical channel. However, bearing in mind the determination of the total structure of the hydrodynamic pressure field, we turn to the integration of Eq. (5).

First of all, we choose certain whole numbers N_1 and N_2 and introduce a dimensionless coordinate net with steps $h=1/N_1$ and $l=1/N_2$. We define the discrete coordinates η_i and ζ_j by the expressions

$$\eta_i=ih; \quad i=N_1, \ldots, \infty; \quad \zeta_j=jl; \quad j=0,1, \ldots, N_2.$$

We introduce the dimensional step in time τ and define the discrete time $t_k=k\tau$. Equation (5) in discrete variables has the form

$$\rho_{i,j}^{k+1}=2\rho_{i,j}^{k}-\rho_{i,j}^{k-1}+\frac{\tau^2}{a^2(i^2h^2-j^2l^2)}[(i^2-N_1^2)\times(P_{i-1,j}^{k}+P_{i+1,j}^{k}-2P_{i,j}^{k})+i(P_{i+1,j}^{k}-P_{i-1,j}^{k})+(N_2^2-j^2)(P_{i,j-1}^{k}+P_{i,j+1}^{k}-2P_{i,j}^{k})-j(P_{i,j+1}^{k}-P_{i,j-1}^{k})].$$

The characteristics of the electric-explosion channel are calculated using Eqs. (2)–(4) given in Ref. 16 [or Eqs. (1)–(3) from Ref. 17]. The pressure of the source-channel of the electric explosion is given on a line:

$$P_{N_1,j}^{k}=P_k; \quad j=0, \ldots, N_2.$$

The pressure calculated in the η, ζ coordinates is transformed to the cylindrical coordinate system $r=a\sqrt{(\eta^2-1)(1-\zeta^2)}$, $z=a\eta\zeta$ and is represented graphically in the form of the instantaneous spatial pattern at a definite instant t^*.

As shown by varying the parameters of the discharge circuit (U_0 is the voltage of the capacitor bank, C is the capacitance of the latter, L is the inductance, and l_0 is the interelectrode gap), their values appreciably affect the dynamic characteristics of the plasma channel of the electric explosion, but make little difference in the qualitative character of the wave field in the medium. The main parameter that substantially affects the features of the wave field is the initial volume gas concentration ε_0. Therefore, to study the structure of the wave field and how the gas-containing medium affects that structure, we chose a discharge regime in which the discharge-circuit parameters were $U_0=22$ kV, $C=22$ μF, $L=5$ μH, and $l_0=40$ mm.[6] The gas concentration was varied from 10^{-5} to 10^{-2}, and $t^*=2$ ms.

Passing to the results of the mathematical modeling of an electric explosion in an equilibrium gas–liquid medium, we should point out the following: As is well known, the pressure emitted by the discharge channel has the character of a pulse with a steep front and a rather rapid, approximately exponential falloff. For a discharge in a pure liquid (Fig. 4, where the calculation corresponds to $\varepsilon_0=10^{-5}$), the pressure field as a whole maintains such a character, and the falloff becomes much slower only because of geometrical divergence. Obvious geometrical causes explain the fact that the pressure is approximately twice as great in the axial direction as in the radial direction.

The pressure field has virtually the same character for an electric explosion in a bubble medium with $\varepsilon_0=10^{-3}$ (Fig.

5). The effect of the gaseous inclusions manifests itself only in a certain reduction of the pressure.

However, the character of the hydrodynamic pressure field qualitatively changes when the gas concentration reaches $\varepsilon_0 = 10^{-2}$ (Fig. 6). In this case, a self-clarifying effect is observed as a result of the essentially nonlinear character of the absorption in the medium. The essence of this is that the pressure falloff is greater, the less is the pressure itself, since the elasticity of the gaseous inclusions substantially increases at high pressure. As a result, the pressure peak propagates with comparatively little absorption, but, on a section of the falloff beginning with a certain threshold pressure, the absorption increases so much that the pressure drops virtually to zero.

Thus, by using an elliptical coordinate system and thereby taking into account the finite size of the source itself, it has been shown that the pressure is significantly greater in the axial direction than in the radial direction. The gas inclusions have virtually no effect when $\varepsilon_0 \leqslant 10^{-5}$. In the interval $\varepsilon_0 = 10^{-4}$–10^{-3}, it manifests itself only in a reduction of the pressure but does not alter the profile of the propagating waves. The character of the wave field changes qualitatively only when $\varepsilon_0 \geqslant 10^{-2}$ and shows up in a sharp increase of nonlinear absorption.

The author is grateful to N. M. Beskaravaĭnyĭ for help in carrying out the calculations.

[1] K. A. Naugol'nykh and N. A. Roĭ, *Electric Discharges in Water* [in Russian] (Nauka, Moscow, 1971), 155 pp.
[2] E. V. Krivitskiĭ and V. V. Shamko, *Transition Processes for High-Voltage Discharges in Water* [in Russian] (Naukova Dumka, Kiev, 1979), 208 pp.
[3] V. G. Kovalev, *Theory, Experiment, and Practice of Electric-Discharge Technologies* [in Russian] (Naukova Dumka, Kiev, 1995), No. 2, pp. 14–19.
[4] V. G. Kovalev, *Abstracts of the Seventh Scientific School on the Physics of Pulsed Discharges in Condensed Media* [in Russian] (Nikolaev, 1995), p. 122.
[5] V. G. Kovalev, *Abstracts of the Second All-Russia Seminar on the Dynamics of Spatial and Nonequilibrium Flows of Liquid and Gas* [in Russian] (Miass, 1993), pp. 58–60.
[6] V. G. Kovalev, in *Dynamics of a Continuous Medium* [in Russian], Collected Scientific Works of the Institute of Hydrodynamics, No. 110, 111–116 (1995).
[7] V. G. Kovalev, Pis'ma Zh. Tekh. Fiz. **20**(18), 61 (1994) [Tech. Phys. Lett. **20**, 750 (1994)].
[8] R. I. Nigmatulin, *Dynamics of Multiphase Media* [in Russian] (Nauka, Moscow, 1987), Pt. 2, 60 pp.
[9] I. S. Okun', Zh. Tekh. Fiz. **41**, 292 (1971) [Sov. Phys. Tech. Phys. **16**, 219 (1971)].
[10] N. M. Beskaravaĭnyĭ, V. V. Dykhta, V. G. Kovalev, and V. V. Tul'skiĭ, *Applied Hydrodynamics of an Electric Explosion* [in Russian] (Kiev, Naukova Dumka, 1992), 200 pp.
[11] V. K. Kedrinskiĭ, Zh. Prikl. Mekh. Tekh. Fiz., No. 4, 29 (1968).
[12] V. E. Nakoryakov, B. G. Pokusaev, and I. R. Shreĭber, *Wave Propagation in Gas–Liquid and Vapor–Liquid Media* [in Russian] (Inst. Tekh. Fiz, Novosibirsk, 1983), 238 pp.
[13] K. A. Naugol'nykh and L. A. Ostrovskiĭ, *Nonlinear Wave Processes in Acoustics* [in Russian] (Nauka, Moscow, 1990), 237 pp.
[14] V. A. Pozdeev, N. M. Beskaravaĭnyĭ, and V. G. Kovalev, *Pulsed Disturbances in Gas–Liquid Media* [in Russian] (Naukova Dumka, Kiev, 1988), 116 pp.
[15] Sh. U. Galiev and O. P. Panova, Problemy Prochnosti, No. 10, 49 (1995).
[16] N. M. Beskaravaĭnyĭ, V. G. Kovalev, and E. V. Krivitskiĭ, Zh. Tekh. Fiz. **64**(2), 197 (1994) [Tech. Phys. **39**, 224 (1994)].
[17] V. G. Kovalev, Zh. Tekh. Fiz. **66**(4), 24 (1996) [Tech. Phys. **41**, 308 (1996)].

Translated by W. J. Manthey

Structure and parameters of the shock layer formed when a supersonic underexpanded jet interacts with a counterpropagating hypersonic flow in the transition regime

N. Yu. Bykov and G. A. Luk'yanov

Institute of High-Performance Computations and Databases, 194291 St. Petersburg, Russia
(Submitted February 12, 1997)
Zh. Tekh. Fiz. **68**, 13–18 (July 1998)

The method of molecular dynamics (the Bird system) has been used to mathematically model a planar, strongly underexpanded supersonic jet that encounters a hypersonic flow of rarefied gas. Particular attention is paid to the structure and parameters of the shock layer close to the plane of symmetry. The results of calculations are presented for currents of a monatomic gas, simulating argon, with a Mach number of the external flow of $M_\infty = 5.48$, a Mach number at the nozzle edge of $M_a = 1$, a ratio of the density at the nozzle edge to the density of the unperturbed flow equal to 130, and various stagnation temperatures of the external flow and of the jet. The evolution of the structure and the parameters of the shock layer as the Knudsen number Kn_∞ varies from 0.02 to 0.35 is considered. The results are compared with the data calculated for the shock layer when argon flows around thermally insulated cylinders. The main features and regularities of the relaxation of the translational degrees of freedom of the gas for external and jet flows are considered. Data are presented on the form of the distribution function over velocities and its evolution as gas moves through the shock layers.
 [S1063-7842(98)00307-9]

When a strongly underexpanded supersonic jet flows counter to a uniform hypersonic flow, the current whose diagram is shown in Fig. 1 appears. In terms of the model of a continuous medium, in describing a gas flow of constant composition on the basis of the Navier–Stokes equations, the decisive parameters for the given type of flows include the Mach numbers M_∞ and M_a, the Reynolds number Re_∞, defined in terms of the parameters of unperturbed flow and the characteristic current dimension L, the Reynolds number Re_a, defined in terms of the parameters at the nozzle edge and its size (height h for a flat nozzle), the ratio $n = p_a/p_\infty$ for the jet discharge (p_a and p_∞ are the pressures at the nozzle edge and in the unperturbed external flow) or the ratio of the corresponding densities $m = \rho_a/\rho_\infty$, the temperature factor $\tau = T_{0a}/T_{0\infty}$ (T_{0a} and $T_{0\infty}$ are the stagnation temperature for jet flow at the nozzle edge and the temperature of the unperturbed external flow), the Prandtl number Pr, the ratio of the specific heats γ, and also the temperature dependences of the viscosity and thermal conductivity.

Figure 1a shows the interaction pattern of an underexpanded supersonic jet with a counterpropagating hypersonic flow when $Re_\infty \to \infty$ and $Re_a \to \infty$. When a supersonic flow runs into a jet that is an obstacle for the external flow, bow shock wave *1* appears. The external and the jet flows are separated by contact surface *2* (the dividing streamline). The jet flowing out of the nozzle with height h has a shock-wave structure characteristic of an underexpanded supersonic jet. Central shock wave *3* (the Mach disk) arises in front of the contact surface in the jet. Between the bow shock wave and the contact surface, a shock layer is formed from the gas of the external flow, with thickness Δ_1. Between the contact surface and the Mach disk, a shock layer is formed from the gas of the jet, with thickness Δ_2. The Mach disk and suspended shock wave *4* bound the free-expansion region of jet *5*, within which the flow parameters are determined exclusively by the parameters of the gas at the nozzle edge and are independent of the flow parameters in the space surrounding this region. The boundary of jet *6* separates the jet flow and separation zone *7* by a complex flow, determined by the conditions downstream. Streamlines *8* qualitatively illustrate the character of the flow in the shock layer and the other regions.

The external flow and the jet have a common critical point k with an identical stagnation pressure $p'_{0\infty} = p'_{0c} = p_{0k}$. In the case of the hypersonic shock layer considered here, $p'_{0\infty}$ is determined for the flow of an ideal gas by the approximate relationship[1]

$$p'_{0\infty} \simeq \rho_\infty u_\infty^2 (1 - 1/\varepsilon), \tag{1}$$

where $\varepsilon = (\gamma+1)/(\gamma-1)$, and u_∞ is the velocity of the unperturbed external flow.

In the case of the flow of a strongly underexpanded jet considered here ($n \gg 1$), the Mach number in front of the Mach disk is large, and p'_{0c} is determined similarly:

$$p'_{0c} \simeq \rho_c u_c^2 (1 - 1/\varepsilon), \tag{2}$$

where ρ_c and u_c are the density and velocity of the jet flow in front of the Mach disk.

Neglecting the thickness of the shock layer, we get from Eqs. (1) and (2) that the equation $p'_{0\infty} = p'_{0c}$ is equivalent to the approximate equation

$$\rho_\infty u_\infty^2 \simeq \rho_{ck} u_{ck}^2, \tag{3}$$

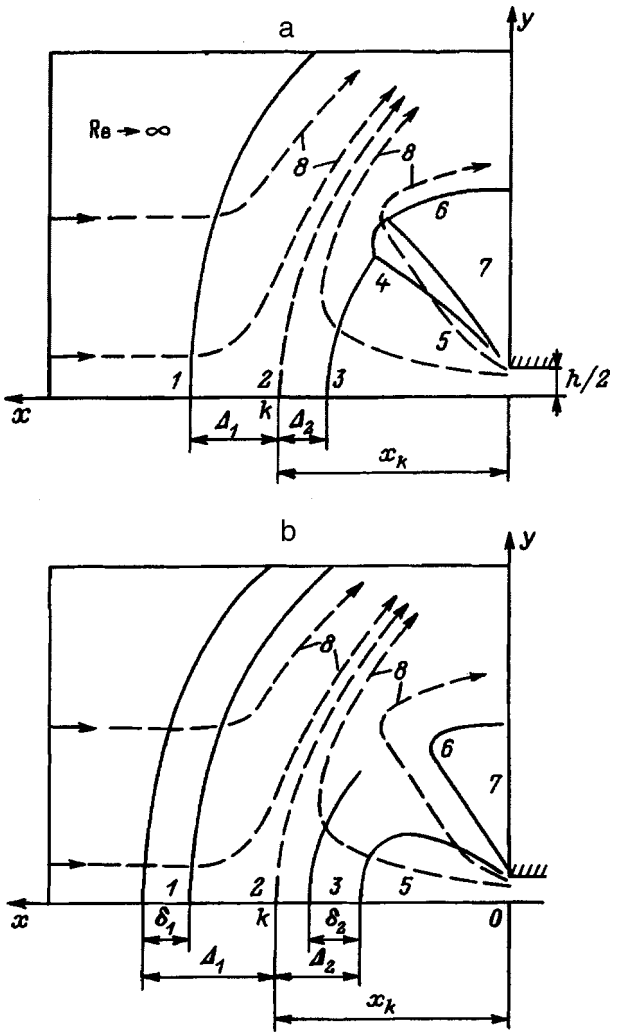


FIG. 1. Flow diagram when a strongly underexpanded jet interacts with a counterpropagating hypersonic flow. (a) Reynolds number $\mathrm{Re}_\infty \to \infty$, (b) transitional flow regime.

where ρ_{ck} and u_{ck} are the density and velocity of the jet flow at point $x = x_k$ in the symmetry plane of the freely expanding jet.

Parameters ρ_{ck} and u_{ck} can be expressed in terms of the parameters at the nozzle edge. When $x_k/h \gg 1$, the flow in the free-expansion region of the jet has the character of the flow from a source with a pole at the nozzle edge, and the relationships for determining ρ_{ck} and u_{ck} have the form[2]

$$u_{ck} \simeq u_{cm} = \sqrt{2\gamma RT_{0a}/(\gamma-1)}, \quad \rho_{ck}/\rho_a \simeq Bh/x_k. \qquad (4)$$

Here u_{cm} is the limiting thermodynamic velocity of the jet flow, and $B = B(\gamma, M_a)$. Using Eq. (4), from Eq. (3) we get an expression for determining x_k/h:

$$x_k/h \simeq B(u_{cm}^2/u_a^2)(\rho_a u_a^2/(\rho_\infty u_\infty^2)). \qquad (5)$$

For an arbitrary value of M_a, the ratio u_{cm}^2/u_a^2 is a function of M_a and γ. When $\mathrm{M}_a = 1$, we have $u_{cm}^2/u_a^2 = \varepsilon$. Thus, for $M_\infty \gg 1$, $n \gg 1$, and given values of M_a and γ, the parameter $i = \rho_a u_a^2/(\rho_\infty u_\infty^2)$ is the similarity parameter. When it is used, parameter n is eliminated from the list of parameters to be determined. Under these circumstances, the position of the critical point k is determined by the approximate relationship

$$x_k/(hi) \simeq B\varepsilon. \qquad (6)$$

The goal of this study is to explain the physical features of the formation of the shock layer for the hypersonic counterpropagating interaction of a uniform flow and a planar strongly underexpanded jet in the transitional (according to the Knudsen number $\mathrm{Kn}_\infty = \lambda_\infty/L$, where λ_∞ is the free path length in an unperturbed external flow) regime. Particular attention is devoted to studying how the number Kn_∞ (or $\mathrm{Re}_\infty = 1.25\sqrt{\gamma}\mathrm{M}_\infty/\mathrm{Kn}_\infty$) affects the structure and the flow parameters in the shock layer.

The flow pattern in the transition regime, shown in Fig. 1b, corresponds to the case in which the thickness of the bow shock wave, δ_1, and of the Mach disk, δ_2, are part of the corresponding shock layers. Increasing the number Kn_∞ densifies the shock layers, gradually blends the shock wave with the compression layer, and reduces the size of the free-expansion region. The dividing streamline is not the interface of the external and the jet flow. Particles of the external flow penetrate into the jet shock layer and, conversely, particles of the jet penetrate into the shock layer of the external flow. There is substantial interest in this case in the structural features of the shock layer as a whole and in the degree of nonuniformity of the flow in it, determined by the evolution of the distribution function over velocities as gas moves in the shock layer.

By analogy with the problem of flow around a cylinder with diameter d, it is natural to take $L = 2x_k$ as a characteristic size of the flow considered here. The Knudsen number in this case takes the form $\mathrm{Kn}_\infty = \lambda_\infty/(2x_k)$.

The Monte Carlo method of direct statistical simulation is used to study the flow. Bird's NTC scheme[3] was used to simulate the collisions of the particles. The simulation took into account only elastic collisions. The mechanics of the collisions corresponded to the VHS model,[3] in which the collision cross section σ is described by

$$\sigma = \pi d^2, \quad d = d_0(c_0/c)^\nu, \qquad (7)$$

where d is the effective particle diameter, and d_0 is the effective diameter when the relative velocity of the colliding particles is $c = c_0$.

The exponent is $\nu = \omega - 0.5$, where ω is the exponent in the temperature dependence of the dynamic viscosity ($\nu = 0.25$).

A rectangular nonuniform network with thickening in the regions adjacent to the nozzle edge was used in the calculations. The steps of the network in the different subregions were less than the characteristic free path length in the flow field, except for the cells immediately adjacent to the nozzle edge, where their size was of the same order of magnitude as λ_a (λ_a is the free path length at the nozzle edge). For the first two calculated versions (see Table I, the number of cells in the region was 5400, while the number of modeling atoms exceeded 150 000. In version 3, 21 600 cells and more than 300 000 ''computer'' particles were considered.

TABLE I. Starting data for versions of the calculation.

No. of the version	M_∞	M_a	τ	Kn_a	Re_a	i	n	m
1	5.48	1	1	0.5	3.2	35.6	1070	130
2	5.48	1	1	0.167	9.6	35.6	1070	130
3	5.48	1	1	0.033	48.5	35.6	1070	130

The cell size and the number of modeling particles were chosen according to special methodological studies. Doubling the number of cells while maintaining the number of modeling particles had no effect on the results of the modeling of the fields of the gas-dynamic parameters within the limits of relative error at the 5% level. Increasing the number of modeling particles by a factor of m while maintaining the number of cells reduced the relative statistical error of the modeling of the density field approximately proportionally to $\sqrt{m}$. For the results presented below, the largest relative error of the modeling of the density field within the shock layers does not exceed 3–5%,

On the left-hand boundary of the calculated region and at the nozzle edge (Fig. 1), the state of the gas was determined by Maxwellian functions with parameters (v_∞, T_∞, v_a, T_a), corresponding to the numbers M, τ, and n from Table I. Conditions for mirror reflection of the particles were established at the lower boundary of the calculated region (the plane of symmetry). The boundary conditions at the upper boundary corresponded to the conditions at infinity. Particles that reached the right-hand boundary of the region were excluded from further consideration. All the calculations were carried out on a Convex C-3800 parallel-vector computer.

In terms of the direct statistical simulation method, the dimensionless characteristics of the flow considered here are determined by the parameters M_∞, M_a, Kn_a (Re_a), i (either n or m), and τ and by the model of the collisions of the particles. This problem is distinctive in that, unlike the problem of flow around a solid, here the characteristic size of the flow (which determines the flow pattern as a whole) is not known ahead of time, and the Knudsen number Kn_∞ is consequently unknown.

The simulation was carried out for three versions of the starting values, shown in Table I.

Figure 2 shows graphs of v/v_∞, n/n_∞ and T/T_∞ along the critical streamline. The distance was measured from the nozzle edge in the relative quantities $x_0 = x/x_k$, which ensures that the stagnation points coincide for all the versions of the calculation. The position of the critical point $x_k/(hi)$ depends on the rarefaction of the flow. For regimes 1, 2, and 3, the number $Kn_\infty = \lambda_\infty/(2x_k)$ corresponds to 0.35, 0.1 and 0.02 ($Re_\infty = 25$, 88 and 140). Increasing the number Kn_∞ in this range slightly displaces the stagnation point toward the nozzle [for versions 1, 2, and 3 of the calculation, $x/(hi)$ equals 1.53, 1.68, and 1.87, respectively]. Version 3 ($Kn_\infty = 0.02$) corresponds to a flow regime with a rather distinct structure of the shock layers, with the thicknesses δ_1 and δ_2 of the shock waves being substantially less than the corresponding thicknesses Δ_1 and Δ_2 of the shock layers (Fig. 1).

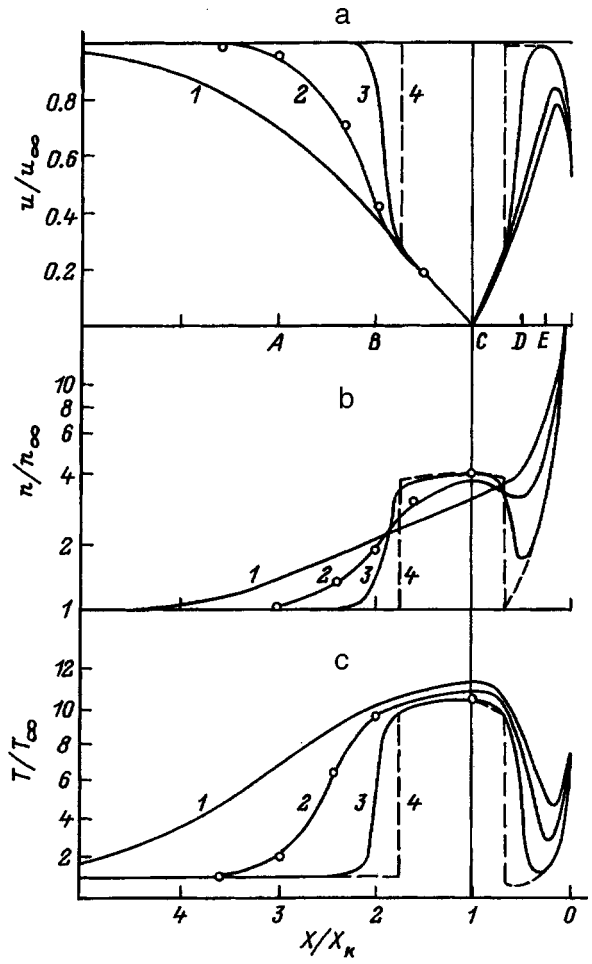


FIG. 2. Comparison of the u (a), n (b), and T (c) profiles in generalized coordinates x/x_k with the analogous profiles for $Re_\infty \to \infty$ (*4*) and the results of a calculation of the flow around a thermally insulated cylinder. *1–3* show the versions of the calculations from Table I; ○ is 5

This regime is an example of a transitional regime that bounds continuum flow regimes. Version 1 of the calculation ($Kn_\infty = 0.35$) corresponds to the totally smeared regime of the shock layer,[4] in which it is impossible in the shock layer to distinguish the structure of the shock wave. This regime is characterized by a monotonic density variation along the x coordinate. Version 2 of the calculation ($Kn_\infty = 0.1$) corresponds in terms of the rarefaction to a transitional regime in which $\delta < \Delta$. This regime is characterized by the appearance of the first signs of gas compression in the shock layer on the density profiles.

The profiles in the shock layer of the external flow are characterized by long tails extending upstream and typical of strong shock waves. The expansion of the jet shock layer toward the nozzle is limited by the density increase of the jet flow. Therefore, as the number Kn_∞ increases, Δ_1 increases substantially faster than does Δ_2.

Dashed curves *4* in Fig. 2 correspond to the parameter profiles in the compressed layers as $Kn_\infty \to 0$, obtained from the Rankine–Hugoniot equations. The velocity profiles in the neighborhood of the stagnation point ($u < 1.4\sqrt{2RT_\infty}$) are virtually linear and coincide for all versions of the calculation. The expansion of the jet shock layer strongly reduces the

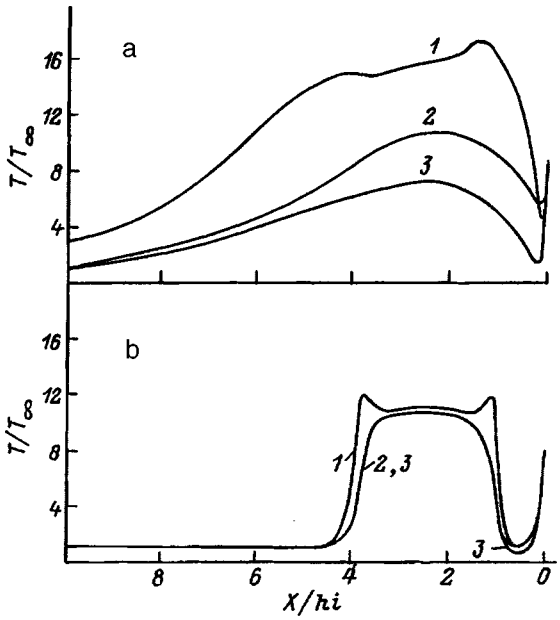


FIG. 3. Profiles of the components of the kinetic temperature T_x (*1*), T_z (*2*) and T_y (*3*) on the stagnation line. (a) shows version 1, (b) shows version 3.

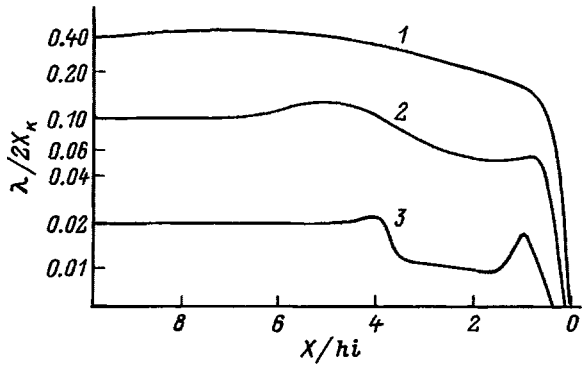


FIG. 4. Variation of the mean free path length λ along the stagnation line. *1–3*—number of the version of the calculation.

free-expansion region of the jet and changes the maximum velocity (the Mach number) in the jet flow. When $\mathrm{Kn}_\infty = 0.02$, the compression n/n_∞ and the heating T/T_∞ of the gas in the shock layer for external flow virtually reach values corresponding to $\mathrm{Kn}_\infty \to 0$. The compression and heating in the jet shock layer are appreciably less than the values obtained from Rankine–Hugoniot equations, and this is explained by the gradient character of the jet flow. As the number Kn_∞ increases, the compression in the shock layer of the external flow decreases, while the heating increases.

There is substantial interest in comparing the results of the calculations shown in Fig. 2 with similar calculations for the case of flow by argon around a thermally insulated cylinder in the transition regime. Such calculations are carried out in Ref. 5 by direct statistical simulation.

A comparison showed that there is fairly good agreement between the results of the calculations of the u, n and T profiles in the shock layer of the external flow virtually to the stagnation point for all the numbers Kn_∞ considered here $[\mathrm{Kn}_\infty = \lambda_\infty/(2x_k) = \lambda_\infty/d]$. Figure 2 shows the results of the calculations of Ref. 5 for flow around a thermally insulating cylinder with $\mathrm{Kn}_\infty = 0.1$. Some difference in the n/n_∞ and T/T_∞ values is observed only in the neighborhood of the stagnation point.

The flow in the shock layer has an essentially nonequilibrium character. The degree of nonequilibrium over the translational degrees of freedom is illustrated by data concerning the profiles of the kinetic temperature components T_x, T_y, and T_z $[T = (T_x + T_y + T_z)/3]$ for the versions of the calculation corresponding to $\mathrm{Kn}_\infty = 0.35$ and 0.02 and shown in Figs. 3a and 3b, respectively. When $\mathrm{Kn}_\infty = 0.35$, the degree of nonequilibrium is extremely great, encompasses the entire compressed layer, and spreads far upwards in the external and jet flow. The physical cause of translational nonequilibrium in the shock layer is the low collision frequency of the particles when the gas stagnates. When stagnation occurs and the kinetic energy of directed mass motion is transformed into the kinetic energy of thermal motion of the molecules, it is more efficient to transfer energy into a longitudinal (relative to the streamline) translational degree of freedom. Because of the flat geometry of the flow $T_z > T_y$, since $\partial T/\partial z = 0$, while $\partial T/\partial y < 0$, and there is an outflow of heat from the stagnation line in the direction of the y axis. It should be pointed out that the opposite result is obtained in Ref. 5: $T_z < T_y$ (for a relatively small difference in T_z and T_y). In our opinion, this result of Ref. 5 is erroneous.

The degree of nonequilibrium in the external flow along the stagnation line varies nonmonotonically. The ratios T_x/T_y and T_x/T_z vary from units in the unperturbed flow to some maximum value in the forward part of the shock layer and then decrease as the stagnation point is approached. When $\mathrm{Kn}_\infty = 0.35$, the maximum of the ratios T_x/T_y and T_x/T_z equals approximately three and is reached when $x/(hi) \simeq 10$. In the neighborhood of the stagnation point, $T_x/T_y \simeq 2.3$ and $T_x/T_z \simeq 1.5$.

Within the free-expansion region adjacent to the nozzle edge, the character of the translational nonequilibrium has a different physical nature. The equilibrium between the different translational degrees of freedom (and the corresponding difference between T_x, T_y, and T_z) breaks down because the collision frequency of the particles decreases in the process of supersonic expansion and acceleration of the gas.[2,3] In this case, the longitudinal temperature T_x ''freezes.'' As shown by calculations,[2,3] for axisymmetric expansion, the temperature corresponding to the distribution of the thermal velocities perpendicular to the mass motion is close to equilibrium. The main cause that this temperature component decreases downstream when collisions of the particles are rare is the geometrical factor of the flow expansion.[2] In a flat jet, the expansion factor acts only with respect to T_y, while the T_z component in the region of rare collisions ''freezes,'' with the minimum value of T_z somewhat larger than the minimum T_x (Fig. 3). The degree of nonequilibrium in the free-expansion region of the jet monotonically increases downstream.

Translational nonequilibrium thus develops quite differently in the jet shock layer than in the shock layer of the external flow. A gas with some degree of translational non-

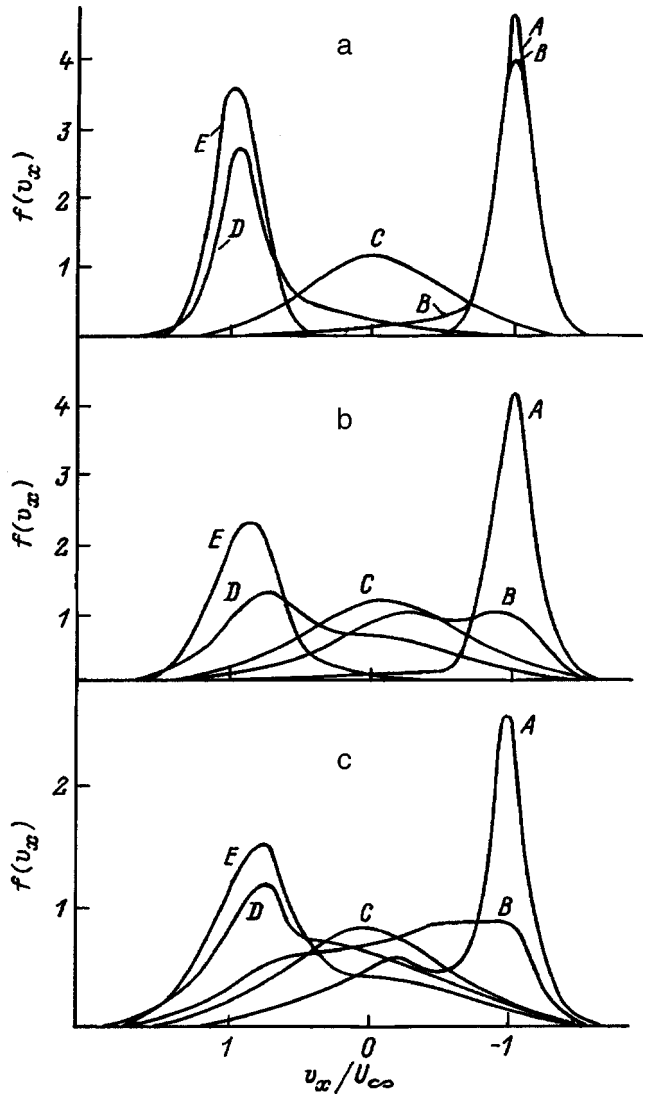


FIG. 5. Evolution of the distribution function $f(v_x)$ over longitudinal velocities when the gas moves through the shock layers along the stagnation line. (a)–(c) show versions 3, 2, and 1 of the calculations, respectively; *A*, *B*, *C*, *D*, and *E* are points on the stagnation line in Fig. 2.

equilibrium, determined by the conditions of the free jet expansion, runs into the front of the jet shock layer. The stagnation of the gas and the rapid and anticipating increase of T_x (relative to T_y and T_z) begins in the jet shock layer. Within a relatively narrow forward zone, the translational relaxation mechanism inherent to a supersonic freely expanding flow changes to the relaxation mechanism inherent to the stagnation of a gas in a shock layer. The presence of a T_x peak for all the versions of the calculation is typical of strong shock waves.[3,5] For version 3 of the calculation ($\mathrm{Kn}_\infty = 0.02$), deviations from equilibrium are observed virtually only in zones occupied by shock waves.

Figure 4 shows data concerning the variation of the mean free path length λ along the stagnation line. When $\mathrm{Kn}_\infty = 0.02$, the compression of the gas in the shock layer corresponds to decreasing λ by about a factor of 2. As Kn_∞ increases, the range of variation of λ within the shock layer decreases.

The data shown in Fig. 5 represent the character of the

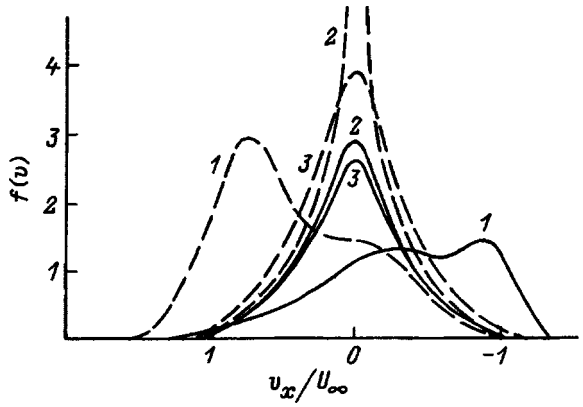


FIG. 6. Comparison of the distribution functions over velocities v_x (*1*), v_y (*2*) and v_z (*3*) at points *B* (solid curves) and *D* (dashed curves) for version 2 of the calculation.

evolution of the distribution function $f(v_x)$ of the particles over velocities as the gas moves through the shock layers along the stagnation line. For versions 1–3 of the calculation, Figs. 5a–5c show the smoothed values at points *A*, *B*, *C*, *D* and *E* (Fig. 2) with coordinates $x/x_k = 3$, 2, 1, 0.5, and 0.25, respectively. When $\mathrm{Kn}_\infty = 0.02$ (Fig. 5a), the distribution function $f(v_x)$ at points *A* and *E*, corresponding to unperturbed external flow and the free-expansion region of the jet, is Maxwellian. At points *B* and *D*, corresponding approximately to the center of the shock waves, the distribution functions become asymmetric, and rather long tails of particles with longitudinal velocities $v_x/u_\infty > -0.5$ for point *A* and $v_x/u_\infty < 0.5$ for point *D* appear in them. The tails also contain particles that possess longitudinal velocities directed opposite to the direction of mass motion. At stagnation point *C*, the distribution function is close to Maxwellian.

When $\mathrm{Kn}_\infty = 0.1$ (Fig. 5b), points *A* and *E* correspond to the front part (in the direction of motion of the gas) of the shock layers. The deviation of $f(v_x)$ from a Maxwellian distribution function is relatively small, but a tail has already begun to form. Points *B* and *D* correspond approximately to the centers of the corresponding shock layers. The distribution functions are very different from the equilibrium functions and have a distinctly bimodal character (they have two maxima). At point *C*, $f(v_x)$ has a Maxwellian form.

When $\mathrm{Kn}_\infty = 0.35$, points *A* and *E* lie inside the shock layers. The distribution functions at these points are substantially different from Maxwellian, possess powerful and extensive tails, and have a bimodal character. The deviation from equilibrium increases at points *B* and *D*, but the form of the distribution function over longitudinal velocities at point *C* is, as before, quasi-Maxwellian.

The degree of anisotropy of the distribution functions over velocities v_x, v_y and v_z is illustrated by the data in Fig. 6. For $\mathrm{Kn}_\infty = 0.1$, this figure shows the form of the distribution functions $f(v_x)$, $f(v_y)$, and $f(v_z)$ at points *B* ($x/x_k = 2$) and *D* ($x/x_k = 0.5$). Unlike the explicitly nonequilibrium and asymmetric functions $f(v_x)$, functions $f(v_y)$ and $f(v_z)$ have a quasi-Maxwellian form. The difference between $f(v_y)$ and $f(v_z)$ determines the difference between T_y and T_z at the given points on the stagnation line.

[1] Yu. P. Golovachev, *Numerical Modeling of Viscous Gas Flows in a Shock Layer* [in Russian] (Nauka, Moscow, 1996), 376 pp.
[2] V. G. Dulov and G. A. Luk'yanov, *Gas Dynamics of Effusion Processes* [in Russian] (Nauka, Novosibirsk, 1984), 234 pp.
[3] G. A. Bird, *Molecular Gas Dynamics and the Direct Simulations of Gas Flows* (Clarendon Press, Oxford, 1994), p. 451.
[4] W. D. Hayes and R. R. Probstein, *Hypersonic Flow Theory* [Academic Press, New York, 1966; IL, Moscow, 1962, 607 pp.].
[5] D. R. Grawford and F. W. Vogenitz, in *Proceedings of the Ninth International Symposium on Rarefied Gas Dynamics, Göttingen, 1974*, p. B24.

Translated by W. J. Manthey

Pressure dynamics during shock loading of aqueous foams

E. I. Vasil'ev, S. Yu. Mitichkin, V. G. Testov, and Hu Haibo

Mechanics Institute, M. V. Lomonosov Moscow State University, 117192 Moscow, Russia
(Submitted November 6, 1996; resubmitted February 16, 1998)
Zh. Tekh. Fiz. **68**, 19–23 (July 1998)

The interaction of shock waves with aqueous foams has been numerically modeling on the basis of a gas–drop model. A modified Godunov scheme of second-order accuracy has been used to solve the gas-dynamics equations. The measured density distributions of the foam over height and the results of calculational modeling of the foam structure were used as starting data. The results of the pressure calculations are compared with experimental oscilloscope traces.
 [S1063-7842(98)00407-3]

INTRODUCTION

The first stage of investigations of the propagation of shock waves (SWs) in aqueous foams involved studying the damping properties of a foam screen.[1,2] It was established that the pressure behavior behind the SW front is rather complex. To interpret the experiments, a simplified representation of the foam in the form of a ''pseudogas'' with an effective adiabatic index Γ was used in most cases. Such an approach sometimes makes it possible to estimate the equilibrium parameters behind a transmitted wave. Subsequent investigations showed that syneresis and the structural characteristics of the foam have a substantial effect on SW propagation,[3,4] but modeling in a vertical shock tube is preferable. The use of the modified Godunov scheme, which is monotonic and provides second-order accuracy, makes it possible to calculate the behavior of the pressure in the equilibrium region with a good degree of accuracy.[4] To describe the pressure behavior in the relaxation zone, it is necessary to create mathematical models of the foam and of its deformation and breakdown processes, accompanied by the formation of a gas–drop medium. There are no direct theoretical studies of the breakdown of foam structure, and the existing experimental results[3,5,6] do not give a complete representation of the breakdown mechanism. The use of photographic recording showed that, when SWs with Mach numbers of $M_i > 1.3$ are incident on a foam column, at the beginning of the relaxation zone in foams with a density of $\sim 30\ kg/m^3$, the foam breaks down rather quickly and can be regarded as a gas–drop medium whose initial parameters are determined by the structure of the foam and by syneresis. The study of the structural characteristics and the structure of the syneresis model make it possible to give the initial parameters of the gas–drop medium: the concentration of large and small drops, their sizes, and the degree of polydisperseness.

This paper discusses the possibility of using a gas–drop model to numerically model SW propagation along a vertical column of aqueous foam. The SW in this case is incident on the foam column from above and therefore moves in the direction of increasing density of the foam.

MODEL CONCEPTS

In the flow behind the SW, in the gas–drop medium formed after the foam breaks down, stripping and shattering of the drops, heat exchange between the liquid and vapor phases, and evaporation and condensation of the liquid phase occur, as well as coalescence of the drops. It is assumed that the formation of small drops is accompanied by their rapid acceleration and by heating to the temperature of the carrier gas. The large drops are ensembles of independent particles; because of the difference of the velocities, small drops are deposited onto the large ones.

In our experiments, the Mach number was $M_i < 1.5$, and the foam was rather dense. Evaporation is insignificant in this case, and we shall model the foam by a medium consisting of three components: a collection of liquid drops (the large drops) characterized by position x_k, mass m, velocity v, temperature T_k, and diameter d; an ideal gas with constant heat capacity, density ρ_g, pressure p_g, velocity u, and temperature T_g; and a collection of relatively small drops, which are in velocity and temperature equilibrium with the gas and are characterized by density ρ_p. The last component is mathematically described as a second gaseous component, with its own heat capacity. The Lagrange representation of a finite number of large drops (Lagrange coordinate ξ) is used to describe the motion of the large drops, and the Euler representation (with x being the coordinate along the tube) is used to describe the mixture of gas and small drops. The equations for the mass, momentum, and total energy of noninteracting large drops for homogeneous flow have the form

$$\frac{d}{dt}x_k = v, \tag{1}$$

$$\frac{d}{dt}m = j_0 + j_1, \tag{2}$$

$$\frac{d}{dt}mv = f_0 + j_0 v + j_1 u, \tag{3}$$

$$\frac{d}{dt}(mv^2/2 + mc_l T_k) = f_0 v + \varepsilon + j_0(v^2/2 + c_l T_k) + j_1(u^2/2 + c_l T_g). \tag{4}$$

Equation (2) takes into account two mass-exchange mechanisms: j_0 is the intensity with which the incoming flow entrains mass in the form of small drops from the surface of an individual large drop (stripping); j_1 is the deposition onto a large drop of small drops that fall into a large ''protector'' drop. In Eq. (3), f_0 is the frictional force with which the gas acts on a large drop. In the energy equation, Eq. (4), when the gas and the large drops have different temperatures, heat exchange ε is present, and c_l is the heat capacity of the liquid phase.

We regard the mixture of the gas with the small drops as a single-temperature and single-velocity medium, which we call an effective gas, with density $\rho=\rho_g+\rho_p$.

The symbols on the right-hand sides of Eqs. (2)–(4), as g_2, g_3, and g_4, describing the behavior of the small drops and the gas, have the form

$$\frac{\partial}{\partial t}\rho_p+\frac{\partial}{\partial x}\rho_p u=-ng_2, \tag{5}$$

$$\frac{\partial}{\partial t}\rho_g+\frac{\partial}{\partial x}\rho_g u=0, \tag{6}$$

$$\frac{\partial}{\partial t}\rho u+\frac{\partial}{\partial x}(\rho u^2+p_g)=-ng_3, \tag{7}$$

$$\frac{\partial}{\partial t}\rho e+\frac{\partial}{\partial x}(\rho e+p_g)u=-ng_4. \tag{8}$$

The concentration n of large drops is determined from the initial concentration $n_0(\xi)$ and their current position

$$n=n_0\frac{\partial}{\partial x}\xi.$$

The adiabatic index Γ of the effective gas depends on the local densities of the gas and the small drops, $\Gamma=1+\rho_g R/(\rho_g c_v+\rho_p c_1)$ (Here R is the universal gas constant, and c_v is the heat capacity of the gas at constant volume). The expression for the volume energy density of the effective gas is $e=u^2/2+p_g/[\rho(\Gamma-1)]$.

To solve the equation, it is necessary to be given the initial conditions in front of the SW for the gas, the density of the small drops, and the mass and concentration of large drops. However, it is more convenient to give the initial distribution of the diameter of the large drops and the density ρ_k of the liquid of the drops, smeared over the volume of the entire mixture ($\rho_k=\alpha\rho_l$, where ρ_l is the density of the liquid, and α is the volume concentration of large drops). The expressions for the mass and concentration of large drops with diameter d have the form $m=\rho_l\pi d^3/6$ and $n=\rho_k/m$.

An expression for the intensity with which the gas flow entrains mass from the surface of a drop by the stripping mechanism, satisfactorily describing the experimental data obtained in Refs. 7 and 8, is

$$j_0=-k\sqrt{0.125\,d^3|u-v|\rho_g^{1/3}},$$

where Engel's constant k (the stripping factor) for the mixtures of air and vapor with drops of water often encountered in practice is $k=1.3-2\ \mathrm{kg^{5/6}/(m^{3/2}\cdot sec^{1/2})}$.

The main criterion that characterizes the condition for the stripping mechanism to occur is the Weber number $\mathrm{We}=p_g d(u-v)^2/\sigma$. Drops begin to be stripped in a gas flow only when the external forces exceed the internal forces by a definite amount ($\mathrm{We}>\mathrm{We}_c$).[9] In this case, under breakdown conditions, we neglect heat exchange ε.

References 9 and 10 recommended the conditions for determining the critical Weber number We_c: $\mathrm{We}_c\mathrm{Re}^{-0.5}=k'$, where $k'=0.5-1.5$, the Reynolds number is $\mathrm{Re}=\rho_g d|u-v|/\mu_g$, and μ_g is the dynamic viscosity of the gas.

An estimate of the Weber number shows that, provided that $\mathrm{M}_i=1.35-1.50$, the stripping mechanism occurs in a wide range of variation of the drop diameter. When $\mathrm{M}_i=1.20$, the stripping mechanism takes place only for drops whose diameter exceeds 30 μm, whereas, when $\mathrm{M}_i=1.05$, stripping does not occur in the entire range of variation of the drop diameter considered here.

The intensity j_1 of the precipitation of small drops onto a large one and the force of friction f_0 between a large drop and the gas are determined from[9]

$$\begin{aligned} j_1&=0.25\pi d^2\rho_p|u-v|,\\ f_0&=0.125\pi d^2\rho_g C_d|u-v|(u-v). \end{aligned} \tag{9}$$

In the case of deformation and shattering of the drop, the following relationship can be used for the resistance factor C_d of the drop:[9]

$$C_d=\begin{cases} 24.0/\mathrm{Re} & \mathrm{Re}\leqslant 0.49,\\ 27.0\cdot\mathrm{Re}^{-0.834} & 0.49\leqslant\mathrm{Re}\leqslant 80,\\ 0.27\cdot\mathrm{Re}^{0.217} & 80\leqslant\mathrm{Re}\leqslant 10^4,\\ 2.0 & 10^4\leqslant\mathrm{Re}. \end{cases}$$

The intensity of the heat flow to the surface of the drop is[9]

$$\varepsilon=\begin{cases} \pi d\lambda_g Nu(T_g-T_k), & \mathrm{We}\leqslant\mathrm{We}_c,\\ 0, & \mathrm{We}>\mathrm{We}_c, \end{cases}$$

where λ_g is the thermal conductivity of the gas; Nu is the Nusselt number, which is determined from the Rants–Marshall formula $\mathrm{Nu}=2+0.6\mathrm{Re}^{1/2}\mathrm{Pr}^{1/3}$; and Pr = 0.708 is the Prandtl number for air.

Equations (1)–(8) were numerically solved by the Lagrange–Euler method with second-order accuracy in space and time, using a modification of the Godunov method.[4]

RESULTS OF MODELING THE INTERACTION OF THE SWs WITH THE GAS–DROP MIXTURE

The density distribution of the foam in height, the fraction of the liquid contained in small drops, and the size of the large drops were given as the initial conditions in the foam. The first of these parameters is given from the experimentally measured density distribution over height of the foam column.[11] In the bottom part of the foam column, with a height less than 20 mm, where the average density was not measured experimentally, the density distribution takes into

account numerical calculations of the syneresis.[12] The second of the initial parameters, i.e., the fraction of small drops, is given from information obtained when the structure of the foam is modeled.[12] It was assumed that the small drops are formed when the foam films break down. The last parameter, i.e., the initial diameter d_0 of the large drops, is determined on the basis of the dependence on the local density of the foam and equals the transverse size in the middle of the channel.[12]

For a height of the foam column of $H=190$ mm, a holding time of the foam of $T=2$ min, and an initial mean density of the foam of 30 kg/m^3, the density distribution ρ_f of the foam, the diameter of the large drops, and the fraction of liquid contained in the small drops varies with height h as follows:

h(mm) = 190 171 152 133 114 95 76 57 38 19 15 11 8 4 0,
$d_0(\mu\text{m})$ = 15 18 22 28 30 31 32 33 34 35 38 41 45 50 70,
ρ_f(kg/m^3) = 3 6 12 16 20 23 25 27 29 32 40 56 75 98 120,
$\rho_p/\rho_f(\%)$ = 16 14 12 11 11 10 10 10 10 9 8 5 4 3 2.

Since the SWs last for a rather long time for the given configuration of the shock tube, the process of interaction of the SWs with the foam column can be separated from the subsequent process of interaction with the rarefaction wave. From the calculational viewpoint, isolating the interaction process of an SW of infinite duration with the foam column from the formulation of the problem as a whole makes it possible to concentrate the calculational modeling in the region of interaction of the SW with the foam. In such a version, the calculation is carried out only for the end part of a shock tube 400–500 mm long containing the foam column. The intensity of the SW in this case is given directly in terms of the Mach number M_i.

The following values of the thermodynamic parameters of the phases were used in the calculation: $T_{g0}=293$ K, $P_{g0}=1.01\times10^{-5}N/m^2$, $\rho_{g0}=1.21\ kg/m^3$, $\gamma_{g0}=1.4$, $c_{g0}=7045$ m^2/(sec$^2\cdot$deg), $\mu_{g0}=1.85\times10^{-5}$kg/(m$\cdot$sec), $\lambda_{g0}=2.5\times10^{-2}kg\cdot$m/(sec$^3\cdot$deg), $\rho_l=10^3 kg/m^3$, $c_l=4180$ m^2/(sec$^2\cdot$deg), and $\sigma_l=0.033$ N/m.

A comparison of the calculated pressure curves with oscilloscope traces for $M_i=1.35$ is shown in Fig. 1. The pressure sensors were located 258 (sensor *1*) and 143 mm (sensor *2*) from the end. It can be seen that the calculated pressures both behind the transmitted wave and behind the reflected compression waves (Fig. 1a) agree fairly well with the experimental values. Varying the stripping factor k within the limits indicated above has virtually no effect on the results of the pressure calculations.

Discrepancies occur in the foam at sensor *2* (Fig. 1b) in part of the front of the transmitted compression wave and in the maximum pressure amplitude at the end sensor *3* (Fig. 1c) behind the reflected wave.

The calculated transmitted compression wave front at the second sensor (50 mm from the foam–air interface) was more abrupt than the experimental wave front, on which it is possible to distinguish a two-step pressure increase: a precursor and a relaxation zone. Variation of the precipitation factor [introduced into Eq. (9)] showed that this parameter has little effect on the growth rate of the pressure at the transmitted wave front at sensor *2*. To make the calculated and experimental results agree in this zone, the foam breakdown process must be taken into account.

At the initial stage of the reflection, the calculated behavior of the pressure at the end sensor agrees well with experiment, although the maximum calculated pressure is somewhat greater than the experimental value. The instant at which the calculated pressure reaches its peak somewhat anticipates the experimental value, and the subsequent falloff occurs more rapidly. The influence of the precipitation factor was also considered in the calculation. Decreasing the precipitation factor (from unity) increases the maximum pressure at the end, and the reflected compression wave front at all the other sensors becomes more abrupt in this case, which does not agree with experiment. To study the effect of the direct incidence of large drops on the end wall, the dynamic pressure of these drops was taken into account in the calculation. It became clear that this fact also has virtually no effect on the overall behavior of the pressure.

The gas–drop medium formed by the breakdown of the

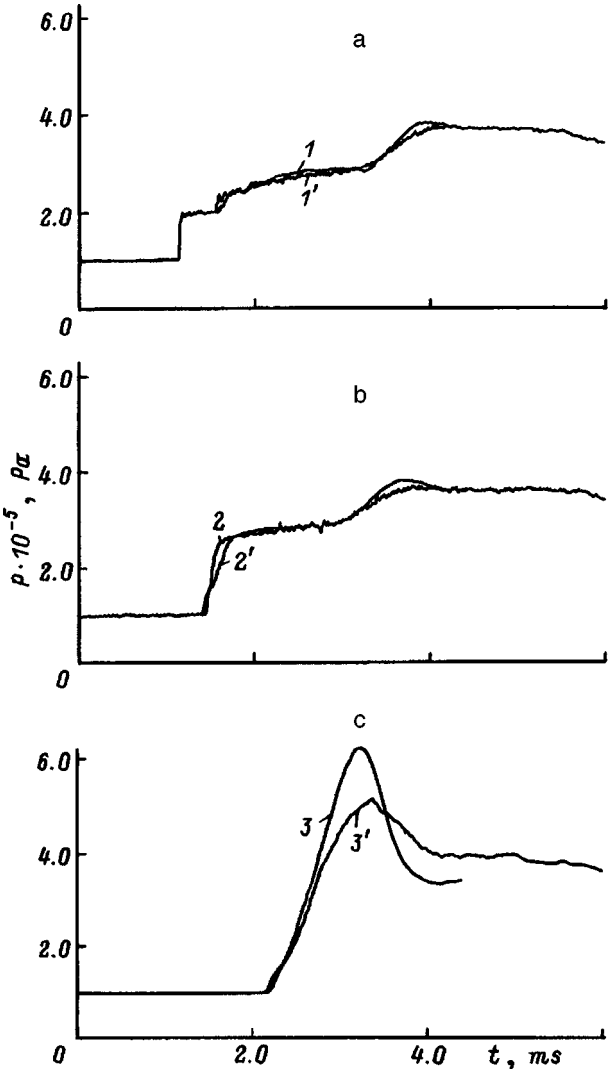


FIG. 1. Curves and oscilloscope traces of pressure *P*. *1, 2, 3* are the sensor numbers, *1–3*—calculation, *1′–3′*—experiment.

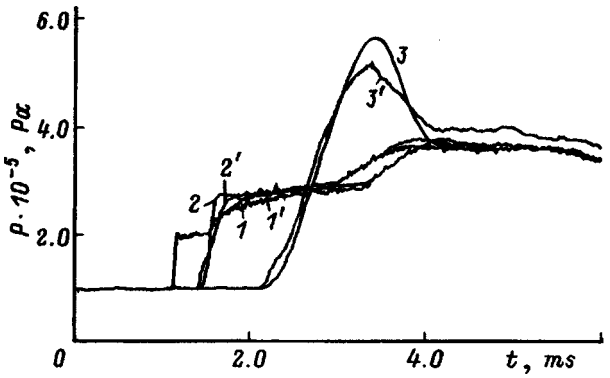


FIG. 2. Graphs and oscilloscope traces of the pressure. *1–3*—calculation for a model polydisperse mixture, *1′–3′*—experiment.

foam structure by SWs is not a monodisperse but a polydisperse mixture in the size of the drops. It was assumed in the next series of calculations that the mixture consists of several groups of drops with different initial sizes and concentrations. The Lagrange coordinates are given for each group of drops. Because the velocities differ in the flow, drops of different size overtake each other and quickly mix in space, and a polydisperse medium is actually formed almost immediately after the compression wave front arrives.

Let us consider a calculational example that models a polydisperse medium in which all the initial conditions except for the initial diameter of the large drops are the same as for Fig. 1. In each cross section by height, there are groups of drops with dimensions $4d_0(h)/3$, $d_0(h)$, and $0.5d_0(h)$ and equal mass concentrations.

Figure 2 shows calculated graphs of the pressure for the model polydisperse mixture and the corresponding experimental oscilloscope traces. The calculations agree better with the experimental results for the model polydisperse mixture than for a monodisperse mixture.

An $x-t$ diagram corresponding to the process described above (the equal-density lines of the gas phase) is shown in Fig. 3. Points *1–3* on the x axis correspond to the positions of the corresponding sensors. By comparison with the effective-gas model,[4] the boundary of the foam column moves downstream more slowly before encountering the compression wave reflected from the end. After it interacts with the compression wave reflected from the end, the boundary of the foam column is almost stationary and lies about 40–50 mm from the end.

The gas–drop model of the interaction of SWs with foam described above makes it possible under the conditions considered here to numerically predict the experimental results. It should be pointed out that, when SWs interact with foam, the density of the foam in the bottom part, unlike the other part of the foam column, sharply increases to $\rho_f > 100\ kg/m^3$. Calculations for the model polydisperse gas–drop mixture when the lower part of the foam column is excluded and the local density of the foam does not exceed 36 kg / m^3 showed that the lower layer of high-density foam has virtually no effect on the behavior of the pressure. Large drops with a size of $d > 50\ \mu$m in the bottom layer are apparently unable to reach equilibrium with the gas, both in the transmitted compression waves and in those reflected from the end wall. In the absence of intense stripping with them and intense heat transfer because of the relatively small specific surface, the large drops of the bottom region can behave relatively neutrally and have no appreciable effect on the flow. However, such a neutral layer can play a certain role in compression at the end wall. It can be assumed that, when the compression is sufficiently intense and a bottom layer of foam with high density is present at the end, a layer of ''bubbly liquid'' is formed that causes, after the phases of pressure increase and relief at the end, a prolonged excess of pressure relative to the equilibrium value (oscilloscope tracing *3′* in Fig. 2).

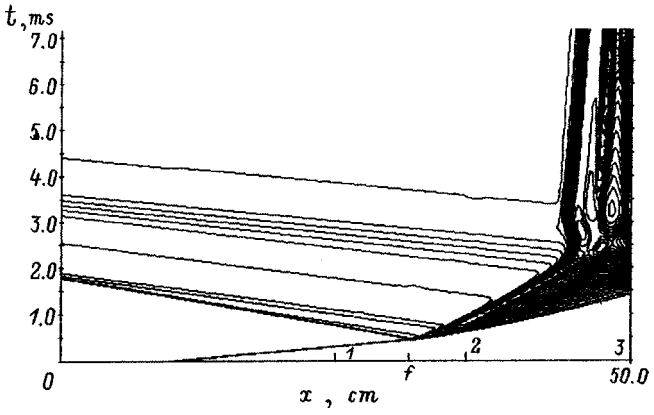


FIG. 3. $x-t$ diagram of SW propagation in foam (f is the initial position of the boundary of the foam).

CONCLUSION

1. Under the conditions considered here, the gas–drop model of the interaction of SWs with foam gives a fairly good description of the behavior of the pressure.

2. The calculated relaxation zone in the transmitted wave is shorter than in experiment, since the processes of deformation and breakdown of the foam structure are neglected in the model.

This work was carried out with the support of the Russian Fund for Fundamental Research. The authors are grateful to V. A. Levin for support.

[1] V. M. Kudinov, B. I. Palamarchuk, B. E. Gel'fand, and S. A. Gubin, Prikl. Mekh. **13**, No. 3, 92 (1977).
[2] A. A. Borisov, B. E. Gel'fand, V. M. Kudinov *et al.*, Acta Astron. **5**, 1027 (1978).
[3] A. B. Britan, I. N. Zinovik, S. Yu. Mitichkin, and V. G. Testov, Zh. Tekh. Fiz. **66**(2), 1 (1996) [Tech. Phys. **41**, 111 (1996)].
[4] E. I. Vasil'ev, S. Yu. Mitichkin, V. G. Testov, and Khu Khaĭbo, Zh. Tekh. Fiz. **67**(11), 1 (1997) [Tech. Phys. **42**, 1241 (1997)].
[5] A. B. Britan, I. N. Zinovik, and V. A. Levin, Fiz. Goreniya Vzryva **28**, No. 5, 108 (1992).

[6] A. B. Britan and I. N. Zinovik, Zh. Prikl. Mekh. Tekh. Fiz. **35**, No. 5, 78 (1994).
[7] A. A. Ranger and J. A. Nicholls, AIAA J. **7**, 285 (1969).
[8] O. Engel, J. Res. Natl. Bur. Stand. **60**, 245 (1958).
[9] A. I. Ivandaev, A. G. Kutushev, and R. I. Nigmatulin, Itogi Nauki Tekh. VINITI. Ser. MZhG. **16**, 209 (1981).
[10] R. A. Dickerson and T. A. Coultas, AIAA Paper No. 611, 24 (1966).
[11] V. A. Kulikovskii, V. A. Levin, S. Yu. Mitichkin *et al.*, Dok. Akad. Nauk **345**, 607 (1995).
[12] S. Yu. Mitichkin, A. V. Pertsov, V. G. Testov, and Hu Haibo, Kolloid. Zh. (1997) (in press).

Translated by W. J. Manthey

High-pressure magnetron sputtering: gas-phase processes

A. G. Znamenskiĭ and V. A. Marchenko

Institute of Problems of Microelectronics Technology and Especially Pure Materials, Russian Academy of Sciences, 142432 Chernogolovka, Moscow Region, Russia

(Submitted April 28, 1997)
Zh. Tekh. Fiz. **68**, 24–32 (July 1998)

Thermalization of the energetic atoms emitted by a target at pressures up to 100 Pa causes heating and motion of the gaseous medium. The temperature and velocity of the gas have been measured as a function of the magnetron-discharge parameters. This paper presents the dependences of the deposition rate, the thickness profile, and the film structure on the pressure and the discharge power. It shows that they are well described by a simple diffusion model of the transport of the thermalized atoms of the target that takes into account the motion and heating of the gaseous medium. © *1998 American Institute of Physics.* [S1063-7842(98)00507-8]

INTRODUCTION

Elevated pressures (tens of pascal) are frequently used in ion-plasma sputtering to reduce the energy of the particles that arrive at the surface of the growing film. This regime has been most widely used when depositing films of metastable compounds and ultrathin layered structures (see Ref. 1 and the citations therein) and when the discharge contains negative ions accelerated in the dark cathode space in the direction of the substrate to energies corresponding to the target potential.[2] The last case includes sputtered films of multicomponent oxides: ferroelectrics, optical materials, high-temperature superconductors, etc. The pressures used in this case reach 50 Pa,[3,4] and the deposition rates are low, of the order of 10^{-2}nm·sec^{-1}. To monitor and optimize the film-growth process under these rather difficult conditions, it is useful to consider the features of the transport of atoms from the target to the substrate at elevated pressures.

When the free path length of the atoms is much less than the target–substrate distance, the particles emitted by the target collide with the gaseous medium. They lose momentum and energy (heating the medium and causing it to move) and thermalize, and their subsequent transport is accomplished by diffusion. Part of the diffusing sputtered atoms are absorbed by the target, which reduces their flow to the substrate. Heating of the gas reduces its density (at a given pressure) and increases the thermalization length proportionally to the temperature. The moving medium carries the diffusing atoms away from the target. The last two effects must increase the deposition rate onto the substrate. It is possible to estimate under what conditions they are substantial by the following means: Let d denote the thermalization length, measured from the target surface (the symbols d_{zy} will also be used below for the distance between objects, where the subscripts denote the following object: t for the target, s for the substrate, and p for the probe). The diffusion time to the target from this distance is $\tau \approx d^2/D$, where D is the mutual diffusion coefficient (depending on the temperature and pressure). If the medium and the diffusing atoms are displaced from the target by a distance $u\tau \approx d$ in this time, where u is the velocity of the gas, the motion of the medium will have an appreciable effect on the sputtering rate. When $ud/D \gg 1$, all the atoms will be displaced toward the substrate, there will be no back-deposition onto the target, and the sputtering rate will approach the rate for the collisionless regime. In the opposite limit, the motion of the medium will have no effect on the sputtering rate.

In the literature, the sputtering rate in the collision regime of particle transport is calculated both by numerical methods that directly calculate the trajectories of the individual atoms[5,6] and in the diffusion approximation, where the thermalized atoms form a diffusion source distributed in space.[7–10] It is assumed that the gaseous medium is stationary and that its temperature is homogeneous in space, which is justified for small discharge powers. The most detailed concept of a virtual diffusion source is developed in Ref. 9. Based on a Monte Carlo calculation of the collisional motion of the sputtered atoms in the reaction gas and an experimental measurement of their concentration as a function of the distance to the target for a hollow-cathode discharge, the following results, in particular, are obtained: When the free path length of the sputtered atoms is much less than the target–substrate distance, the dropout ratio of atoms that return to the target without thermalizing and do not participate in diffusional transport is independent of the pressure and equals $G(\mu) \approx 0.09(3 - \ln\sqrt{\mu})$, where μ is the mass ratio of the target and gas atoms. The flux of nonthermalized atoms from the target to the substrate can be neglected when the target–substrate distance is $d_{ts} \geqslant 4d$. This condition also determines the limit in pressure of the diffusional transport region. The concentration of diffusing atoms in space is mainly determined by the position of the maximum of the distribution function of the extended diffusion source. Replacing it with a δ function makes no significant change in the concentration.

Experimental data on the dynamics and heating of the gas during sputtering are available only for pressures $P < 10$ Pa. A reduction of the gas density close to a planar

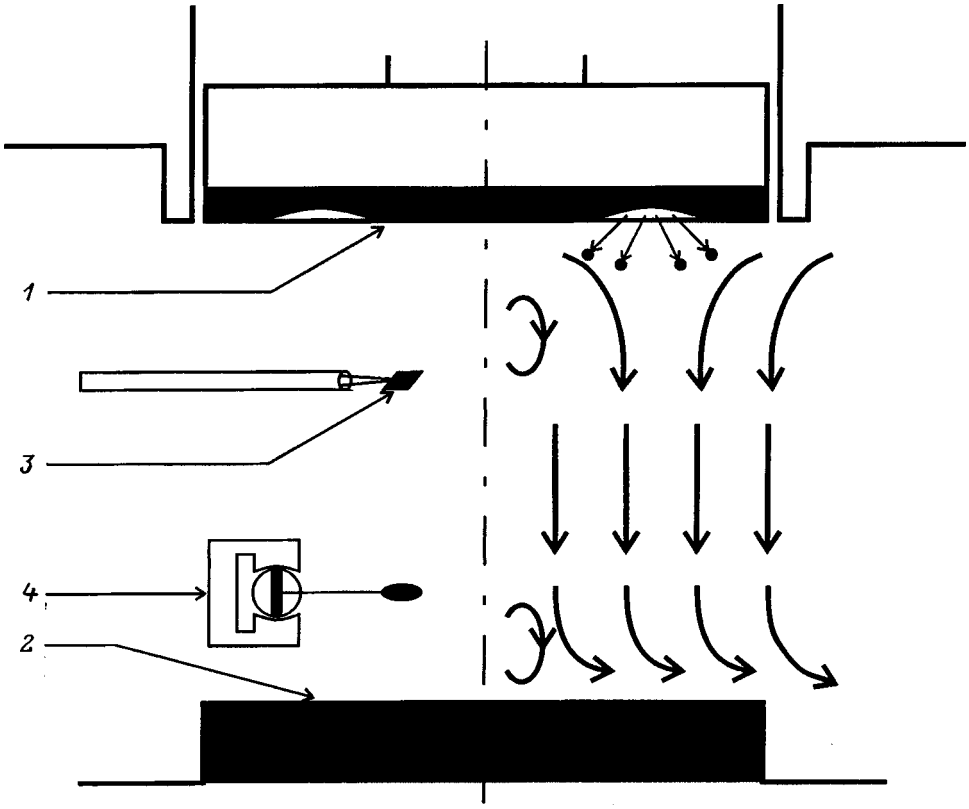


FIG. 1. The sputtering system. The right-hand side of the figure shows the direction of the gas flows.

magnetron because of heating was observed in Refs. 11 and 12, and a wind during sputtering from a cylindrical magnetron was observed in Ref. 13. These experiments were carried out at relatively low pressures of 0.2–4 Pa; the heating of the gas had little effect on the deposition rate onto the substrate.[11] The data of Ref. 10 show that the deposition rate at a pressure of 9 Pa increases faster than does the power, and the authors associate this with heating of the gaseous medium. The experimental data on the transport of sputtered atoms at pressures of up to 4 Pa are generalized in Ref. 12.

As far as we know, the temperature and velocity distributions of the gas during magnetron sputtering and their influence on the deposition rate at elevated pressures (tens of pascal) have not been studied. Moreover, stagnation of the sputtered atoms in the gas, equivalent to cooling of a vapor of these atoms, can supersaturate the vapor and cause the gaseous phase to condense — a phenomenon that has been thoroughly studied in the production technology of ultradisperse powders.[14] The particles that are formed, as they settle out on the surface of the film, change its structure.

To clarify how essential the processes enumerated above are for the growth of films at elevated pressures, we measured the temperature and velocity distributions of the gas, the deposition rate, and the thickness profile of the films as a function of the pressure and power in an ordinary planar magnetron sputtering system. To prevent the sputtering rate from being affected by chemical reactions that occur on the target surface during reactive sputtering, a metal target and a pure inert gas were used.

EXPERIMENTAL TECHNIQUE

The sputtering system and the placement of the probes are shown in Fig. 1. Target *1*, 75 mm in diameter, is fabricated from a Zr–30 at.% Y alloy, and the working gas is argon. The mean diameter of the erosion zone is 55 mm, and its width is 15 mm. The magnetron operated in the power-stabilization regime. In the power range $W=20-500$ W and the pressure range $P=10-110$ Pa, the working voltage is $U=130-200$ V, whereas, when $P=2.5$ Pa, $U=160-280$ V. Silicon wafers served as substrates. The sputtering rate was calculated from the thickness of the film, measured with a profilometer after chemically etching a bare strip. The design of the sputtering apparatus (Leybold Z-400) made it possible to displace the magnetron horizontally and vertically without breaking vacuum. The pressure in the chamber was measured with ionization and thermal vacuum gauges calibrated at high pressures using an oil-filled *U*-tube manometer.

The gas temperature was measured with chromel–alumel thermocouple *3* welded from wires 40 μm in radius. To ensure thermal equilibrium with the gas and to reduce the fraction of thermal power escaping through the leads into the holder, the thermocouple junction was placed in a 5×5-mm envelope made from metal foil. The thermocouple leads entered ceramic tubes 6 mm from the foil. The plane of the foil is parallel to the plane of the target. An analysis of the thermal balance of this design showed that, in the pressure and temperature interval used here, the heat loss through the leads understates the gas temperature by no more than

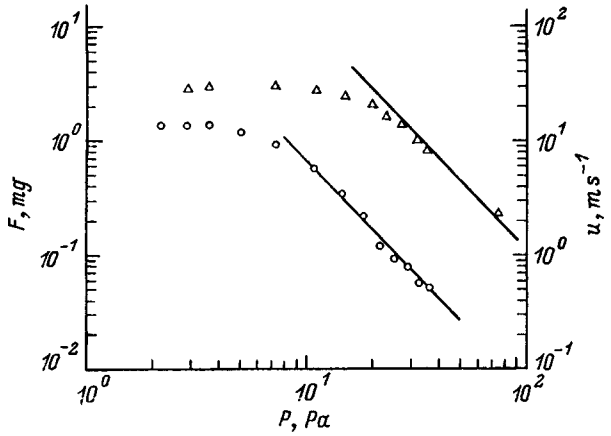


FIG. 2. Force measured by the probe under the erosion zone versus Ar pressure. $W=200$ W, △ — $d_{tp}=10$ mm and $d_{ts}=38$ mm; ○ — $d_{tp}=38$ mm and $d_{ts}=54$ mm; the solid curves show the $F\sim P^{-2}$ dependence.

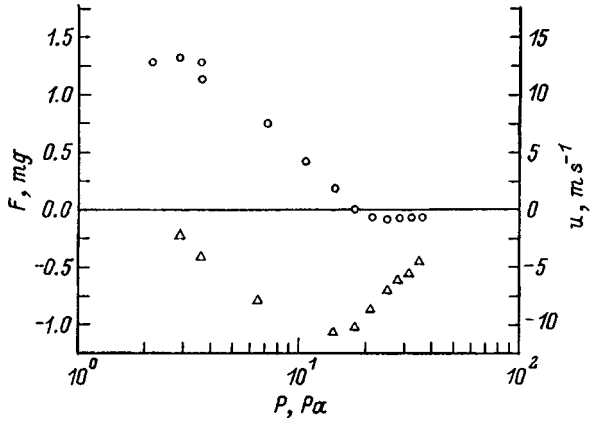


FIG. 3. The same as in Fig. 2, but with the probe on the axis of the magnetron. A positive sign corresponds to a force directed away from the magnetron.

5 °C. To reduce heating of the thermocouple by currents from the plasma, the measurement circuit was ungrounded; i.e., the thermocouple and the foil had a floating potential, which could be measured with a voltmeter connected between one of the thermocouple leads and the chamber housing.

Force probe *4*, based on the measurement system of a microammeter, was used to measure the wind velocity. A circular disk made from aluminum foil with a radius of $R=0.3$ cm was fastened to the needle. Preliminary calibration of the probe included measuring the dependence between the weights placed on the foil and the current through the microammeter coil needed to return the needle to its starting position. The probe was placed in the sputtering apparatus at various distances from the magnetron and substrate holder *2*. For small Reynolds numbers $\mathrm{R}=\rho u R/\eta$, the force F acting on a thin disk when there was orthogonal gas flow around it was calculated from the Stokes formula $F=16\eta \mathrm{R} u$, where u is the flow velocity far from the disk, η is the dynamic viscosity, and ρ is the density of the gas.[15] An experimental check of the applicability of the corresponding formulas for a sphere and a long cylinder[16] shows that they are valid to within $\approx 10\%$ up to $\mathrm{R}\approx 5$. For the area used here, a wind velocity of $u=10$ m/s in argon corresponds to a force of 1 mg. At this velocity and a pressure of $P=20$ Pa, $\mathrm{R}=0.6$.

EXPERIMENTAL RESULTS AND DISCUSSION

1. *Gas flows*. The results of the force measurement as a function of pressure at constant power are shown in Figs. 2 and 3 for four different probe positions (measurements at constant pressure show that, in all of space except for the region adjacent to the center of the substrate holder, the force is proportional to the discharge power). The right-hand axes in these figures indicate the velocities of the gaseous medium formally computed from the Stokes formula, corresponding to the measured forces. The dependence of force on pressure, measured under the erosion zone (Fig. 2), displays saturation in the region of low pressures and a falloff close to $F\sim P^{-2}$ at high pressures. The beginning of the $F\sim P^{-2}$ dependence on both curves of Fig. 2 corresponds to the product $P\cdot d\approx 300$ Pa·mm. According to Ref. 1, this value is close to the length for which the energy of the argon atoms reflected from the target decreases by a factor of 10. The data in Fig. 3 show that, on the axis of the magnetron close to the target surface ($d_{tp}=10$ mm), the force measured by the probe is directed toward the target at all pressures; closer to the substrate holder (18 mm above it), the direction of the force depends on the pressure and is also directed toward the target at high pressures. It is natural to assume that, at low pressures, when atoms sputtered and reflected from the target, possessing appreciable energy and small scattering cross section (several times less than thermal),[17,18] experience on their path toward the probe or the substrate holder a small number of scattering events, the probe records a change in their momentum when they interact with the measurement area. Close to the center of the magnetron, where the flux sputtered and reflected from the target is negligible, energetic argon atoms, most likely reflected from articles in the chamber, act on the probe, reversing the direction of the force. Such a picture agrees with the results of Ref. 13 for the gas dynamics at pressures of 0.2–4 Pa in the neighborhood of a cylindrical magnetron.

At high pressures, when thermalization occurs at lengths appreciably less than the target–sensor distance, its readings are determined by the directed motion of the adjacent medium, i.e., by the wind (vertical component). At the pressures where $F\sim P^{-2}$, the measurements were made for different locations of the probe relative to the target and substrate. The directions of the gas flows in the target–substrate space are schematically shown in Fig. 1. Briefly, the gas dynamics can be interpreted as follows: The momentum imparted to the gas by atoms knocked out of the erosion zone of the target and reflected by Ar atoms (former ions) causes it to flow out of the erosion zone. The flow out of this region is compensated by flows from the periphery and center. As one goes away from the magnetron, the gas under the entire surface of the target moves in the same direction with a velocity that is maximal for the erosion zone. The flow distribution around the substrate holder is characteristic of gas flow around a flat obstacle, with the appearance of vorticity close to the center.

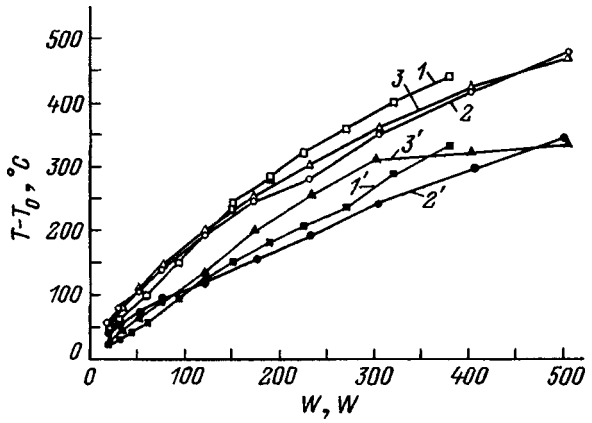


FIG. 4. Gas temperature versus discharge power. $d_{tp}=5$ mm; *1–3* — under the erosion zone, *1′–3′* — under the center of the magnetron; *1, 1′* — 36; *2, 2′* — 17; *3, 3′* — 7.2 Pa.

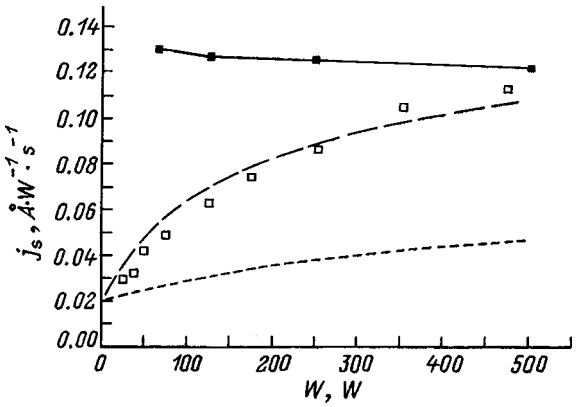


FIG. 6. Sputtering rate versus power. □ — $P=33$ Pa and $d_{ts}=3.8$ cm; ■ — $P=1$ Pa and $d_{ts}=4$ cm; the dotted curve shows the calculated values neglecting wind, and the dashed curve shows the values with wind.

Since the flow velocities of the gas are substantially less than the velocity of sound, the influence of the velocity on the density is negligible for steady-state motion.[15] The steady-state velocity is determined by setting the sum of the external forces applied to the gas equal to zero. This includes the momentum change of the energetic particles per unit time and the force of viscous friction of the gas from the internal articles of the chamber,

$$\oint \int \eta \partial u/\partial n dS + dP_{\text{ext}}/dt = 0.$$

The integration is carried out over all the internal surfaces, and the derivative of the velocity is taken along the normal to the surface.

2. *Heating of the gas*. Measurements of the radial temperature distribution of the gas under the magnetron showed that it is inhomogeneous, with its maximum under the erosion zone; when the target–probe distance is increased, the inhomogeneity decreases. Part of the experimental data are shown in Figs. 4 and 5. The data of Fig. 5 show that the coordinate dependence of the gas temperature under the erosion zone is qualitatively different from that under the center of the target: under the center, the temperature decreases as one approaches the surface of the target. As one goes away from the target, the temperatures under the erosion zone and under the center of the target become equal. As the power is increased, the measured temperature increases more weakly than linearly (Fig. 4).

3. *Sputtering rate as a function of power and pressure*. The sputtering rates reduced to unit power as a function of the power (at a constant Ar pressure of $P=33$ Pa) and of pressure (at a constant power of $W=125$ W) for $d_{ts}=38$ mm are shown in Figs. 6 and 7. The sputtering rates are shown in Å/W·sec and correspond to the maxima of the thickness profiles of the films. As can be seen from Fig. 6, the reduced sputtering rate increased by more than a factor of 3 in the range of powers used here. For comparison, the same figure shows the sputtering rates reduced to 1 W at a pressure conventional for magnetrons (at the somewhat greater value of $d_{ts}=40$ mm); these are weakly dependent ($\approx 7\%$) on power, in agreement with existing concepts.

To quantitatively estimate how wind and the heating of the gas affect the deposition rate, let us consider the diffusion of thermalized target atoms in the simplest diode-sputtering geometry. Let the target and the substrate be infinite parallel planes the distance between which is d_{ts}. Coordinate x is measured from the target in the direction of the substrate.

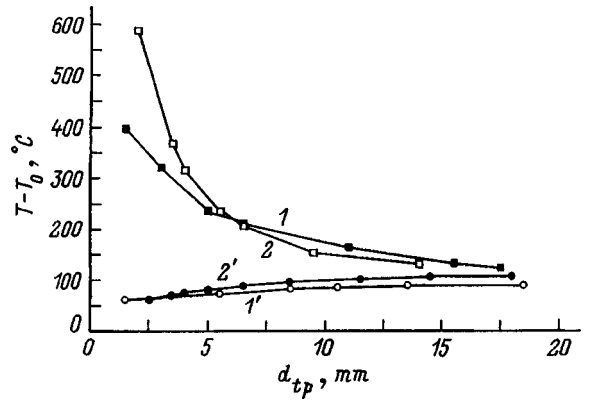


FIG. 5. Measured temperatures versus probe–target distance. $W=100$ W; squares — under the erosion zone, circles — under the center of the magnetron; *1, 2′* — 30; *2, 1′* — 2.6 Pa.

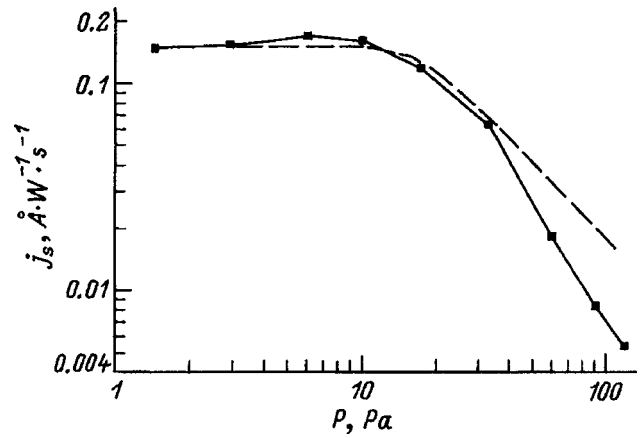


FIG. 7. Sputtering rate versus pressure. $d_{ts}=38$ mm. $W=125$ W. Dashed curve shows calculated values.

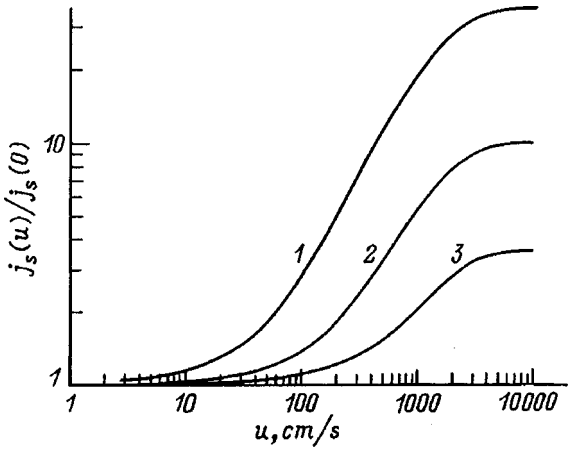


FIG. 8. Sputtering rate versus gas velocity. d_{ts}, cm: *1* — 36, *2* — 10, *3* — 3.6; $P=33$ Pa, $T=300$ °C, $d=1$ cm.

Atoms knocked out of the target thermalize at a distance d from the target, creating a source of power N, equal to the sputtering rate; they diffuse from this distance and are absorbed on the target and the substrate surfaces. The diffusion coefficient of the target atoms in the gas is D, and their concentration is $c(x)$; the velocity u of the gaseous medium is directed from the target to the substrate, and D and u are independent of the coordinate. The diffusion equation for the flux of atoms j in the steady state under these conditions has the form $j=-Ddc/dx+cu$. The boundary conditions are $c(0)=c(d_{ts})=0$. Since in the steady state the fluxes j_t and j_s of atoms to the target and the substrate are equal in modulus to the deposition rates and are independent of the coordinate, using the condition $j_s-j_t=N$, we get

$$j_s=N\frac{1-\exp(-ud/D)}{1-\exp(-ud_{ts}/D)}, \tag{1}$$

$$j_t=N\frac{\exp(-ud_{ts}/D)-\exp(-ud/D)}{1-\exp(-ud_{ts}/D)}, \tag{1a}$$

$$c_1(x)=(j_s/u)(1-\exp[u(x-d_{ts})/D]), \tag{2}$$

$$c_2(x)=(j_t/u)[1-\exp(ux/D)], \tag{2a}$$

where $c_1(x)$ is the concentration of target atoms in the space between the diffusion source and the substrate, and $c_2(x)$ is that between the diffusion source and the target.

When $u\to 0$

$$j_s=N\frac{d}{d_{ts}}, \tag{3}$$

$$j_t=-N\left(1-\frac{d}{d_{ts}}\right). \tag{3a}$$

The result of a calculation of the deposition rate as a function of the wind velocity in this simple model is shown in Fig. 8 for three target–substrate distances at constant power of the diffusion source and the indicated temperature, gas pressure, and thermalization length. It can be seen that the experimentally observed velocities, reaching 15 m $\cdot\sec^{-1}$, must appreciably increase the deposition rate relative to the rate in a stationary medium.

By comparison with the conditions for deriving Eq. (1), the actual situation differs not only in the geometry of the sputtering system, but also in the nonuniform distribution of the velocity of the gaseous medium and the temperature, on which the diffusion coefficient and the thermalization length depend. Calculations for a three-layer medium with constant velocities and diffusion coefficients that differ between the layers showed that the deposition rate is mainly determined by the characteristics of the medium close to the plane of the diffusion source; the variation of the parameters in the other layers is substantially weaker. Therefore, to obtain calculated dependences of the sputtering rate on the power and pressure, we used the experimental values of the velocity and temperature of the gas at distances from the target equal to the thermalization lengths. As mentioned above, in the diffusion regime, when the wind velocity is negligible, the deposition rate of the film is expressed by the very simple formula given in Eq. (3). N, the power of the diffusion source, was taken to be equal to the sputtering rate at low pressure, at which deposition of the sputtered atoms onto the target is negligible, $N=j_s(0.4\text{ Pa})=0.15$ Å $\cdot\text{W}^{-1}\cdot\text{s}^{-1}$ (Fig. 8). Strictly speaking, it would be more correct to write $N=j_s(0.4\text{ Pa})[1-G(\mu)]$. Taking into account that the dropout ratio $G(\mu)$ of sputtered atoms that return to the target without thermalizing is independent of pressure,[9] to simplify the analysis, we neglected it. This can result in a small numerical error ($\approx 25\%$) in calculating the dependence of the sputtering rate on the pressure but does not alter its functional form. A second reason for neglecting $G(\mu)$ is given below. Since the thermalization length depends on the particle density of the gas–scatterer, i.e., on its temperature and pressure, it is expedient to determine it by extrapolating the sputtering rate (Fig. 6) to zero power, at which there is no heating of the gas $j_s(33\text{ Pa})\approx 0.02$ Å $\cdot\text{W}^{-1}\cdot\text{s}^{-1}$. Taking into account $d_{ts}=3.8$ cm and using Eq. (3), we get that the effective value for thermalizing the sputtering of the atoms is $Pd\approx 16(T_a+\Delta T)/T_a$ Pa·cm, where T_a is the temperature of the gas in the chamber, and ΔT is the heating of the gas when the discharge is switched on. The successive calculation of the thermalization length, using the interaction potential obtained from experimental data on scattering, was carried out only for a relatively high average initial energy of the ejected atoms, $\langle E\rangle=20$ eV.[1] For Zr in Ar, the result is $Pd\approx 50$ Pa·cm. In our case, the cathode potential is $U=130-200$ V, and $\langle E\rangle=5-6.5$ eV;[19,20] the value of Pd must be lower. In an earlier paper,[17] using the thermal scattering cross section, the thermalization length was calculated as a function of the initial energy. With $\langle E\rangle=5$ eV, the result for Zr in Ar was $Pd\approx 5$ Pa·cm. Considering these results, the value of $Pd\approx 16$ Pa·cm that we obtained seems plausible. It seems that the calculated dropout ratio $G(\mu)$ also depends on the chosen interaction potential that determines the scattering cross sections, and the approximation $G(\mu)\approx 0.09(3-\ln\sqrt{\mu})$ obtained in Ref. 9 for the Lennard–Jones potential can fail to correspond to the experimental conditions.

For calculations using Eq. (1), the diffusion coefficient of Zr in Ar is assumed to be in the form $D=(1.56\times 10^4 P)\times[(T_a+\Delta T)/273]^{1.5}$, following from the Langevin formula

$D_{12}=(3\lambda/8)(\pi kT/2M_r)^{0.5}$ [$\lambda=(\sqrt{2}n\pi d_{12}^2)^{-1}$ is the path length, n is the particle density in the gas, $d_{12}=(d_1+d_2)/2$ is the mean diameter of the particles, and $M_r=m_1m_2/(m_1+m_2)$ is the reduced mass]. In accordance with the results of the measurements, the wind velocity was assumed to be proportional to the discharge power and inversely proportional to the square of the pressure. Because data on the dependence of velocity on the coordinate was limited, to take into account its reduction as one goes away from the target, the velocity was assumed to be inversely proportional to the coordinate. The result of the calculation from Eq. (1) is shown in Figs. 6 and 7 by the dashed curves. The dotted curve in Fig. 6 shows the values computed for zero wind velocity. It can be seen that, despite the roughness of the model, taking the heating and motion of the gaseous medium into account results in satisfactory agreement with the experimental dependence of the sputtering rate on the power. It is clearly inadequate to take into account only the change of the thermalization length as a result of heating while neglecting the wind (Fig. 6, dotted curve).

It can be seen by comparing the experimental and calculated pressure dependences of the sputtering rates (Fig. 7) that they satisfactorily agree below $P=33$ Pa, whereas the measured sputtering rate is less than the calculated results above that pressure. The calculated sputtering rate is almost inversely proportional to the pressure at high pressures. Such behavior is caused by the fact that, as the pressure increases, the wind velocity decreases more rapidly than does the diffusion coefficient ($u\sim P^{-2}$, Fig. 2, $D\sim P^{-1}$), the gas temperature at constant power is independent of the pressure (Fig. 4), and the diffusion of the sputtered atoms approximates diffusion in a stationary medium.

The experimental data on the pressure dependence of the sputtering rate for $33<P<120$ Pa show a decrease of the rate close to $j_s\sim P^{-2}$ (Fig. 7). To explain similar behavior observed when sputtering $YBa_2Cu_3O_7$ films, Sakuta *et al.*[21] assumed that the dropout ratio of atoms that return to the target without thermalizing is inversely proportional to the pressure. This assumption contradicts the theoretical result obtained by Ref. 9 that G is independent of pressure. A second hypothesis that makes it possible in principle to explain the decrease in sputtering rate with increasing pressure is that the energy of the Ar^+ ions bombarding the target decreases because of collisions in the cathode potential fall region where they are accelerated. Taking into account the threshold character of the energy dependence of the sputtering coefficient,[22] this can cause a reduction in the sputtering rate. However, our measurements of the potential distribution close to the cathode with an electric probe showed that the thickness of the cathode layer in the entire region of pressures used here is less than the mean path length of the ions before charge exchange, and consequently the flux of atoms knocked out of the target at constant discharge power is independent of pressure.

The overestimated calculated values of j_s in the pressure region discussed here most likely have the following cause: As the pressure increases, the thermalization length d decreases, and the plane of the diffusion source is located beside the region of intense plasma close to the erosion zone of the magnetron, where the gas heats up much more strongly than elsewhere. Since $D\sim T^{1.5}$, the approximation that the diffusion coefficient is constant can become too crude. The effective diffusion coefficient for the reverse flow of atoms to the target must appreciably exceed that for the flow of atoms to the substrate. Solving the diffusion equation $j=-Ddc/dx$ for a stationary two-layer medium with different temperatures and a diffusion source at the interface shows that the ratio of the sputtering rate j_s^* for this case to the sputtering rate j_s in a homogeneous medium is

$$\frac{j_s^*}{j_s}=\frac{d_{ts}}{d+(T_1/T_2)^{1.5}(d_{ts}-d)}\rightarrow(T_2/T_1)^{1.5}$$

$$\text{for}\quad d_{ts}\gg d, \tag{4}$$

where T_1 is the temperature of the layer close to the target, and T_2 is the temperature elsewhere.

At distances from the target less than 5 mm, the gas temperature increases appreciably (Fig. 5). Using the estimate $Pd\approx 16$ Pa·cm, we get that the approximation of a homogeneous medium becomes inapplicable when $P>30$ Pa. The temperatures close to the target (Fig. 5) are such that the experimentally observed decrease of the sputtering rate is approximately provided by comparison with that calculated in the model of a homogeneous medium.

Temperature inhomogeneity also reduces the concentration of thermalized target atoms by comparison with the calculated value for a homogeneous medium. For a stationary medium, the maximum concentration is given by

$$c_0=\frac{Nd(d_{ts}-d)}{D_2d+D_1(d_{ts}-d)}. \tag{5}$$

When $d\ll d_{ts}$, for a homogeneous medium ($D_1=D_2=D$), $c_0=Nd/D$. For a two layer medium, $c_0=Ndd_{ts}/(D_2d+D_1d_{ts})$.

4. *Surface structure and thickness profile of the films.* The structure of films sputtered at $P=33$ Pa depends on the discharge power. When $W=25$ W, the entire surface of the films is homogeneous and smooth. When $W\geqslant 38$ W, a smeared dull spot is observed in the central part of the films. The surface of the film has a weakly expressed structure outside the central region, but the structure becomes stronger as one approaches the spot. Its form in the electron microscope is shown in Fig. 9. A similar evolution of the surface structure is observed in films sputtered at constant power ($W=125$ W) as the argon pressure increases. For pressures up to 10 Pa, the films are smooth and structureless; at 17 Pa$\leqslant P\leqslant$90 Pa, a sootlike deposit is present at the center and a weakly expressed surface structure elsewhere. When $P=120$ Pa, the film surface again is smooth, with no spot (below we shall cite the pressure and power corresponding to transitions of the film surface from smooth to dull and back as in the regimes a, b and c in the order they are mentioned). Additional sputtering with the center of the magnetron displaced relative to the center of the substrate holder revealed that the spot always was located close to the center of the holder and not under the center of the magnetron. In our opinion, this is evidence that the inhomogeneous film struc-

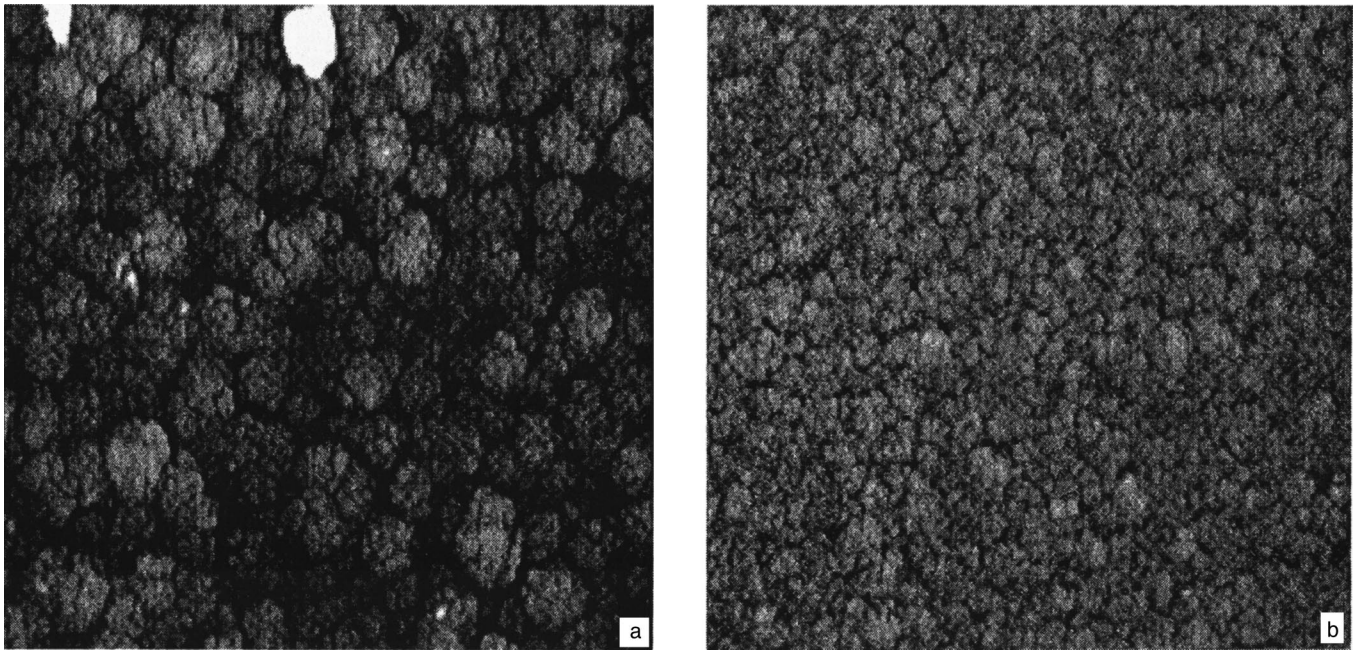


FIG. 9. Form of the surface of rough films. (a) at the dull spot, (b) in the peripheral region. The size of the field of the picture is 8×8 μm.

ture is caused by a singularity in the dynamics of the gas flow close to the center of the substrate holder. It is well known[15] that, when a gas flows around a flat body with a moderate velocity (Reynolds number $R\ll 1$), the flow moves along the surface in a radial direction everywhere close to the surface except for the central region; at the central (critical) point, the velocity is zero. When $R\sim 10$, the gas flowing around such a body becomes turbulent. As shown by measurements of the gas velocity under our conditions, it is minimal close to the center of the substrate holder, and there is vorticity in the gas flow (Fig. 1). We assume the following scenario for depositing the particles: the atoms emitted by the target thermalize, and, when their concentration is sufficient, the vapor condenses. The condensate is entrained by the gas flow in the direction of the substrate and can precipitate on the surfaces that the flow is passing around. Close to the center of the substrate, where there is a singularity in the distribution of the gas velocity, the particles spend more time close to the substrate than they do over the periphery, the precipitation probability of the particles is greater at this site, and this causes them to have a greater concentration.

The most intense condensation of the vapor must occur in the region of the highest concentration of metal atoms, i.e., close to the diffusion source. An estimate of the concentration from Eq. (2) for $x=d$ gives $c_{\mathrm{Zr}}=1.6$, 2.8, and 2.1 $\times 10^{13}$ cm^{-3} for the a,b and c regimes, respectively. We assume that the limiting stage of particle growth in the vapor is the formation of diatomic molecules of the sputtered metal. According to chemical kinetics,[23] it occurs in the triple collisions $Me+Me+\mathrm{Ar}=Me_2+\mathrm{Ar}$, whose probability is proportional to $c_{\mathrm{Zr}}^2\cdot n_{\mathrm{Ar}}$. For the threshold regimes, in the same order of enumeration as before, we have $c_{\mathrm{Zr}}^2\cdot n_{\mathrm{Ar}}=2.4$, 3.9, and 6×10^{42} cm^{-6}. The concentration estimates must be overestimated for regime c, because the approximation of a homogeneous medium in which Eq. (2) is obtained is explicitly not satisfied.

Thickness profiles of films sputtered at high and low pressures differ qualitatively: at low pressures, the profile is appreciably more convex (Fig. 10). For films sputtered at $P=33$ Pa, the experimental results for the central part are not shown because of the large scatter due to surface roughness in this region.

A calculation of the thickness profile for the collisionless regime with an emission probability proportional to the cosine of the ejection angle and the depth of the erosion zone (Fig. 10, solid curve), gives a convex distribution close to that observed at $P\leqslant 2$ Pa. The thickness profile for higher pressures (Fig. 10, dotted curve) was calculated in terms of the diffusion-source model by a technique close to that explained in Ref. 10. The diffusion source was modeled by a flat ring located under the erosion zone at a distance equal to

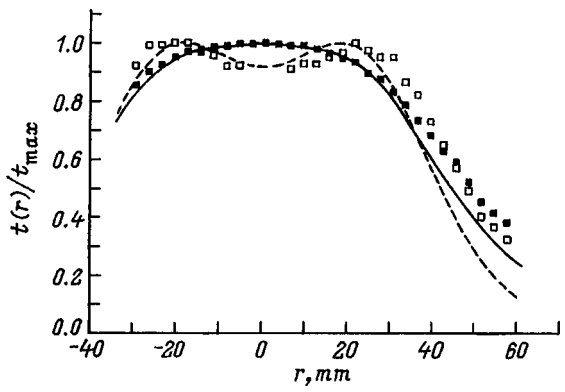


FIG. 10. Thickness profiles of films. $d_{ts}=36$ mm. The squares are experimental data, the curves show the result of calculation; P, Pa: ■ — 2, . . . ,□ — 33.

the thermalization length. As the distance from the diffusion source to the target was varied within the limits 5–20 mm and the width of the ring was varied from zero to 20–40 mm, respectively, the thickness profiles obtained by calculation for $d_{ts}=36$ mm possess a dip at the center. A three-dimensional diffusion source can be represented in the form of a sum of the flat rings considered here; the overall thickness profile of the film will also have a dip at the center. Consequently, diffusional transport of atoms results in a less convex thickness profile than does collisionless transport. The same result is obtained for pressures of 10 Pa when the calculation is done by the Monte Carlo method.[6]

We should point out that, in the range of pressures and gas velocities considered here, the velocity inhomogeneity of the gas flow determines the distribution of finely dispersed particles on the substrate surface, but does not cause a significant difference of the thickness profile of the film from that calculated for a stationary medium. This can be because the profile shape is weakly sensitive to the distance of the diffusion source from the target.

CONCLUSION

It has been shown in this paper that, for magnetron sputtering at elevated pressures (tens of pascal), the retardation of the energetic particles emitted by the target heats the gaseous medium and puts it into motion. The structure of the gas flow has been explained. The increase of the sputtered atoms by the moving medium appreciably increases the sputtering rate. As the power increases, the sputtering rate increases faster than linearly. The increase of the concentration of sputtered atoms in the medium causes them to condense in the vapor phase and causes finely dispersed particles to settle out on the growing surface. The threshold concentration below which no particles are detected on the film surface has been estimated.

[1] R. E. Somekh, J. Vac. Sci. Technol. A **2**, 1285 (1984).
[2] J. M. Harper, J. J. Cuomo, R. J. Gambino *et al.*, J. Vac. Sci. Technol. **15**, 1597 (1978).
[3] C. T. Eom, J. M. Phillips, and R. J. Cava, Mater. Res. Soc. Symp. Proc. **341**, 229 (1994).
[4] Th. Schurig, S. Menkel, Z. Quan *et al.*, Physica C **262**, 89 (1996).
[5] A. M. Myers, J. R. Doyle, J. R. Abelson *et al.*, J. Vac. Sci. Technol. A **9**, 614 (1991).
[6] F. A. Chudnovskii, B. Sh. Elkin, A. A. Fursenko *et al.*, in *Proceedings of the Third German–Soviet Bilateral Seminar on High-Temperature Superconductivity*, Karlsruhe, 1990, pp. 451–457.
[7] J. A. Valles-Abarca and A. Gras-Marti, J. Appl. Phys. **55**, 1370 (1984).
[8] A. G. Zhiglinsky, A. M. Izmailov, V. V. Kuchinsky *et al.*, Izv. Vyssh. Uchebn. Zaved. Fiz. **25**, 869 (1982).
[9] A. G. Zhiglinsky and V. V. Kuchinsky, *Mass Transfer in the Interaction of a Plasma with a Surface* [in Russian] (Énergoizdat, Moscow, 1991), 206 pp.
[10] I. Petrov, I. Ivanov, V. Orlinov *et al.*, J. Vac. Sci. Technol. A **11**, 2733 (1993).
[11] S. M. Rossnagel, J. Vac. Sci. Technol. A **6**, 19 (1988).
[12] S. M. Rossnagel, IEEE Trans. Plasma Sci. **PS-18**, 878 (1990).
[13] D. W. Hoffman, J. Vac. Sci. Technol. A **3**, 561 (1985).
[14] I. D. Morozov, L. I. Trusov, and V. N. Lapovok, *Physical Phenomena in Ultradisperse Media* (Énergoizdat, Moscow, 1984), 372 pp.
[15] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* [Pergamon Press, Oxford, 1987; Nauka, Moscow, 1988, 736 pp.].
[16] H. von Schlichting, *Grenzschicht-Theorie* [Verlag G. Braun, Karlsruhe; Nauka, Moscow, 1974, 480 pp.].
[17] W. D. Westwood, J. Vac. Sci. Technol. **15**, 1 (1978).
[18] R. S. Robinson, J. Vac. Sci. Technol. **16**, 185 (1979).
[19] R. A. Brizzolara, C. B. Cooper, and T. K. Olson, Nucl. Instrum. Methods **35**, 36 (1988).
[20] Y. Okabe, M. Sasao, M. Wada *et al.*, Jpn. J. Appl. Phys. **30**, L1428 (1991).
[21] K. Sakuta, Y. Sakaguchi, T. Awaji *et al.*, IEICE Trans. **E74**, 1960 (1991).
[22] R. Behrisch, Ed., *Sputtering by Particle Bombardment I.*, vol. 47 of Applied Physics [Springer, Berlin, 1981; Mir, Moscow, 1984].
[23] V. N. Kondrat'ev and B. E. Nikitin, *The Kinetics and Mechanism of Vapor-Phase Reactions* (Nauka, Moscow, 1974), 558 pp.

Translated by W. J. Manthey

Stability of charged drops of spheroidal shapes with respect to axisymmetric deformation

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

Yaroslavl State University, 150000 Yaroslavl, Russia
(Submitted June 16, 1997)
Zh. Tekh. Fiz. **68**, 33–36 (July 1998)

The stability of a strongly charged spherical drop with respect to deformations of its shape to prolate and oblate spheroids has been studied. It is shown that drops can become unstable and can break apart provided that the virtual shape is a prolate spheroid. Deforming a drop to an oblate spheroid does not cause it to break apart. © *1998 American Institute of Physics.* [S1063-7842(98)00607-2]

INTRODUCTION

The study of the stability of strongly charged drops with respect to small deformations is of interest in connection with numerous applications in physics, geophysics, scientific instrumentation, and technology.[1,2] According to Refs. 3 and 4, it is easiest to excite the instability of the normal axisymmetric mode $\sim \pm P_2(\cos\theta)$ of a drop corresponding to the deformation of the original spherical drop to prolate or oblate spheroids [$P_2(\cos\theta)$ is a Legendre polynomial]. This is confirmed by the results of full-scale measurements in clouds. Thus, it is reported in Ref. 5 that, when 1783 drops were photographed in mutually perpendicular directions with two cameras, a spherical shape was noted in 569 cases, a prolate spheroidal shape in 496 cases, and an oblate spheroidal shape in 331 cases. According to numerical calculations,[6] the shape of a drop suspended in the atmosphere depends on its size and can be either spherical or spheroidal. In this connection, it seems expedient to study the question of how the variation of the potential energy of a charged drop undergoing virtual deformations to the shapes of prolate and oblate spheroids depends on the deformation amplitude (the eccentricity of the spheroid) and the charge on the drop. It should be pointed out that a similar problem was solved earlier.[7] However, the analysis in Ref. 7 was carried out only for a weakly spheroidal drop, based on the method of expansion in a small parameter, for which the value of the characteristic deformation was chosen. The results obtained in Ref. 7 have a particular character, while the main sum of Ref. 7 should be regarded as the formulation of the problem.

1. Let a spherical drop of radius R with surface tension σ and charge Q initially undergo a virtual deformation to a prolate spheroid with eccentricity e. The potential energy U of such a spheroidal drop will consist of the energy of the surface-tension forces and the energy of the electric field of its own charge[7,8]

$$e^2=1-\frac{b^2}{a^2},\quad b=R(1-e^2)^{1/6},\quad a=R(1-e^2)^{-1/3},$$

$$U=\frac{e\sqrt{1-e^2}+\arcsin(e)+\operatorname{arccos}\,h\left(\dfrac{1}{\sqrt{1-e^2}}\right)\sqrt{1-e^2}W}{(1-e^2)^{1/6}e(W+2)},\tag{1}$$

where a and b are the semimajor and semiminor axes of the spheroid, the Rayleigh parameter $W=Q^2/(4\pi R^3\sigma)$ characterizes the stability of the drop with respect to its own charge, the drops become unstable when $W\geqslant 4$,[3] and energy U is normalized to the potential energy U_0 of the initial spherical drop

$$U_0=4\pi R^2\sigma+\frac{Q^2}{2R}.\tag{2}$$

Figure 1 shows the $U=U(W,e)$ dependence calculated from Eq. (1). It can be seen from this figure that, for subcritical values of the Rayleigh parameter W (for $W>4$), the potential energy of the drop decreases as the eccentricity increases, passes through a minimum at $e\approx 0.9$, and then quickly increases as e increases further. In Fig. 2, curve *1* shows the dependence of parameter W on eccentricity e corresponding to constant energy $U=1$ of the drop.

Curve *2* in the same figure shows a similar dependence, obtained from the condition $\partial U/\partial e=0$

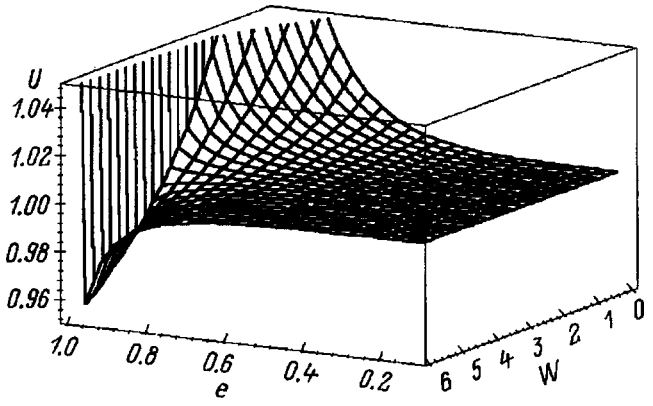


FIG. 1. Energy U of charged drop having the shape of a prolate spheroid versus Rayleigh parameter W and eccentricity e.

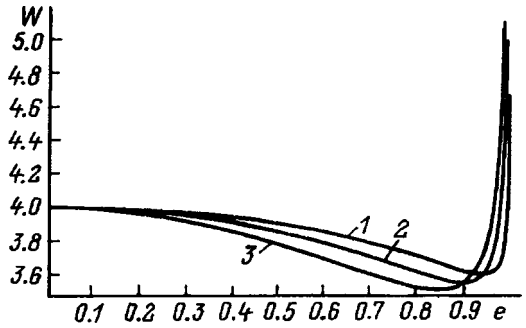


FIG. 2. Rayleigh parameter W versus the eccentricity e of a prolate spheroidal drop, obtained from the conditions $U=1$, $\partial U/\partial e=0$, and $\partial^2 U/\partial e^2=0$ (*1–3*, respectively).

$$\frac{\partial U}{\partial e}=-\frac{1}{3}\left[2e^3-4\frac{e^2\arcsin(e)}{\sqrt{1-e^2}}-3We-\operatorname{arccos}h\left(\frac{1}{\sqrt{1-e^2}}\right)We^2-3e+3\frac{\arcsin(e)}{\sqrt{1-e^2}}+3\operatorname{arccos}h\left(\frac{1}{\sqrt{1-e^2}}\right)W\right]\times[e^2(1-e^2)^{2/3}(W+2)]^{-1}. \quad (3)$$

Curve *2* connects the values of W and e corresponding to the extreme variation of the potential energy of a deformed drop in an actual deformation process (in accordance with the principle of least action).

Curve *3* in Fig. 2 is obtained from the condition $\partial^2 U/\partial e^2=0$,

$$\frac{\partial^2 U}{\partial e^2}=-\frac{1}{9}\left[2e^5-28\frac{e^4\arcsin(e)}{\sqrt{1-e^2}}-4W\times\operatorname{arccos}h\left(\frac{1}{\sqrt{1-e^2}}\right)e^4-24We^3-27e^3+18We+39\frac{e^2\arcsin(e)}{\sqrt{1-e^2}}+30\operatorname{arccos}h\left(\frac{1}{\sqrt{1-e^2}}\right)We^2+18e-18\frac{\arcsin(e)}{\sqrt{1-e^2}}-18\operatorname{arccos}h\left(\frac{1}{\sqrt{1-e^2}}\right)W\right]\times[(W+2)e^3(1-e^2)^{5/3}]^{-1}, \quad (4)$$

and separates the (W,e) plane into two regions: W and e values corresponding to the geometrical site of points located above curve *3* characterize an unstable state of a prolate spheroidal drop; all the (W,e) states below it correspond to stable spheroidal shapes from which the drop returns to the original spherical shape. Thus, a charged drop with virtual deformation to a prolate spheroid can decay only to the region of (W,e) values in Fig. 2 that satisfy the condition

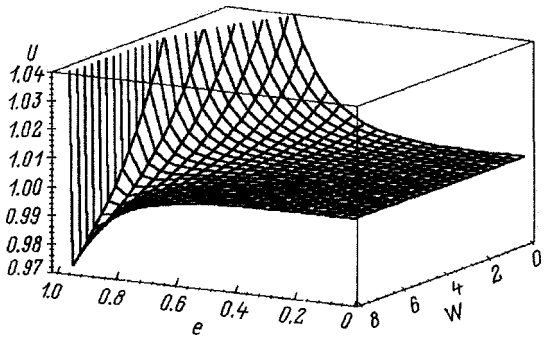


FIG. 3. Energy U of a charged drop having the shape of an oblate spheroid versus Rayleigh parameter W and eccentricity e.

$$\frac{\partial^2 U}{\partial e^2}\geqslant 0, \quad \frac{\partial U}{\partial e}\geqslant 0. \quad (5)$$

It can be seen from Fig. 2 that this condition is satisfied only for extremely significant deformations: $e\geqslant 0.9$. What the decay will be— into two parts of comparable size[9] or with emission of a large number of highly dispersed, strongly charged droplets[10]—is impossible to say in the idealization used here. This will depend on the viscosity and conductivity of the drop and the external medium.[11,12]

2. Repeating the discussion above for a charged drop having the shape of an oblate spheroid, it is easy to find the expression for its potential energy:[7,8]

$$U=\frac{1}{2}\left[2e+\ln\left(\frac{1+e}{1-e}\right)-\ln\left(\frac{1+e}{1-e}\right)e^2+2\sqrt{1-e^2}\times\arccos(\sqrt{1-e^2})W\right][e(1-e^2)^{1/3}(W+2)]^{-1},$$

$$e=\sqrt{1-\frac{c^2}{a^2}}, \quad a=\frac{R}{(1-e^2)^{1/6}}, \quad c=R(1-e^2)^{1/3}, \quad (6)$$

where a is the semimajor and c the semiminor axis of the spheroid.

Figure 3 shows the dependence of the potential energy of an oblate spheroidal drop on Rayleigh parameter W and eccentricity e, calculated from Eq. (6). It is easy to see that it is qualitatively similar to the analogous dependence for a prolate spheroid, shown in Fig. 1.

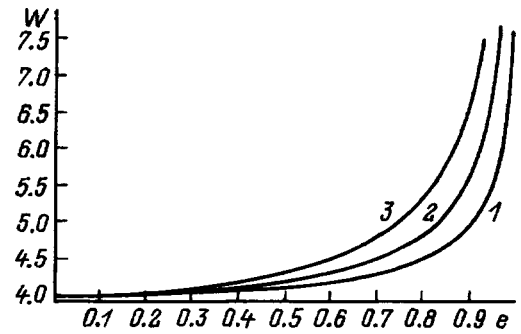


FIG. 4. Rayleigh parameter W versus the eccentricity e of an oblate spheroidal drop, obtained from the conditions $U=1$, $\partial U/\partial e=0$, and $\partial^2 U/\partial e^2=0$ (*1–3*, respectively).

In Fig. 4, curve *1* shows the dependence of W on e corresponding to constant energy $U=1$ of the drop. Curve *2* in the same figure shows the dependence of Rayleigh parameter W on eccentricity e corresponding to the extremal variation of the potential energy or to an actual deformation process of a drop because of the principle of least action, obtained from the condition $\partial U/\partial e=0$:

$$\frac{\partial U}{\partial e}=-\frac{1}{6}\Big\{(6-10e^2+4e^4)W\ \arccos(\sqrt{1-e^2})+(-e^4-2e^2+3)\sqrt{1-e^2}\ \ln\left(\frac{1+e}{1-e}\right)+6(-1+e^2)\sqrt{1-e^2}e(W+1)-4\sqrt{1-e^2}e^3\Big\}\times[e^2(W+2)(1-e^2)^{11/6}]^{-1}. \quad (7)$$

Curve *3* in Fig. 4 shows the dependence of W on e obtained for the condition $\partial^2 U/\partial e^2=0$:

$$\frac{\partial^2 U}{\partial e^2}=\frac{1}{9}\Big[(10e^4-33e^2+18)\sqrt{1-e^2}W\ \arccos(\sqrt{1-e^2})+(e^6+11e^4-21e^2+9)\ln\left(\frac{1+e}{1-e}\right)+(-21e^5-18e+39e^3)W-2e^5+36e^3-18e\Big]\times[e^3(W+2)(1-e^2)^{1/3}(-1+e^2)^2]^{-1}. \quad (8)$$

The geometrical site of points (W,e) located below curve *3* corresponds to the states of an oblate spheroidal drop from which it returns to the original spherical shape. Since all the points of curve *2* along which W and e vary for an actual deformation of a drop are located below curve *3*, a charged drop does not break up when it is symmetrically deformed to an oblate spheroid. This circumstance was pointed out earlier in Ref. 13 on the basis of other starting premises. A charged oblate spheroidal drop can break up only when it is asymmetrically deformed.[14]

3. Figure 5 shows how the potential energy of a prolate spheroidal drop (with $e^2>0$) and an oblate spheroidal drop (with $e^2<0$) depends on the square of the eccentricity e^2 for fixed values of the Rayleigh parameter $W=$ const. It is easy to see that, while the behavior of the $U=U(e)$ dependence is qualitatively similar for oblate and prolate spheroids, the minimum of function $U=U(e)$ is much deeper for a prolate spheroid, which explains why the probabilities of observing drops of oblate and prolate spheroidal shapes are different in full-scale measurements.[5]

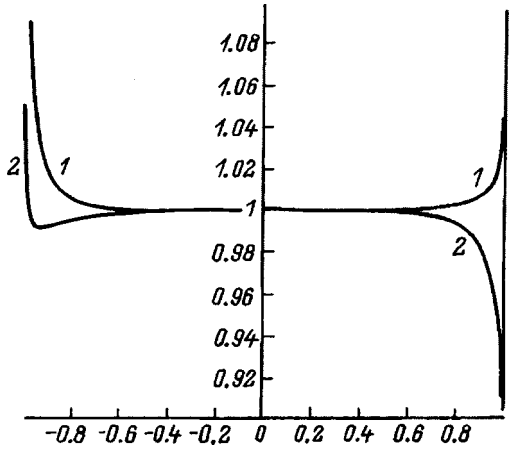


FIG. 5. Energy U of spheroidal drops of oblate ($e^2<0$) and prolate ($e^2>0$) shapes versus the square of the eccentricity, e^2, for $W=3$ (*1*) and 6 (*2*).

CONCLUSION

An analysis based on the principle of minimizing the potential energy of an isolated charged drop shows that strongly charged spherical drops deformed to the shape of a prolate spheroid can become unstable with respect to breaking up into daughter drops, whereas drops deformed to the shape of a oblate spheroid are stable.

[1] A. I. Grigor'ev, S. O. Shiryaeva, and S. I. Shevchenko, Nauchn. Prib. **1**, No. 3, 25 (1991).
[2] A. I. Grigor'ev and S. O. Shiryaeva, Izv. Ross. Akad. Nauk Mekh. Zhidk. Gazov, No. 3, 3 (1994).
[3] Lord Rayleigh (John William Strutt), Philos. Mag. **14**, 184 (1982).
[4] A. I. Grigor'ev, Zh. Tekh. Fiz. **55**, 1272 (1985) [Sov. Phys. Tech. Phys. **30**, 736 (1985)].
[5] D. M. A. Jones, J. Meteor. **16**, 504 (1959).
[6] H. R. Pruppacher and J. D. Klett, *Microphysics of Clouds and Precipitation* (D. Reidel Publish. Co., 1978), 714 pp.
[7] G. Ailam and I. Gallily, Phys. Fluids **5**, 575 (1962).
[8] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, 2nd ed. (Pergamon Press, Oxford, 1984) [Russ. original, Nauka, Moscow, 1982, 620 pp.].
[9] S. O. Shiryaeva, V. A. Koromyslov, and A. I. Grigor'ev, Elektron. Obrab. Mat. No. 1, 35 (1995).
[10] A. I. Grigor'ev and S. O. Shiryaeva, Zh. Tekh. Fiz. **61**(3), 19 (1991) [Sov. Phys. Tech. Phys. **36**, 258 (1991)].
[11] S. O. Shiryaeva, Pis'ma Zh. Tekh. Fiz. **22**(4), 84 (1996) [Tech. Phys. Lett. **22**, 171 (1996)].
[12] S. O. Shiryaeva, A. I. Grigor'ev and V. A. Koromyslov, Pis'ma Zh. Tekh. Fiz. **22**(9), 64 (1996) [Tech. Phys. Lett. **22**, 371 (1996)].
[13] O. A. Bassaran and L. E. Scriven, Phys. Fluids A **1**, 795 (1989).
[14] A. I. Grigor'ev, A. A. Firstov, and S. O. Shiryaeva, in *Proceedings of the Ninth International Conference on Atmospheric Electricity, St. Petersburg, 1992*, pp. 450–453.

Translated by W. J. Manthey

Classical limiting values of energy production in the D–^{3}He plasma of an ambipolar reactor

V. I. Khvesyuk, N. V. Shabrov, D. V. Semenov, and A. N. Lyakhov

Scientific-Research Institute of Power Machinery, N. É. Bauman Moscow State Technical University, 107005 Moscow, Russia

(Submitted October 23, 1996)
Zh. Tekh. Fiz. **68**, 37–43 (July 1998)

[S1063-7842(98)00707-7]

INTRODUCTION

In connection with the exhaustion of mineral fuel reserves, the question of searching for new energy sources that will provide for the global needs of mankind is becoming acute. The main requirements on such sources are the presence of large fuel resources in nature, ecological cleanliness, and high energy efficiency. One promising source is future thermonuclear reactors. The weak side of this approach is technological unreadiness. However, analysis shows that, along with this, the question of choosing a thermonuclear fuel is becoming acute. The fusion of deuterium and tritium,

$$D_1^2+T_1^3\rightarrow He_2^4+p_1^1+n_0^1+17.6\ \text{MeV}, \tag{1}$$

is the most attractive from the viewpoint of achieving plasma parameters that ensure a positive power output (plasma temperature, Lawson parameter), because reaction (1) has the highest rate of all known combinations of the light elements. However, the use of this fuel poses very serious (and possibly insoluble) ecological and technological problems, because 80% of the fusion energy in the D–T reaction is due to the neutrons. The interaction of fast neutrons with the nuclei of the structural materials of the reactor causes large induced radioactivity of the materials, whose level coincides with the radioactivity of a nuclear reactor of the same power. Changes in the spatial structure of the lattice of metal alloys cause them to swell and lose their strength and consequently shortens the lifetime of the structural elements to 5–6 yr.[1]

Thus, the use of D–T fuel makes it extremely problematic to ensure both the required level of ecological cleanliness and the energy efficiency of a fusion reactor. In this connection, there is great interest in analyzing the feasibility of low-radioactivity controllable fusion in the reaction

$$D_1^2+He_2^3\rightarrow He_2^4+p_1^1+17.6\ \text{MeV}. \tag{2}$$

As can be seen, no neutrons are formed and there are no radioactive nuclei in this reaction. However, because of the relatively low rate of reaction (2), it is necessary to take into account the parallel branches of the reaction,

$$D_1^2+D_1^2\rightarrow T_1^3+p_1^1+4.0\ \text{MeV}, \tag{3}$$

$$D_1^2+D_1^2\rightarrow {}_2^3\text{He}+n_0^1+3.52\ \text{MeV}, \tag{4}$$

while the appearance in reaction (3) of tritium must be considered, along with reaction (1). Thus, neutrons appear when D–^{3}He fuel is used, but, as shown by calculations, they are responsible for less than 5% of the total power of the nuclear fusion.[2]

However, the rate of reaction (2) (and, consequently, the power evolved for an identical concentration) is a factor of several tens less than that of reaction (1). Therefore, the energy fluxes in a multicomponent plasma of D, ^{3}He, T, ^{4}He, p, and e must be carefully analyzed to obtain a reliable estimate of the energy efficiency.

The simple integral estimates of Ref. 2 show that positive energy output can be achieved in a D–^{3}He plasma confined by a magnetic field when the condition

$$T_{\text{fuel}}=T_{\text{D}}=T_{\text{He}}=60-90\ \text{keV}. \tag{5}$$

is satisfied.

The value of β_Σ in this case should be rather large ($\beta_\Sigma\sim 0.4$–0.9). We use the power multiplication factor Q_{pl} as a parameter to characterize the energy production in the plasma:

$$Q_{\text{pl}}=\frac{P_{\text{fus}}}{P_{\text{req}}}, \tag{6}$$

where $P_{\text{fus}}\text{W/m}^3$ is the power density evolved as a result of the fusion reaction, and P_{req} is the power density that must be introduced into the plasma to maintain the given fuel temperature T_{fuel}.

The energy fluxes in the technical apparatus are neglected here. To ensure that a thermonuclear reactor is competitive, the following values of the parameters that characterize the energy production in a thermonuclear plasma must be achieved:[3]

$$Q_{\text{pl}}>10, \tag{7}$$

$$P_{\text{fus}}>2\ \text{MW/m}^3. \tag{8}$$

The main purpose of this paper is to estimate the theoretical limiting values of Q_{pl} for the central cell of the Tandem Mirror Reactor. Multicomponent classical kinetic models neglecting turbulence effects were used to solve the problem.

An ambipolar confinement system was chosen because it provides high β_Σ values and simple geometry, and a direct conversion system can be used in it; nonclassical transfers of particles and energy across the magnetic field are absent in axisymmetric systems.

Kinetic calculations carried out for $\beta_\Sigma = 0.7$ showed that inequalities (7) and (8) are not satisfied for a D–^{3}He plasma when a conventional ambipolar reactor system is used. This is because a large fraction of β_Σ comes from the reaction ash: protons and alpha particles.

It has been shown that using a system of forced selective ash removal in the given energy interval makes it possible to substantially increase $Q_{\rm pl}$ all the way to values that characterize a commercial reactor. Based on these results, the requirements on an ash-pumping system have been formulated.

A KINETIC MODEL OF A MULTICOMPONENT D–^{3}He PLASMA

The carriers evolved in the energy-fusion reaction are the ash nuclei: protons, alpha particles, etc. Therefore, a key problem in solving the formulated task is to correctly calculate two processes: energy transport from the ash components to the fuel ions and electrons and the accumulation of ash in the plasma. It is obvious that these processes are closely interrelated. The former is actually unambiguously determined by the confinement time of the protons and alpha particles. An increase of τ_p and τ_α increases the fraction of energy transferred from the ash to the fuel and to the electrons. Because of this, $P_{\rm req}$ decreases in reaction (2). However, an increase of τ_p and τ_α means a simultaneous increase in the ash accumulation β_Σ, and this reduces $P_{\rm fus}$, since the fuel concentration decreases. This means that the conditions for confinement and power exchange of the ash in the plasma can have a substantial effect on $Q_{\rm pl}$, with certain optimal conditions being required to achieve the maximum values of $Q_{\rm pl}$.

It is obvious that the appropriate kinetic equations must be solved to calculate the components of the ash, since their distribution functions are substantially different from Maxwellian, while direct losses in velocity space can be taken into account only by using a kinetic model. As far as the fuel nuclei and the electrons are concerned, their distribution functions over energy can be considered uniform with good accuracy, and balance equations based on the corresponding Spitzer and Pastukhov formulas can therefore be used for them.

The general system for calculating a multicomponent plasma is as follows: The main given parameters of the system are the fuel temperature $T_{\rm fuel}$; the ratio $n_{\rm D}/n_{\rm He}$ of the components of the fuel; B_0 and B_p, the vacuum values of the magnetic fields in the central section and in the plug; β_Σ, the sum value of the beta of the plasma; and the ratio $e\Delta\varphi/T_{\rm fuel}$ of the ion-confining potential to the fuel temperature.

The distribution functions of the components of the ash and the corresponding power fluxes, both between various plasma components and escaping from the central cell, are determined by solving the kinetic equations with the appropriate boundary conditions. Moreover, the calculations give the power $P_{\rm fus}$ released by the thermonuclear reactions, the electron temperature T_e, the floating potential φ_0 that confines the electrons of the central cell, and the power $P_{\rm req}$ required to maintain the given fuel temperature. These data are used to determine the $Q_{\rm pl}$ value that characterizes the energy production in the plasma.

The kinetic equations are written on the assumption that plasma in the central cell of an ambipolar reactor is homogeneous. A feature of the model under consideration is that it allows for elastic nuclear scattering by combining the Fokker–Planck operator with the Boltzmann operator. This is associated with the fact that the Coulomb and nuclear scattering cross sections are comparable even at an energy of 5 MeV, whereas the elastic nuclear scattering cross section becomes greater at large energies (because it is energy-independent in the given region). In the final analysis, this affects the energy distribution of the fast ash particles between the fuel nuclei and the electrons. The kinetic equation has the form[4–6]

$$\frac{\partial f}{\partial t} = \left(\frac{\delta f_a}{\delta t}\right)_{FP} + \left(\frac{\delta f_a}{\delta t}\right)_B + S_a - L_a, \tag{9}$$

where $f_a(v,r,t)$ is the distribution function of particles of species a (here a is p, ^{4}He, T, or ^{3}He), S_{ta} is an operator describing the sources, and L_a is an operator describing the sinks of particles of species a. The normalization condition for the distribution function is

$$n_a = \int_0^\infty f_a(v)\,dv, \tag{10}$$

using

$$f(v)_{v \geqslant v_{\max}} = f(v_{\max}) = 0. \tag{11}$$

Here the Fokker–Planck collisional operator is

$$\left(\frac{\delta f_a}{\delta t}\right)_{FP} = \Gamma_a \left[-\frac{\partial}{\partial v_i}\left(f_a(v,t)\frac{\partial h_a(v)}{\partial v_i}\right) + \frac{1}{2}\frac{\partial^2}{\partial v_i \partial v_j}\left(f_a(v,t)\frac{\partial^2 g_a(v)}{\partial v_i \partial v_j}\right)\right], \tag{12}$$

where

$$\Gamma_a = \frac{z_a^2 e^4}{4\pi\varepsilon_0^2 m_a^2}.$$

The Rosenbluth potentials take the form[4,5]

$$g(v) = \sum_b \frac{z_b^2}{z_a^2}\lambda^{ab}\int_{v'} f_b(v,t)\,|v - v'|\,dv', \tag{13}$$

$$h(v) = \sum_b \frac{z_b^2}{z_a^2}\lambda^{ba}\frac{m_a + m_b}{m_b}\int_v \frac{f_b(v',t)\,dv'}{|v - v'|}, \tag{14}$$

$$\lambda^{ba} = \ln\left[\frac{m_a m_b}{m_a + m_b}\frac{2\alpha c \lambda_D}{z_a z_b e^2}\max\left(\frac{2\langle E\rangle}{m}\right)_{a,b}^{1/2}\right] - 0.5, \tag{15}$$

where λ^{ba} is the Coulomb logarithm, α is the fine-structure constant, c is the speed of light,

$$\lambda_D^{-2} = \frac{3e^2}{2\varepsilon_0}\sum_b \frac{m_b z_b^2}{\langle E\rangle_b}$$

is the Debye radius, $\langle E\rangle_b$ is the mean energy of particles of species b, and ε_0 is the permittivity of free space.

The summation is carried out over all species, including the test particles; the integration is carried out over all of velocity space; $(\delta f_a/\delta t)_B$ is the Boltzmann collision operator, which describes nuclear near collisions. Here n_b is the density of background particles, $\sigma(v\rightarrow v',dv)$ is the scattering cross section, which transports a particle with coordinates v' to an element of space dv with coordinate v. In spherical coordinates, $dv=v^2dvd\mu$. Then in simplified form,

$$\left(\frac{\delta f_a}{\delta t}\right)_B=\sum_b\left(\frac{n_b}{v\gamma}\int_v f_a(v')\sigma_t(v'\rightarrow v)v'dv'-f_a(v,\mu)\sigma_t(v)vn_b\right). \tag{16}$$

Here $\gamma=2m_am_b/(m_a+m_b)^2$, and $\sigma_t(v')=\sigma(v'\rightarrow v,dv)v'^2\gamma/vdv$ is the total collision cross section. The summation in Eq. (16) is carried out over the fuel ions, while the total collision cross sections are determined for all pairs of particles a–b.

In order to close the system of equations, it is necessary to determine the potential difference $\Delta\varphi_a$ between the plasma and the wall of the vacuum chamber. We write the energy-balance equations for a multicomponent plasma as

$$\frac{d(1.5n_aT_e)}{dt}=-\frac{n_a}{\tau_e}(T_e-e\Delta\varphi_e)-\frac{1.5n_eT_e}{(\tau_R)_e}+\sum_i Q^{i-e}-Q^{\mathrm{rad}}. \tag{17}$$

Here T_e is the electron temperature, Q^{a-b} is the power transferred from component b to component a, Q^{rad} is the radiational losses (bremsstrahlung and cyclotron radiation of the electrons), $\tau_{i,e}$ is the longitudinal (axial) confinement time of ions or electrons (the Pastukhov time), and τ_R is the radial confinement time.

For Maxwellian particles, the balance equation for the particles is written as[4,5]

$$\frac{dn_a}{dt}=-n_a\left(\frac{1}{\tau_a}+\frac{1}{\tau_R}\right)+S_a. \tag{18}$$

The electron concentration is determined from the conditions for ensuring that the plasma is quasi-neutral. The ambipolar equation, which relates the flow of ions and electrons from the confinement region, is used for this

$$\frac{dn_e}{dt}=\sum_i z_i\frac{dn_i}{dt}. \tag{19}$$

In solving the kinetic equations, boundary conditions that take into account the presence both of magnetic plugs and of the confinement potential are imposed on the distribution function. Since the plasma in the central cell is considered homogeneous, a rectangular shape is assumed for both the magnetic well and the confinement potential. The following important quantities can be determined by solving the kinetic equations along with the distribution function over energy: the mean energy of the particles, the energy flux from the ash particles to the electrons and fuel nuclei, the particle flux into the loss cone, and their distribution function.

We shall solve Eq. (9) with the following boundary conditions: The confinement region in velocity space is determined by the plug ratio and by the electron potential in the thermobarrier. Taking into account the charges and masses of the particles, the boundary of the confinement region has the form of a hyperboloid of one sheet or a hyperboloid of two sheets with the boundary equation

$$\mu_l^2=1-\frac{1-2ez_a\Delta\varphi_a/m_av^2}{R}, \tag{20}$$

where $\mu=\cos\Theta$, Θ is the pitch angle, $\Delta\varphi_a$ is the potential jump at the boundary of the central cell, and $R=B_p/B_0$ is the plug ratio.

Since the time of flight of the particles along a trap is much less than the mean time between collisions, it can be assumed that

$$f_a(v,\mu_l(v),t)=0,\ \text{for}\ \mu\geqslant\mu_l(v), \tag{21}$$

i.e., a particle incident into the loss region is instantaneously lost from the trap. Because the distribution function is small in the high-energy region,

$$f_a(v_{\max},\mu,t)=0. \tag{22}$$

The distribution function is independent of the pitch angle when $v=0$, i.e.,

$$\frac{\partial f_a(v=0,\mu,t)}{\partial\mu}=0. \tag{23}$$

Since the distribution function is symmetric in velocity space, we get

$$f_a(v,\mu,t)=f_a(-v,\mu,t), \tag{24}$$

$$\frac{\partial f_a(v,\mu=0,t)}{\partial\mu}=0, \tag{25}$$

$$\frac{\partial f_a(v,\mu=1,t)}{\partial\mu}=0. \tag{26}$$

The main goal of the calculations is to determine whether inequalities (7) and (8) are satisfied under the assumed conditions. Two series of calculations were carried out at the first stage. It was assumed in the first series that there is no radial transport; i.e., the time of radial transport of the particles is $\tau_R\rightarrow\infty$.

The corresponding results are shown in Fig. 1. It can be seen that Q_{pl} does not exceed 1.3, and thus inequality (7) is not satisfied. It should be pointed out that the radial transport must be taken into account, since the longitudinal confinement times are very long, greater than the classical transverse transport time.

In this connection, a second series of calculations was carried out that involved studying how the intensity of the radial transport of particles τ_R affects Q_{pl} and P_{fus}. It was assumed in this case that τ_R is identical for all the compo-

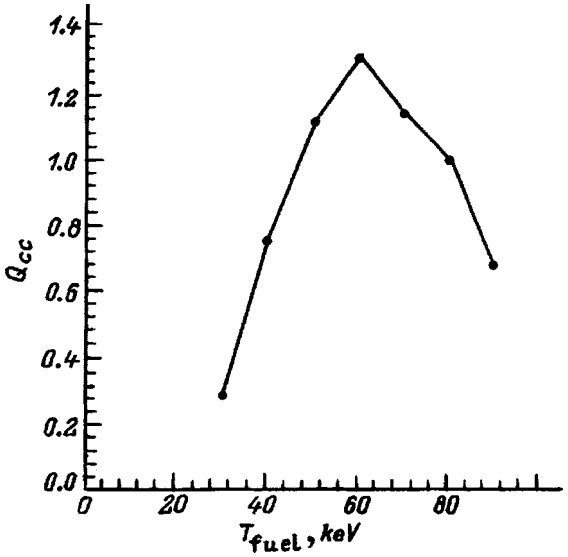


FIG. 1. Plasma power multiplication factor versus fuel temperature. $\beta_\Sigma = 0.7$, vacuum magnetic field $B = 5$ T.

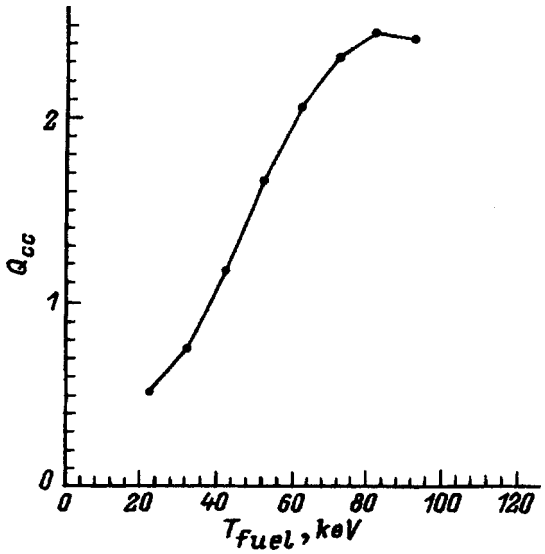


FIG. 3. Plasma power multiplication factor versus fuel temperature for a fixed characteristic radial-transport time. $\beta_\Sigma = 0.7$, $B = 5$ T, $\tau_R = 20$ sec, $\Delta\Phi e = 4\, T_{\text{fuel}}$.

nents of the plasma (electrons, fuel, ash). To do this, the kinetic equations were supplemented with the following parametric operator:[4]

$$L_R^a = \frac{f_a}{\tau_R}. \tag{27}$$

Figures 2 and 3 show the τ_R dependences obtained for Q_{pl} and P_{fus}. It can be seen that an optimal value exists, at which Q_{pl} is maximal.

The following section presents the results of the third series of calculations, which describes the possibilities of increasing Q_{pl} in comparison with the first two versions.

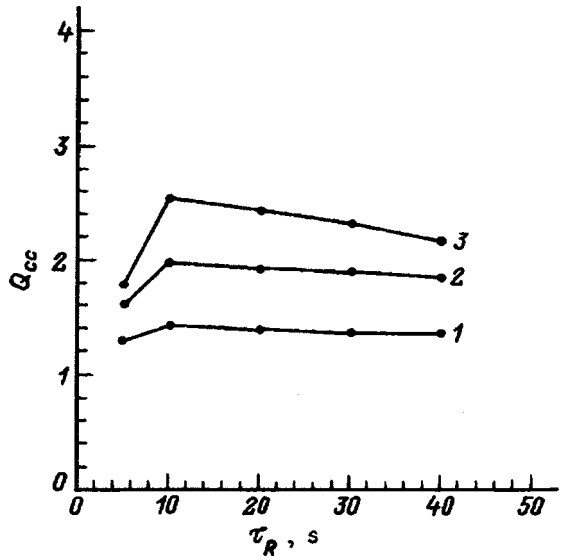


FIG. 2. Plasma power multiplication factor versus characteristic radial-transport time τ_R. T_{fuel}, keV: *1* — 70, *2* — 90, *3* — 80.

ANALYZING THE KINETICS OF A D–^{3}He PLASMA TAKING INTO ACCOUNT ASH PUMPING

The cause of the low energy efficiency (the low Q_{pl} value) is associated with the accumulation of the fusion reaction products.[2,4] By imparting their energy to the fuel ions and electrons, the ions p and α are retarded and confined by the magnetic system. The low-energy particles in this case play a parasitic role, extracting the thermonuclear power for themselves. An analysis of the results indicates one possibility of increasing Q_{pl}. This is to selectively remove fusion reaction products, most importantly protons and alpha particles, from the plasma. By exciting weak external nonsteady-state fields for protons and alpha particles, conditions are provided under which their radial transfer time across the magnetic field is significantly less than for the other components (see Appendix A).

The main goals of the calculations were to establish the limiting values of Q_{pl} in the presence of ash pumping and to determine the requirements imposed on the parameters of the system for removing the ash particles.

We shall assume below that the reactor is equipped with a special system that ensures selective pumping of the ash components. Relative to the ash-pumping system, we assume that it ensures removal of particles of a definite energy range. We assume that there is a mechanism that varies both the width of this range and its position on an energy scale.

Postulating the indicated properties, we can estimate the limiting energy production in a D–^{3}He plasma (the limiting values Q_{pl} and P_{fus} in the plasma). In fact, the preferred ash-removal zone is the region where the particles have relatively low energies (in comparison with their generation energies ε_0). Then the particles transport the maximum possible energy to the plasma in the time it takes them to slow down, while simultaneously ensuring relatively small β_{ash} values.

On the other hand, if the pumpout mechanism is univer-

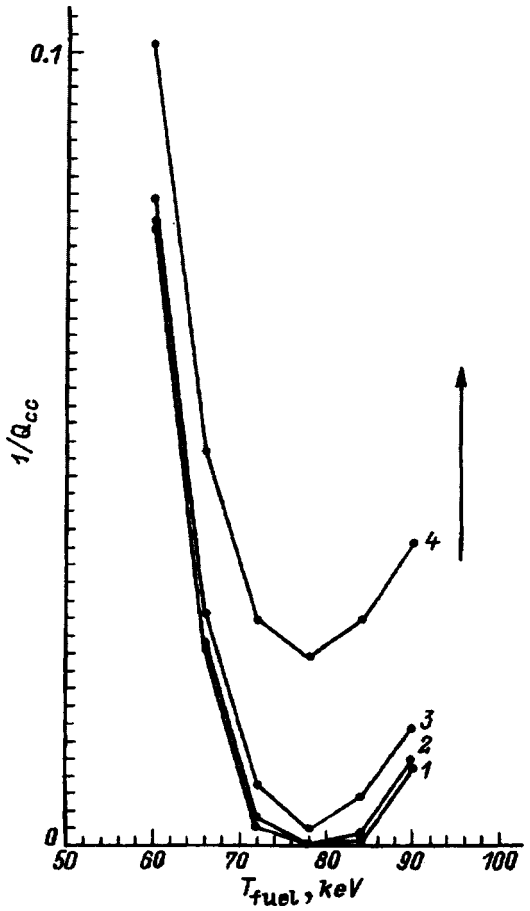


FIG. 4. Effect of the width of the pumping-energy range on the dependence of the plasma power multiplication factor on the fuel temperature. The low-energy boundary of the pumpout region is $\varepsilon_1=400$ keV, $\beta_\Sigma=0.7$, $B=5$ T, $\tau_R=20$ sec, $\Delta\Phi e=4T_{\text{fuel}}$. $\Delta\varepsilon$, keV: *1* — 5, *2* — 20, *3* — 80, *4* — 320.

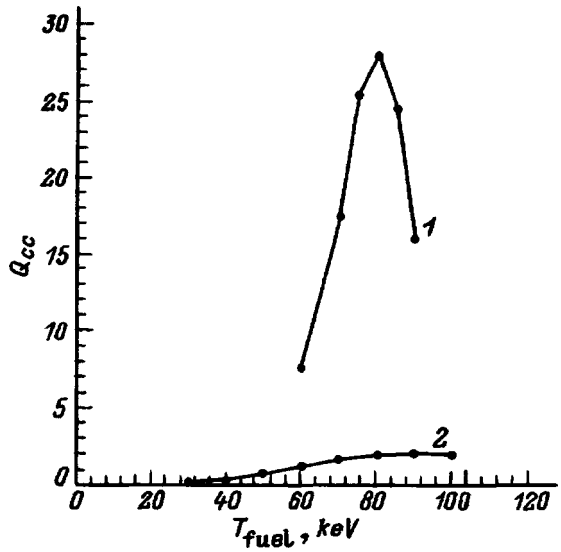


FIG. 5. Plasma power multiplication factor versus fuel temperature. $\beta_\Sigma=0.7$, $B=5$ T, $\tau_R=20$ sec, $\Delta\Phi e=4T_{\text{fuel}}$. *1* — with pumping in the energy range 300–400 keV, $\tau_p=0.02$ sec; *2* — with no pumping.

sal, i.e., if it depends only on the energy of the particles and is independent of their individual properties (mass and charge), the pumping interval must be appreciably above the fuel temperature. Otherwise the fuel loss, and hence P_{req}, will increase sharply.

The mathematical model of Eqs. (9)–(27) was used in the calculations. The kinetic equations for protons and alpha particles contain an additional operator, corresponding to the removal of these particles,

$$L_{ap}=\frac{f_a}{\tau_{ap}}. \tag{28}$$

The corresponding losses are also taken into account in the energy-balance and particle-balance equations. Two quantities were varied in the calculations: the pumpout time τ_{ap} and the width of the energy interval $\varepsilon_2-\varepsilon_1$ in which the pumping was done. The pumpout region is located in the lower part of the energy scale in such a way that it substantially exceeds the fuel temperature T_{fuel}. As a consequence, it is possible to reliably assume that the perturbations do not affect the transverse confinement of the fuel particles.

The pumpout parameters are found[6] (Fig. 4) for which the energy efficiency is highest. The Q_{pl} parameter corresponding to this regime is shown in Fig. 5.

Let us dwell on an important feature inherent to Q_{pl} — the maximum close to the temperature $T_{\text{fuel}}=80$ keV. A detailed analysis of the behavior of $P_{\text{fus}}(T_{\text{fuel}})$ on temperature and on all the quantities that enter into P_{req} showed the following: On one hand, P_{fus} increases in the entire temperature interval studied here, since $\langle\sigma v\rangle=f(T_{\text{fuel}})$ increases. True, this increase slows down, since n_{D} and n_{He} decrease with temperature. However, Q_{pl} increases until $T_{\text{fuel}}=80$ keV. On the other hand, beginning with $T_{\text{fuel}}=70$ keV, the losses to cyclotron radiation sharply increase, which causes P_{req} to increase, and Q_{pl} consequently decreases.

It was determined from calculations that the main fraction of the thermonuclear power (about 90%) reaches the first wall of the reactor with the radiation. This means that, for such a version of the reactor, a relatively low-temperature cycle of electrical power production can be implemented with a correspondingly low efficiency. A direct conversion system is unsuitable in this case.

Pumping out the ashes substantially alters the picture of the power fluxes in a thermonuclear plasma. The relative losses with radiation from the plasma are appreciably reduced, and the fraction of power carried away from the plasma by particles via the loss cone increases. It may be promising in this case to establish direct conversion systems capable of appreciably increasing the efficiency of the electrical power source.

CONCLUSION

This paper has shown that it is necessary to selectively pump the ash out of a D–^{3}He thermonuclear plasma to attain the required energy efficiency of a low-radioactivity thermonuclear reactor.

The use of selective pumpout substantially alters the energy balance in the plasma. In this case, the relative radia-

tional losses are reduced, the escape of fuel ions into the loss region is decreased, and the requirement for additional heating is significantly reduced.

This work was financed by Grant MNF J5Y100.

APPENDIX SELECTIVE ION PUMPING FROM AN AXIALLY SYMMETRIC MAGNETIC TRAP

We earlier developed a method of selectively removing the ions from an axially symmetric magnetic trap. The method was described in detail in Ref. 7. Here we briefly describe the main principles of the method. The proposed mechanism for radially removing ions is based on the following concepts:

1) Generation of a weak perturbing magnetic field, which introduces asymmetry into the main field of the trap. When a definite type of nonsteady-state perturbations are chosen, the problem of moving a charged particle across a magnetic field reduces to the problem of moving a nonlinear pendulum perturbed by a periodic force.

2) Rotation of the perturbing multimode magnetic field around the axis of the system with frequencies close to the azimuthal drift frequency of the particles to be pumped out. This ensures the resonance properties of the perturbing action and selects the particles in terms of rotation direction and in terms of energy.

It is well known[8] that even weak periodic actions on a nonlinear pendulum cause a qualitatively new property to appear — chaotic dynamics. This opens up the possibility of organizing stochastic collisionless ion transport across a magnetic field.

When several modes of the perturbing field are superimposed, the topologies of the motion of the particle on the phase plane can be described by the standard mapping

$$\tilde{I}=I+K_0 \sin \Theta, \quad \tilde{\Theta}=\Theta+\tilde{I}, \ \mathrm{mod}(2\pi). \tag{A1}$$

Here K_0 is the Chirikov coefficient,

$$K_0=4\pi\frac{AC}{\Delta\omega^2}, \tag{A2}$$

where $\Delta\omega$ is the frequency shift between two adjacent modes of the perturbing field,

$$A=\frac{4m}{eB}\left(v_\parallel^2+\frac{1}{2}v_\perp^2\right)\frac{r^*}{(2L^2-r^2)^2}, \tag{A3}$$

$$C=\frac{e}{m}\left(v_\parallel^2+\frac{1}{2}v_\perp^2\right)\frac{B^*R_{\mathrm{per}}}{B^2L^2}. \tag{A4}$$

Here e and m are the charge and mass of the particle being pumped out, B and L are the main magnetic field and its doubling parameter, R_{per} is the radius at which the windings of the perturbing field are located, B^* is the perturbing field, $v_\parallel$ and $v_\perp$ are the parallel and longitudinal components of the particle velocity, and r and r^* are the radial coordinate and the value at which the azimuthal drift frequency is accurately equal to the rotation frequency of the perturbing field.

The criterion for global stochasticity is $K_0>4.5$. Changes of the radial coordinate over the entire width of the stochastic layer are possible for a particle in the stochastic regime. A particle that was formerly in resonance with the first perturbation is displaced along a radius and is caught by the next perturbation, etc. A particle far from resonance with the perturbation will experience only a weak flutter of the drift surface. The width of the layer is determined by the magnetic field of the perturbation and by the frequencies of the adjacent perturbations.

Under the action of external perturbations, a stochastisized particle can reach such a radial coordinate that it either falls onto the limiter or is lost through the plug because of the decrease of the confining potential far from the axis of the system. To organize such a particle sink, the perturbation parameters are chosen so that the stochastic layer overlaps the entire distance from the axis of the system to the limiter.

The stochastic diffusion coefficient is derived from the standard mapping and has the form

$$D_r=\pi\frac{C^2}{\Delta\omega}. \tag{A5}$$

This relationship was derived in the chaotization approximation of the phases of particle motion in one iteration period of the standard mapping. The time to remove particles from the central section of the reactor is $\tau=0.02$–0.05 sec, with the necessary values of the perturbation field being 1–2% of the main value.

[1] W. Hefele, J. P. Holdren, G. Ressler, and G. L. Kulcinsski, *Fusion and Fast Breeder Reactors*, International Institute for Applied System Analysis Luxemburg (Austria), 1976, Vol. A-2361.
[2] I. N. Golovin, V. I. Khvesyuk *et al.*, in *Proceedings of the Course and Workshop Held at Villa Manastero, Varenna, Italy, 1989*, pp. 673–692.
[3] V. V. Kostenko, V. I. Khvesyuk, and N. V. Shabrov, At. Energ. **68**(3), 188 (1990).
[4] D. V. Semionov, V. I. Khvesyuk, and D. L. Poletaev, in *Proceedings of the International Conference on Open Plasma Confinement Systems for Fusion, Novosibirsk, 1993*.
[5] I. N. Golovin, V. I. Khvesyuk, D. V. Semionov *et al.*, in *Proceedings of the Second Wisconsin Symposium on He-3 and Fusion Power, Madison, Wisconsin, 1993.*
[6] I. N. Golovin, V. I. Khvesyuk, and D. V. Semionov, *Sixth International TOKI Conference on Plasma Physics and Controlled Nuclear Fusion*, TOKI City, Japan, 1994.
[7] V. I. Khvesyuk, N. V. Shabrov, and A. N. Lyakhov, in *Proceedings of the International Conference on Open Plasma Confinement Systems for Fusion* (World Scientific, Singapore, 1993).
[8] R. Z. Sagdeev, D. A. Usikov, and G. M. Zaslavsky, *Nonlinear Physics: from the Pendulum to Turbulence and Chaos* [Harwood, Chur, 1988; Nauka, Moscow, 1988].

Translated by W. J. Manthey

Study of the leader of a spark discharge over a water surface

V. P. Belosheev

S. I. Vavilov State Optical Institute All-Russia Science Center, 199034 St. Petersburg, Russia
(Submitted December 2, 1996)
Zh. Tekh. Fiz. **68**, 44–50 (July 1998)

This paper discusses an experimental study of the leaders of incomplete spark discharges with a capacitance of 0.1 and 1 μF over a water surface when the initial voltage is 3–6 kV in discharge gaps 8 and 22 cm long, having side branches and without branches. The distributions of the field, the current, the current density, the conductivity, and the electron concentration along the leader have been determined, as well as the changes in the velocity and length of the leader as it evolves. It has been established that the evolution of the leader has a self-maintained character, and that the product of the storage capacitance and the initial potential difference between the head of the leader and the water surface is an invariant of its spatial evolution. © *1998 American Institute of Physics.* [S1063-7842(98)00808-3]

INTRODUCTION

It was proposed in Ref. 1 to use a spark discharge over a water surface (SDWS) as a source of UV radiation for deactivating microorganisms in a volume of water. The formation of an SDWS for these purposes was studied in Ref. 2. At the same time, the experiments presented there showed that, being in essence a sliding discharge that evolves by a leader mechanism, an SDWS has an essential feature in comparison with a sliding discharge over a solid insulator, associated with the Ohmic conductivity of water. This circumstance, in particular, makes it possible to form a leader over water in a wide range of initial voltages and accordingly with extremely different parameters: current, length, and velocity, under conditions of complete or incomplete discharge. The last case is particularly suitable for studying the evolution of a leader.

In terms of a study of the leader mechanism of SDWS evolution, Ref. 2 concentrated on the formation phase of a leader during the competition of several initial leaders. This paper is devoted to a study of the dynamics of the evolution of a leader and the parameters of its channel. This information can be useful for understanding the leader mechanism of evolution of long spark discharges in the atmosphere, including lightning, since it is rather hard to obtain all the necessary experimental information concerning large-scale discharges in the atmosphere.[3]

EXPERIMENTAL LAYOUT AND CONDITIONS

SDWS leaders were formed in rectangular $23\times4\times1.5$-cm Plexiglas cell *1* (Fig. 1a), filled with tap water *2* with a conductivity of $\simeq1\times10^{-4}$ S/cm. Conical anode *3*, composed of steel wire 0.1 cm in diameter, had a radius of curvature of $\sim2\times10^{-2}$ cm and was placed in the initial state at a distance of $\simeq1$ cm from the water surface. Cathode *4* was a stainless steel wire 0.16 cm in diameter and 8 or 22 cm long, placed in the water at a distance of 0.3–0.8 cm from the water surface. When the thickness of the water layer was less, breakdown occurred. Storage capacitor *5*, with a value of 0.1 or 1 μF, was charged to an initial voltage of $U_0=3-7.5$ kV, measured by kilovoltmeter *6*. After this, a leader was initiated by reducing the distance between the anode and the water surface to 0.1–0.3 cm, respectively.

The total current i in the discharge circuit (Fig. 2) was measured by means of shunt *7*, the voltage from which was applied to an oscilloscope. The voltage U_c on the capacitor (Fig. 2) was displayed on the oscilloscope using 1 : 1000 Ohmic voltage divider *8* with resistances of 5 $k\Omega$ and 5 MΩ. Moreover, wire probe *9*, 0.01 cm in diameter and touching the water surface above the cathode, and 1 : 1000 Ohmic voltage divider *10*, with resistances 10 kΩ and 10 MΩ, were used to obtain an oscilloscope tracing relative to ground of the potential of the water surface along the discharge gap, as well as of the leader channel (Fig. 3).

The structure and geometry of the leader channels were studied visually through a microscope with magnification from $8\times$ to $30\times$ with a measurement grid, which made it possible to measure the diameter of the channel in any cross section, as well as from overall photographs. The leaders were photographed on isopanchromatic-22 film with a frame size of 35×24 at a scale of 1 : 2 and 1 : 1. An overall photograph of the leader, evolving on free water with $U_0=6$ kV and capacitance 0.1 μF and with a water layer 0.3 cm thick, is shown in Fig. 4a.

However, most of the experiments were carried out under conditions in which the leader was formed along a slit 0.16 cm wide between two or four quartz bars *11* lying on the bottom of the cell next to the cathode and protruding above the water surface by $\simeq0.05$ cm (Fig. 1a). Figure 4b shows the leader under such conditions for $U_0=6$ kV, $C=0.1\mu$F and a water layer 0.3 cm thick. In this case, besides the total current, the current i_l of the leader itself was displayed on the oscilloscope (Fig. 5) using Rogowski loop *12* wound on a ferrite ring with outer and inner diameters of 2.5 and 1 cm, respectively, and 0.25 cm thick.

The ring, coated with a layer of wax, was partially immersed in the water between the four quartz bars so that the

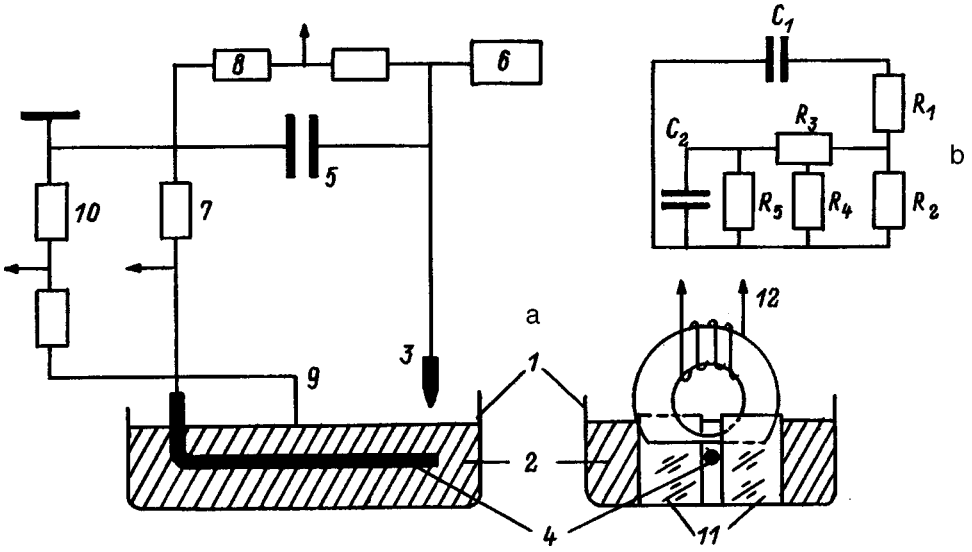


FIG. 1. Layout of experimental apparatus (a) and equivalent circuit diagram (b) of a discharge circuit with a leader of an incomplete SDWS.

leader passed through the ring. The voltage from a coil wound on the other part of the ring and also coated with wax was proportional to the ratio of the number of its turns to the load resistance and was fed to the oscilloscope. The current of the leader could be measured at any of its cross sections by displacing the bars with the ferrite ring along the leader.

EXPERIMENTAL RESULTS

In both versions of SDWS formation, the main experiments were done with a 0.1-μF storage capacitance, a discharge gap 8 cm long, and a water layer 0.3 cm thick over the cathode. Under these conditions, when $U_0 \geqslant 7$ kV, a complete SDWS occurred in the gap, whereas, when $U_0 = 3-6$ kV, leaders of incomplete discharges $\simeq 1-5$ cm long, respectively, occurred. However, to elucidate why the parameters of the leader tended to vary as a function of the initial conditions, the water layer was increased to 0.8 cm in a number of experiments. This made it possible, other conditions being equal, to reduce the length of the leader and the maximum current in the circuit by about a factor of two. A 1-μF storage capacitance was also used. In this case, a complete SDWS developed in a discharge gap 22 cm long with a water layer 0.3 cm thick over the cathode and $U_0 = 6$kV. As the thickness of the water layer was increased to 0.8 cm, only the leader of an incomplete discharge occurred in the gap. Its length on free water reached 20 cm, while the maximum current in the circuit was 17 A.

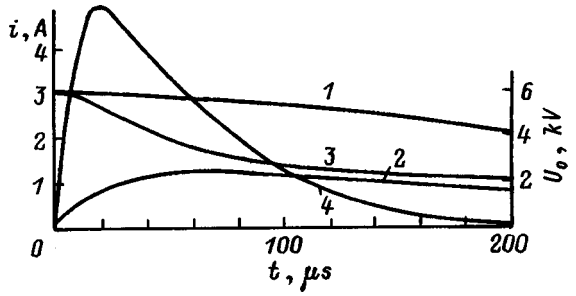


FIG. 2. Oscilloscope tracings of the voltage on the capacitor and the total current in the circuit. Slit: *1* — U_c, *2* — i; free water: *3* — U_c, *4* — i.

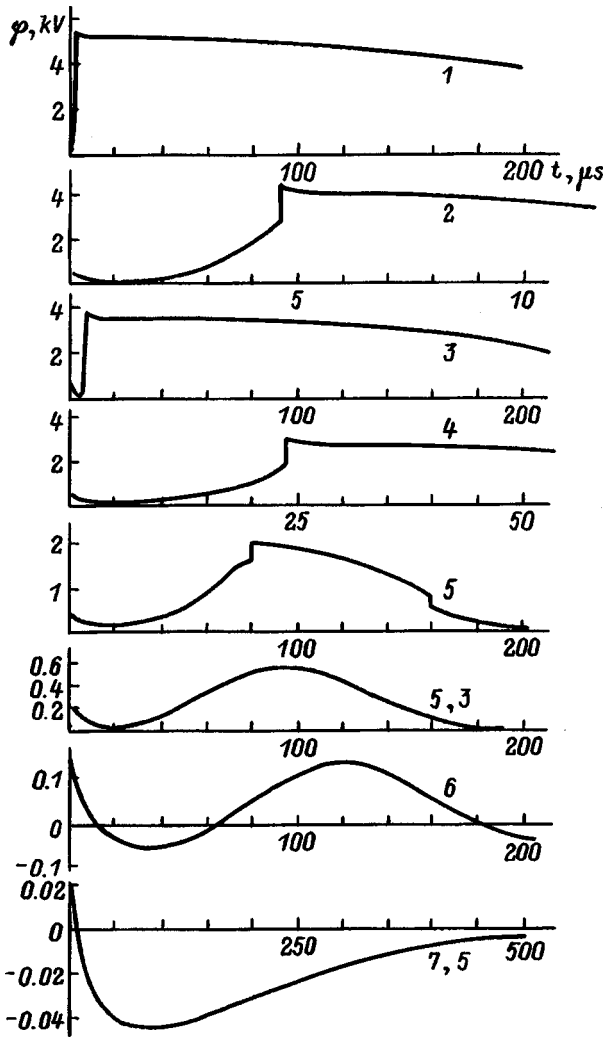


FIG. 3. Oscilloscope tracings of the potential in the discharge gap. The numbers on the curves show the distance from the anode to the probe in centimeters.

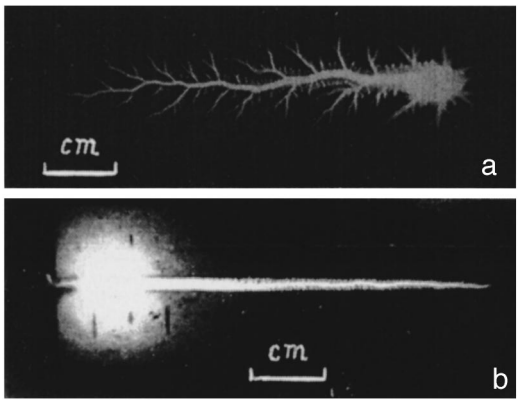


FIG. 4. Overall views of leaders. (a) free water, (b) slit.

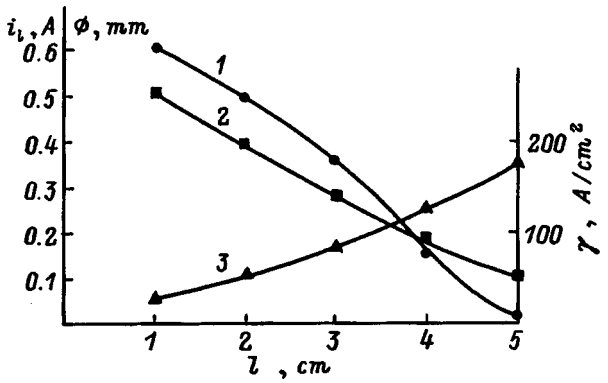


FIG. 6. Distribution of current, current density, and channel diameter along the length of a leader from the anode. *1* — i, *2* — $\varnothing$, *3* — γ.

In the range $U_0=3-6$ kV, on free water or in a slit, a leader developed after the distance between the anode and the water surface was reduced from 1 to 0.1–0.2 cm, respectively. In this case, a corona discharge initially developed in the air at the anode, and then it crossed the entire air gap in the form of a cone, supported by its base on the water, from which the leader sprouted along the water surface. All the parameters of the leader were unambiguously determined by the capacitance, the initial voltage, the thickness of the water layer, and the conditions on the water surface. With a capacitance of 0.1 μF and a water layer 0.3 cm thick, the length of the leader is $\simeq 1$ cm when $U_0=3$ kV and increases approximately linearly to $\simeq 5$ cm when $U_0=6$ kV. The length of the leader under these circumstances was 10% greater on free water. The brightness of the channel and its diameter decreased monotonically from the anode along the leader. The color of the luminescence was white on the anode side but was bluish-lilac at the head. The diameter of the leader channel on a photographic plate was a factor of 1.3 greater than that observed in the microscope. Figure 6 shows the averaged channel diameter over the length of the leader in the slit when $U_0=6$ kV and the water layer is 0.3 cm thick, with a total current amplitude in the circuit of $i=1.2$ A (Fig. 2). This current contains two components. One of them, $i_{\rm cr}$, passes through the air gap and the layer of water under the anode, at which the corona discharge is developed, while the other, i_l, passes through the air gap, through the leader channel along the entire length, and through the water layer under it. It is the intrinsic current of the leader. On an oscilloscope tracing of the total current i, recorded by the shunt, the corona phase of $i_{\rm cr}$ lasts $\simeq 0.1\mu$s (the ionic conductivity of water is established in a time of $\simeq 1$ ns). The amplitude $i_{\rm cr}=0.2$ A reached in this time depends on U_0 and the conductivity of the water layer. The further increase of i in the circuit is associated with the evolution of the leader. The time to attain the maximum i depends on U_0, the thickness of the water layer, and the conditions on the water surface. The maximum amplitude of i is a factor of 2.5 greater on free water, while the time to reach it is a factor of 2.5 less than when the leader evolves in a slit (Fig. 2). Using oscilloscope tracings of the leader current at different cross sections along its length (Fig. 5), it is possible to construct the current distribution along the channel of the leader from the anode at any instant. Figure 6 shows the current distribution of the leader for $U_0=6$ kV, which reaches its maximum length of 5.2 cm at 80 μs. Along with the dependence of the cross sectional area of the channel on the distance to the anode, this makes it possible to compute the current densities in the leader channel over its length (Fig. 6).

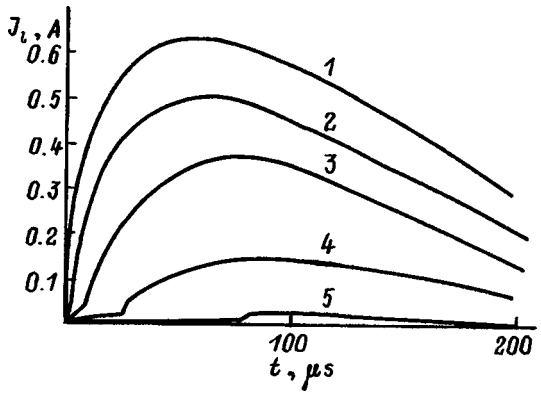


FIG. 5. Oscilloscope tracings of the leader current. The numbers on the curves show the distance from the anode to the Rogowski loop in centimeters.

The oscilloscope tracings of the voltage at the capacitance (Fig. 2) have two characteristic sections: a close-to-linear falloff in the region of maximum i and then a slower falloff, when i decreases to zero, while the voltage decreases to $\simeq 0.3U_0$. The oscilloscope tracings of U_c and i make it possible to determine how the resistances of the discharge circuit depend on time under various conditions on the water surface (Fig. 7).

The oscilloscope readings of the probe potential at different points of the discharge gap (Fig. 3) make it possible to establish the variation of the potential of the water surface over the cathode (φ_B) until the head of the leader arrives at this point of the gap (the concave, growing part of the oscilloscope tracing), the delay time of the arrival of the head of the leader, and its potential (φ_h) (the sharp increase of the signal), as well as the change of the potential of the water surface at those points of the discharge gap that the leader does not reach ($l \geqslant 5.3$cm). Note that the positive overshoot of the beginning of the oscilloscope tracing is associated with the charge of the probe capacitance relative to ground.

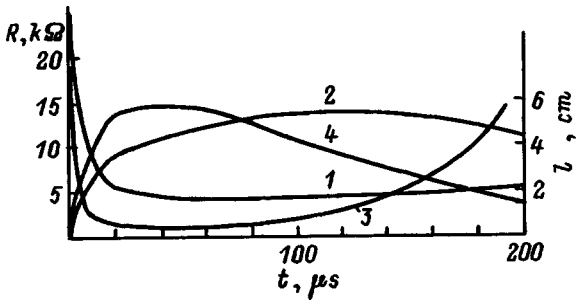


FIG. 7. Gap resistance and length of the leader versus time. Slit: *1* — R, *2* — l; free water *3* — R, *4* — l.

Based on such oscilloscope tracings, the potential distribution over the length of the leader and of the entire gap can be constructed for any instant. Figure 8 shows such a distribution for 80 μs, when the leader has come to a stop, as well as the field along the channel at this instant. These oscilloscope tracings from the probe can also be used to determine how the potential of the head and the water surface in front of it and the potential difference between them (ΔU) depend on the length of the leader (Fig. 9). Moreover, by using data on the delay time of the arrival of the head at various points of the interval (Fig. 10), it is possible to determine the mean velocity of the head in the intervals between these points (the length of the interval is 1 cm). Its dependence on the length of the leader is shown in Fig. 10.

The oscilloscope tracings of the probe potential also make it possible to establish the instant at which the head of the leader breaks away from the probe from the sharp potential drop by 200–300 V at the trailing edge of the oscilloscope tracing (Fig. 3, $l=5$ cm). This makes it possible to construct the time dependence of the length of the leader not only at the stage when it is developing but also at the stage when its length is shortening after the current reaches its maximum (Fig. 7).

DISCUSSION OF THE RESULTS

In considering the evolution of the leader of an incomplete SDWS, five aspects can be distinguished: the initial stage of the appearance of the leader, the structure of the channel, the leader as an element of an RC circuit, plasma formation in the head and its motion, and the parameters of the plasma of the leader channel. We will discuss our results in these terms.

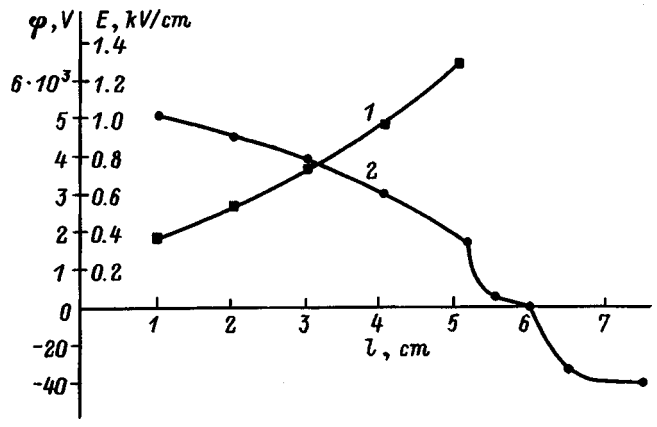


FIG. 8. Potential and field distribution in the discharge gap at the time the leader stops. *1* — E, *2* — φ.

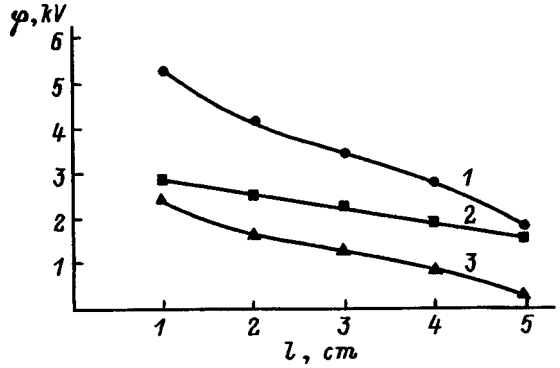


FIG. 9. Potential of the head and the water surface and the potential difference between them versus the length of the leader. *1* — potential of the head, *2* — potential of the water surface, *3* — potential difference between the head and the water.

However, it is first necessary to make a remark concerning terminology. Terms are used in this paper to describe an SDWS leader that are customarily used to describe the leaders of long spark discharges in air. Although it is clear that the breakdown of a discharge gap along the boundary of two media with substantially different ε that is smaller by a factor of 10–100 cannot be wholly analogous to the breakdown of a long air gap, it is nevertheless possible to note a qualitative similarity of the initial pulsed corona from the anode in the two cases, while the seeming absence of streamers from it and the head is associated with their short length, corresponding to the radii of the point of the anode and the head,[3] and with the brightness of the luminescence. These considerations indicate that leaders and not streamers were observed in Ref. 2. Thus, only the literal absence of a head at the end of the SDWS leader distinguishes its structure from that of the leader of a long spark discharge in air.

As can be seen from Fig. 3, the water surface over the cathode had a negative potential at the initial instant. Its

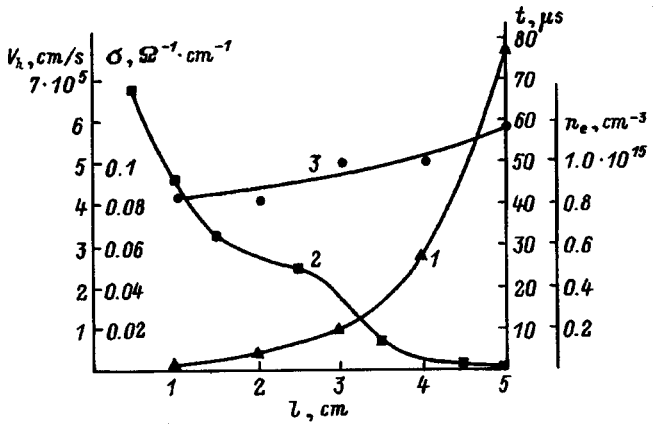


FIG. 10. Time for the head to move and velocity of the head versus the length of the leader and distribution of the conductivity and electron concentration of the plasma over the length of the channel at the instant the leader stops. *1* — t, *2* — V_h, *3* — σ and n_e.

value ($\simeq$40 V 7.5 cm from the anode) was determined by the ratio of the capacitance of the air gap and of the water layer above the cathode, as well as by the value of U_0. From the instant the corona discharge reaches the water surface, its potential increases and is now determined by the ratio of the resistance of the plasma and of the water layer under the base of the corona discharge. The leader begins to develop from this instant.

Judging from Fig. 4a, on free water, when the length of the gap and of the cathode are equal, leaders begin to develop in all directions (the leader to the side opposite the cathode is closed by the anode). However, only one leader subsequently developed along the cathode. If a cathode twice as long as the discharge gap was used, the leaders above the cathode developed in both directions from the anode virtually symmetrically. This is associated with the presence of quasi-one-dimensional potential–charge relief (PCR) on the water surface over the cathode.

In Ref. 2, the cathode was placed in the water vertically, and the initial PCR at the water was less expressed. Therefore up to four leaders initially developed from the anode on the side of the cathode, but, because of competition between them, only one was left after several seconds, causing the discharge gap to break down with a corresponding delay. The distribution of the time delay, having four maxima, was determined in this case by the character of the competition of the leaders. In our case, the PCR helped a leader develop over the cathode, and the competition in time was completed within one microsecond.

As can be seen from Fig. 4a, the structure of the leader on free water consists of a channel, side branches, and offshoots. When a slit was used, only the channel with offshoots was present, the presence of which is evidence that the PCR is not one-dimensional in this case. However, the decrease on free water of the overall density of structural elements of the leader as one goes away from the channel is obviously associated with a corresponding change of the density of the initial polarization charges on the water surface. The side branches and the offshoots over the entire channel correspondingly have about equal lengths and are located at equal intervals. However, their brightness and diameter are greater at the base of the channel and the side branches. This is apparently associated with the fact that they developed only a short time after the head of the leader, and afterwards their evolution ceased as the field decreased, whereas their brightness and diameter are connected with the amplitude and duration of the current that flows through them. The side branches are usually located asymmetrically along the channel, which is evidence of competition during the appearance of bifurcation. Moreover, the branch that developed close to the axis of the gap had a larger velocity because of PCR and then became a section of the leader channel. Thus, PCR is one reason that there is no head of conventional type at the end of the leader. The branching process must cause the leader to move nonuniformly and must cause oscillations of the current, but it is impossible to record them in this case, because of the large number of nonsynchronously developing elements of the leader.

It was shown in Ref. 2 that a discharge circuit with storage capacitance C and incomplete SDWS is an $R(t)\cdot C$ circuit. Under the conditions of this paper, the evolution of the leader ensured that the current in the circuit decreased linearly. This made it possible to solve the differential equation for the current in the circuit and to obtain an analytical expression for the gap resistance that had a minimum. This indicates that the leader during the course of incomplete SDWS plays the role of a nonlinear regulating element in the discharge circuit. In our work, with the lower conductivity of water and a larger discharge gap, the initial resistance was significantly larger and the current was smaller in the discharge circuit, and therefore the development of a leader not only provided rectification but also further increased the current in the circuit (Fig. 2). Its equivalent circuit is shown in Fig. 1b, where C_1 is the storage capacitor, R_1 is the resistance of the plasma in the anode–water gap, R_2 is the resistance of the water layer under the anode, R_3 is the resistance of the leader channel, R_4 is the resistance of the water layer between the leader channel and the cathode, R_5 is the resistance of the water layer in the region of the head of the leader, and C_2 is the capacitance of the water layer, which, taking into account the comparatively high Ohmic conductivity of water, did not appreciably affect the development of the leader. This circumstance distinguishes an SDWS from a sliding discharge over a solid insulator, in which the current of the leader has a capacitive character.

As can be seen from Fig. 7, the resistance of the discharge gap until a leader begins to develop equals $R_1+R_2 \simeq 25\ \mathrm{k}\Omega$. When a leader develops and the current increases, R_1 decreases and can be neglected, while the gap resistance is considered to be dependent on R_2-R_5. Moreover, if R_2 and R_5 can be considered constant, as the length of the leader increases, R_4 decreases, reducing the resistance of the gap and increasing the current of the leader. At the same time that R_3 increases with the growth of the leader, it simultaneously decreases with the increase of the current. In total, the resistance of the gap decreases at the growth stage of the leader current. The length of the leader, the current amplitude, and the gap resistance in this case reach extreme values simultaneously.

In the case of a leader on free water, the side branches, by increasing the contact area of the plasma with the water, further reduce R_4, increasing the current and thereby reducing R_3. In total, the gap resistance becomes still less, and the extremal values are reached faster, but again simultaneously. The subsequent shortening of the leader increases the gap resistance and protracts the falloff of the voltage on the capacitance.

As already indicated, the discharge in the air gap between the anode and the water begins from the point of the anode as a corona and then, judging from the current amplitude of $\simeq$0.2 A, changes into an anomalous glow discharge, with subsequent contraction and tendency toward an arc discharge. The last two forms of discharge have a positive column and a cathode layer at which a large part of the voltage drop occurs and a plasma is generated. Taking into account the cathode material, the kind of gas, and its density, the thickness of the cathode layer in our case is $\simeq 10^{-3}$ cm,[4] which is quite comparable with the observed diameter of the

head of the leader, $\varnothing \simeq 10^{-2}$ cm. This suggests that the leader generated at the boundary of the cathode layer at the water surface subsequently maintains the plasma-formation mechanism of the cathode layer, associated with high field strength and with the presence of fast electrons capable of effectively ionizing the air in front of the head. Since the diameter of the head remains unchanged, the plasma-formation efficiency will always be determined by the potential difference between the head and the water surface (Fig. 9). In this case, the field component normal to the water surface at the boundary of the head at a distance of 1 cm from the anode is $\Delta U/\varnothing = 250$ kV/cm. It is harder to determine the longitudinal component of the field in front of the head, since the potential distribution over the water surface in front of the head, associated with the presence in the water layer in front of the head of a longitudinal current component, can be established only qualitatively. However, if it is recalled that the characteristic offset length of the positive potential in front of the head is $\simeq 1$ cm (Fig. 8), the longitudinal field component will be $\simeq 6$ kV/cm. A comparison of the resulting values of the field components, on one hand, must assume that the plasma formation in the head is associated mainly with the normal field component, and, on the other hand, indicates that it is possible that the low density of streamers in front of the head and their short length is because the longitudinal field component is inadequate for their intense development.

The longitudinal field component, moreover, is the source of the Coulomb force that moves the head, but the motion of the plasma at the boundary of the head rather occurs because of transport processes during plasma formation. These two circumstances explain the fast decrease of the velocity of the leader because of the decrease of ΔU, while the falloff of the voltage on the capacitor is comparatively slow (Figs. 9 and 10).

The initial potential difference ΔU_0 and consequently V_0 are determined by the ratio of R_1 and R_2, as well as by the value of U_0. At the same time, as the leader evolves, the equality $U_c = U_5 + \Delta U + U_3 + U_1$ is valid (U_5, U_3, and U_1 are the voltage drops on R_5, R_3, R_1). Close to the maximum of the current, U_1 can be neglected, and then $U_c \simeq U_5 + \Delta U + U_3$; since U_c and U_5 drop off slowly, while U_3 increases as the leader grows, ΔU and V decrease. Moreover, the initial relative elongation of the leader is large, and therefore U_3 increases quickly, while ΔU and V fall off quickly. However, having gained a certain length, the leader, because of the connection between its length and the current, now shows a stabilizing action on V (Fig. 10). Since the falloff of V decreases the gain in length by the leader, the growth of i and U_3 and consequently the falloff of ΔU and V slow down. Nevertheless, when the length of the leader and the current increase continuously, a time comes when $U_5 + U_3 = U_c$, $\Delta U = 0$, and the leader comes to a halt. When this happens, the current stops increasing, and the resistance of the gap decreases. However, the continuing decrease of U_c now causes a decrease of the current and, consequently, of the power contributed to the leader channel. But, since the steady state of the plasma in a broad sense is ensured by the fact that the contributed power equals the power loss, the plasma begins to decay from the head, where the losses are greater. This reduces the length of the leader and, by increasing the resistance of the gap, decreases the current still further. Thus, as the leader evolves, the connection between its length and the current initially ensures their mutual increase, and then their mutual decrease. The time at which these phases change is determined by the value of U_5, which depends on U_0. Therefore, the outwardly paradoxical situation arises in which, when $U_0 = 6$ kV, the leader comes to a halt if $U_c \simeq 5.5$kV, whereas it is formed and reaches a length of 1 cm when $U_0 = 3$ kV.

The difference in the evolution of the leader over the slit and on free water is associated with the side branches in the latter case. By increasing the current in the channel, they increase its conductivity, decreasing U_3 and the rate of decrease of ΔU_0. The initial falloff of the velocity of the leader is therefore less on free water. However, the large discharge current in this case (Fig. 2) accelerates the falloff of U_c, and the extremal values of the length of the leader, the current, and the gap resistance are attained faster but, as before, simultaneously, and the large mean velocity ensures that the maximum length of the leader is the same in both cases. Such an interconnection of the parameters of the leader in time is evidence that its evolution is self-consistent.

In experiments with a storage capacitance of $C = 1\,\mu$F, a water layer 0.8 cm thick, and the same $U_0 = 6$ kV, the falloff of U_c occurs more slowly than with a capacitance of 0.1 μF, and this causes the mean velocity of the head, the length of the leader, and the current amplitude to be large. This case, taking into account what was explained above, makes it possible to assume that the product $C \cdot \Delta U_0$ is an invariant of the spatial evolution of the leader.

The main purpose of our experiments was to understand the dynamics of the evolution of the leader, but they also make it possible to determine the electrical parameters of the SDWS leader channel. These parameters not only supplement the picture of the evolution of the leader in this case, but are also significant in themselves, since it is virtually impossible under other conditions to determine them experimentally over the entire length of the leader. When an analysis is made of the distribution of γ and σ along the length of the channel at the instant that the leader stops (Figs. 6 and 10), it is necessary to take into account that the measurement of the current and the channel diameter is less accurate in the region of the head because their values are small. The distribution of $n_e = \sigma/e \cdot \mu$ (e is the charge and μ the mobility of the electrons) (Fig. 10) is determined with a value of $\mu = 1000\text{cm}^2/\text{V} \cdot \text{sec}$, most likely corresponding to the field in the region of the head.[3] Therefore, as the field close to the anode weakens in the course of the evolution of the leader, the value of n_e there can be even less, since $\mu \sim E^{-1/2}$.[4,5]

The reduction of n_e and σ in the old sections of the channel is associated with its expansion, which has a diffusion character, and the decrease with time of the power contributed to it. These tendencies of the plasma parameters to vary along the leader channel in our case coincide with the concepts in the literature concerning the value and distribution of these parameters in the channel of a long leader in the atmosphere.[3]

This paper has not discussed the state of the plasma in the leader channel. Although its parameters, determined from the experimental data and taking into account the comparatively slow kinetics of the evolution of the leader, $\simeq 10^{-5}$ sec, apparently make it possible to speak of an equilibrium state of a plasma with a temperature of 4000–6000 K,[5,6] the small diameter of the channel, the presence in it of not only longitudinal but also transverse field and current components, and the generation of plasma along the entire length of the generator of the channel most likely indicate that the plasma in the channel and especially in the region of the head is not in equilibrium. Therefore, additional spectroscopic information is needed to solve this problem.

The above treatment makes it possible to draw the following conclusions:

1. SDWS formation at a medium field in the interval 0.3–1 kV/cm is associated with the evolution of the leader.

2. The evolution of the leader is determined by the potential difference between its head and the water surface and has a self-consistent character. The product of the storage capacitance and the initial potential difference between the head of the leader and the water surface is an invariant of the spatial evolution of the leader.

3. Plasma formation at the head of the leader is associated with the presence in it of an electric-field component normal to the water surface, having high field strength because of the small diameter of the head.

4. The motion of the head of the leader is determined both by the longitudinal field component in front of the head and by the normal field component at the boundary of the head.

[1] V. P. Belosheev, ''Device for decontaminating and purifying drinking and waste waters'' [in Russian], Russian Patent No. 2042641 on 5/14/92.
[2] V. P. Belosheev, Zh. Tekh. Fiz. **66**(8), 50 (1996) [Tech. Phys. **41**, 773 (1996)].
[3] E. M. Bazelyan and I. M. Rozhanskiĭ, *The Spark Discharge in Air* [in Russian] (Nauka, Novosibirsk, 1988).
[4] Yu. P. Raĭzer, *Physics of the Gas Discharge* [in Russian] (Nauka, Moscow, 1987), p. 366.
[5] V. L. Granovskiĭ, *The Electron Current in Gas* [in Russian] (Nauka, Moscow, 1971), p. 102.
[6] R. H. Huddlestone and S. L. Leonard (eds.), *Plasma Diagnostic Techniques* [Academic Press, New York, 1965; Mir, Moscow, 1967, p. 169.]

Translated by W. J. Manthey

Prominent characteristics of the decay of a photoplasma generated by radiation from an annular sliding discharge

N. A. Popov

Computing Center, Russian Academy of Sciences, 117967 Moscow, Russia
(Submitted April 14, 1997)
Zh. Tekh. Fiz. **68**, 51–55 (July 1998)

The dynamics of the absorption of diagnostic microwave radiation in a decaying nitrogen photoplasma generated by a pulsed annular sliding discharge is modeled numerically. It is shown that the microwave absorption coefficient can vary nonmonotonically with time (as observed in the majority of experimental studies) during recombinative plasma decay. The nonmonotonic behavior is attributed to gasdynamic processes induced by the annular sliding discharge, which cause the plasma region to expand along the axis of the sensing microwave beam. © *1998 American Institute of Physics.* [S1063-7842(98)00907-6]

1. Major advances in the production and investigation of externally sustained discharges in molecular gases have stimulated searches for new ways to manage these discharges and for new gas preionization mechanisms. One recently developed technique is to use ultraviolet radiation from the plasma of a discharge sliding along the surface of a dielectric.[1–4] This approach can result in a high density of photoelectrons and the formation of a space charge at relatively low electric field strengths.[1,2]

Two types of discharge are currently under investigation, one on the surface of a flat dielectric[1,2] (forming so-called plasma sheets) and the other on the surface of a dielectric ring (annular discharge).[3–7] In either case a dense plasma is formed at distances of the order of several centimeters as a result of high-intensity UV radiation. For example, the density of photoelectrons in nitrogen at atmospheric pressure for both annular and planar discharges has attained $N_e = 6 \times 10^{12}$ cm^{-3} (Refs. 2–4).

Significant differences from the recombination law have not been established in the investigation of the decay of the resulting plasma in the planar case, whereas the photoplasma created from an annular discharge has been observed to have anomalously long lifetimes tens of times greater than the corresponding electron-ion recombination times.[5] The discharges have been found to have this character not only in nitrogen, but also in argon, helium, air, CO_2, and mixtures of these gases.[5–7]

The procedure used to investigate the decay of a plasma produced by an annular sliding discharge entails the determination of the absolute value and temporal dynamics of the absorption coefficient of diagnostic microwave radiation. This coefficient is equal to the ratio of the transmitted to the incident microwave power and is defined as[8]

$$F = \exp\left(-4\pi/c \int \sigma(y)\, dy \right), \tag{1}$$

where $\sigma(y)$ is the plasma conductivity averaged over the cross section of the microwave beam:

$$\sigma = e^2 N_e \nu_m / m(\nu_m^2 + \omega^2), \tag{2}$$

ν_m is the electron collision frequency, ω is the frequency of the sensing microwave radiation, and c is the speed of light.

The integration is carried out along the direction of electromagnetic wave propagation (along the y axis) within the boundaries of the plasma region.

Proceeding from Eqs. (1) and (2), we can write the following equation for the average degree of ionization along the y axis:

$$N_e^{cp}/N = \ln(1/F)\, \alpha_0 \Delta Y, \tag{3}$$

where $\alpha_0 = mc/4\pi e^2 (\nu/N) \simeq 1.4 \times 10^{-6}$ cm, and ΔY is the length of the plasma region.

The previously observed[5–7] dynamics of the microwave absorption coefficient during the plasma decay stage exhibits nonmonotonic behavior. This fact has led Gritsinin *et al.*[5,9] to the conclusion that secondary ionization reactions effectively take place in the postdischarge period.

Anomalous phenomena have also been recorded in the investigation of plasma decay in the channel of a laser spark in nitrogen and in air. Under the conditions reported in Refs. 10 and 11, for example, the decay times of the electron density of the plasma formed in the laser breakdown of air exceeded by a factor of tens the time corresponding to the recombination decay regime. The density of the generated plasma in this case was determined from the microwave absorption coefficient, whose temporal dynamics was also nonmonotonic.

The objective of the present study is to analyze the specific mechanisms underlying the formation of the photoplasma of an annular sliding charge and to describe the prominant features of the dynamics of the microwave absorption coefficient during the decay of the generated plasma.

2. At a gas pressure $P > 10$ Torr the emergence of a plasma inside the ring is associated with the photoionization of molecules of the mixture by hard UV radiation from a sliding discharge.[3–7] The presence of even a minute impurity of oxygen molecules (which have a relatively low ionization potential) can lead to photoionization because radiation hav-

ing a wavelength smaller than 102 nm is absorbed by the oxygen. In nitrogen-oxygen mixtures the source of ionization radiation is found in molecular N_2 bands in the wavelength interval $98\ \text{nm}<\lambda<102.5$ nm (Ref. 12).

The spectral composition of sliding-charge plasma radiation has been investigated mainly in the range $\lambda>100$ nm (Ref. 13). It has been shown that the radiation spectrum has a complex line structure with a predominance of strong lines from constituent atoms of the dielectric material. Berezhetskaya *et al.*,[3] have performed probe measurements and used to determine the absorption coefficient of ionizing radiation from an annular sliding discharge. For this coefficient in nitrogen at atmospheric pressure they obtained $\chi=0.6-0.8\ \text{cm}^{-1}$.

Ionization in pure nitrogen is associated with the absorption of radiation in the wavelength range $\lambda<80$ nm. In this part of the spectrum the coefficient χ at atmospheric pressure has the value $\chi\simeq 700\ \text{cm}^{-1}$, which is considerably higher than the experimental value. Pravilov[14] has proposed a mechanism involving the photoionization of metastable nitrogen molecules, which, in turn, are formed in the absorption of photons by unexcited $N_2(X^1\Sigma_g)$ molecules. In nitrogen the most intense generation is exhibited by $N_2(a^1\Pi_g)$ metastables during the absorption of radiation in Lyman–Birge–Hopfield bands. However, the radiation absorption coefficient in these bands at $P=1$ atm is not higher than $0.11\ \text{cm}^{-1}$ (Ref. 15), which is far below the experimentally measured value. Moreover, it has been remarked[3] that the coefficient χ is essentially independent of the nitrogen pressure, so that the photoionization observed in Ref. 3 is most likely attributable to the presence of impurities (e.g., an oxygen impurity). Similar conclusion are drawn in Refs. 13 and 16, based on investigations of the mechanism of photoionization of nitrogen by radiation from sliding and spark discharges.

At atmospheric pressure and with a $\simeq 0.1\%$ fraction of oxygen in the mixture the absorption coefficient is approximately $\chi\simeq 0.6\ \text{cm}^{-1}$ (Ref. 12), which is consistent with the measurements results.[3]

The primary ions formed in the absorption of UV radiation by oxygen molecules are O_2^+. They can eventually convert into more complex O_4^+ and $O_2^+N_2$ ions in the reactions[17]

$$\begin{aligned} &O_2^+ + O_2 + O_2 \rightarrow O_4^+ + O_2, \quad 2.4\times 10^{-30}(300/T)^{3.2}\ \text{cm}^6/\text{s},\\ &O_2^+ + N_2 + N_2 \rightarrow O_2^+N_2 + N_2, \quad 8\times 10^{-31}(300/T)^2\ \text{cm}^6/\text{s}, \end{aligned} \tag{4}$$

$$O_2^+N_2 + N_2 \rightarrow O_2^+ + 2N_2,$$

$$10^{-6}(300/T)^5\exp(-2357/T)\ \text{cm}^3/\text{s}, \tag{5}$$

$$O_2^+N_2 + O_2 \rightarrow O_4^+ + O_2, \quad 10^{-9}\ \text{cm}^3/\text{s}. \tag{6}$$

The rates for the dissociative recombination reactions of O_2^+ and O_4^+ ions

$$O_2^+ + e \rightarrow \text{product}, \quad 2\times 10^{-7}(300/T_e)^{0.7}\ \text{cm}^3/\text{s}, \tag{7}$$

$$O_4^+ + e \rightarrow \text{product}, \quad 2\times 10^{-6}(300/T_e)^{0.5}\ \text{cm}^3/\text{s} \tag{8}$$

differ more than tenfold,[12,17] underscoring the importance of the question as to which ion species is predominant.

We must also take into account the additional electron annihilation channels associated with attachment to O_2 molecules:

$$e + O_2 + O_2 \rightarrow O_2^- + O_2, \quad 1.9\times 10^{-30}\ \text{cm}^6/\text{s},$$

$$e + O_2 + N_2 \rightarrow O_2^- + N_2, \quad 8.5\times 10^{-32}\ \text{cm}^6/\text{s}.$$

At $P=200$ Torr in a $N_2+0.1\%O_2$ mixture (under conditions corresponding those in Ref. 5) attachment processes becomes decisive at $t>\tau_\alpha=300\ \mu$s.

Proceeding from Eqs. (4)–(8), we estimate the ratio of the main species of ions for the $N_2+0.1\%O_2$ mixture at $P=200$ Torr. At times $t\geqslant 5$ ns the concentration $[O_2^+N_2]$ is

$$[O_2^+N_2]=[O_2^+][N_2]k_4/k_5\simeq 0.01[O_2^+],$$

where k_1 is the rate constant of the ith process. With this fact in mind, at $t\gg(k_6[O_2])^{-1}\simeq 1\ \mu$s we obtain

$$[O_4^+]/[O_2^+]=[O_2][N_2]k_4k_6/(N_ek_5k_8). \tag{9}$$

If only two main ion species are present, the dissociative recombination coefficient for a quasineutral plasma ($N_e\simeq[O_4^+]+[O_2^+]$) can be written in the form

$$\beta=(k_7+k_8[O_4^+]/[O_2^+])/(1+[O_4^+]/[O_2^+]). \tag{10}$$

It is evident from Eqs. (9) and (10) that the ratio $[O_4^+]/[O_2^+]$ and the quantity β increase with increasing pressure and, according to (4)–(6), decrease as the temperature of the gas increases. Under the experimental conditions reported in Ref. 5 ($T=350$ K, $P=200$ Torr, and $N_e\simeq 2\times 10^{11}\text{cm}^{-3}$) we obtain $[O_4^+]/[O_2^+]\simeq 0.1$ and $\beta\simeq 3.5\times 10^{-7}\text{cm}^3/\text{s}$.

At a constant recombination rate β the dynamics of the electron density in the decay stage is described by the expression

$$N_e(t)=N_e^0/(1+N_e^0\beta(t-t_0)). \tag{11}$$

Under the conditions of Ref. 5 at $\beta=3.5\times 10^{-7}\text{cm}^3/\text{s}$ the characteristic plasma decay time should not exceed 15 μs. On the other hand, measurements[5–7] give values 20–30 times higher.

Gritsinin *et al.*[5,9] attribute the increase in the characteristic decay times of the generated photoplasma to the onset of associative ionization processes involving metastable electron-excited atoms and molecules. The possible candidates for these molecules in nitrogen are $N_2(A^3\Sigma_u)$ and $N_2(a^1\Pi_g)$. Their interaction can lead to the formation of charged particles in the reactions[17]

$$N_2(A^3\Sigma_u) + N_2(a^1\Pi_g) \rightarrow N_4^+ + e, \tag{12}$$

$$N_2(a^1\Pi_g) + N_2(a^1\Pi_g) \rightarrow N_4^+ + e. \tag{13}$$

However, investigations[6,7] have shown that the presence of up to 0.5% oxygen molecules in nitrogen scarcely affects the way in which the plasma decays. Under the experimental conditions in Refs. 6 and 7 at $P=100$ Torr the concentration of oxygen molecules attained $[O_2]=(1-2)\times 10^{16}\text{cm}^{-3}$, resulting in deactivation of the $N_2(A^3\Sigma_u)$ and $N_2(a^1\Pi_g)$ states (in times of 30 μs and 1 μs, respectively). It is important to note that in other gases metastable states capable of contributing to associative ionization reactions [e.g., $Ar(^3P_0)$,

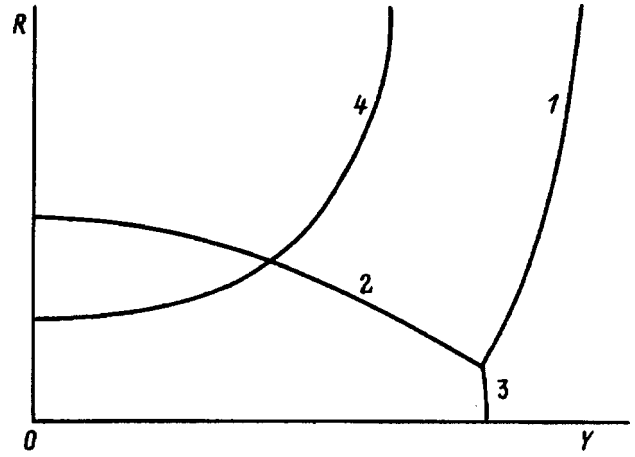


FIG. 1. Positions and configurations of the principal discontinuities in the reflection of a toroidal shock wave from the symmetry (*Y*) axis. *1*) Incident shock; *2*) reflected shock; *3*) Mach wave; *4*) second shock.

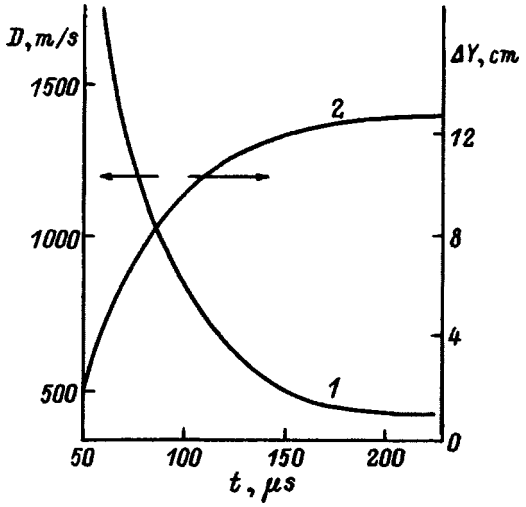


FIG. 2. Mach wave velocity[19] (curve *1*) and length of the plasma region (curve *2*) along the *Y* axis versus time.

$\mathrm{He}(^3S_1)$, $\mathrm{He}(^3S_0)$, etc.] are effectively quenched by molecular oxygen. Consequently, the allowance for reactions involving these atoms and molecules does not offer any explanation for the experimentally observed rise in the density of electrons within times $t \geqslant 70-100\ \mu$s.

3. The influence of the above-described UV radiation sources utilizing an annular sliding discharge is accompanied by rapid heat release and active gasdynamic processes. The papers reviewed by us[3–7,18–20] report investigations of annular discharges of the same structure (described in detail in Refs. 3 and 4). The ring had a radius of 5 cm, the pulse duration was $10-20\ \mu$s, and the applied voltage was $U=14-21$ kV. The width of the diagnostic microwave beam was $d \simeq 2\lambda$, where λ is the wavelength of the microwave radiation [$\lambda=2$ cm (Ref. 5) and $\lambda=0.8$ cm (Refs. 3, 4, 6, and 7)].

Several studies of gasdynamic processes initiated by an annular discharge[18–23] have been concerned primarily with the dynamics of shock wave formation and cumulative effects accompanying the reflection of this shock wave from the axis of the ring. Figure 1 shows a typical example[21–23] of the relative position of the main discontinuities of the gasdynamic parameters after the reflection of a toroidal shock wave from the symmetry axis (*Y* axis). The axis *OR* lies in the plane of the ring.

The acceleration of a convergent annular shock wave decreases the local values of the angle of inclination of the shock front relative to the symmetry axis, so that the reflection of the shock front from the axis is an irregular process with the formation of a Mach shock wave propagating along the *Y* axis (curve *3* in Fig. 1). It must be emphasized that a large fraction of the energy of compression in the reflection of an annular shock wave is imparted to the gas expanding along the axis of the ring in the direction of least counterpressure.[21] This phenomenon results in the formation of a strong Mach wave.

Gasdynamic processes can significantly influence the dynamics of the plasma region formed as a result of photoionization processes. The expansion of this region in the direction of the *Y* axis takes place because the gas flow behind the shock wave (curve *1* in Fig. 1) transports the plasma from the periphery of the ring into the axial zone. As a consequence, the electron density increases at large distances from the plane of the ring, an effect than can be interpreted as expansion of the plasma region along the *Y* axis. Inasmuch as the degree of ionization of the gas changes only very slightly at the shock front, it can be assumed with reasonable accuracy that the distribution of N_e/N along the *Y* axis is uniform within the boundaries of the plasma region.[1)] The variation of the electron density in this case is mainly attributable to electron-ion recombination processes.

The velocity of the boundary of the plasma region depends on the velocity of the gas flow behind the front of the resulting Mach wave (curve *3* in Fig. 1). According to calculations,[21,22] this velocity depends linearly on the coordinate *Y*. Consequently, the instantaneous dimensions of the plasma region can be defined as

$$\Delta Y(t)=\Delta Y_0 \exp \int_0^t V_f(\tau)/Y_s(\tau)\,d\tau. \tag{14}$$

Here $Y_s(t)=\int_0^t D(\tau)\,d\tau$ is the distance traversed by the shock wave in the time t, ΔY_0 is the initial length of the plasma region, and $V_f(t)$ is the velocity of the gas immediately behind the shock front[24]:

$$V_f(t)=\frac{D(t)-C_s^2/D(t)}{\gamma+1}, \tag{15}$$

$D(t)$ is the velocity of the shock wave, C_s is the sound velocity in the undisturbed gas, γ is the adiabatic exponent, and $\Delta Y_0 \simeq 2.5$ cm (Ref. 3).

Bedin *et al.*[19] (working under conditions similar to those in Refs. 18 and 20) have measured the dynamics of the Mach wave velocity $D(t)$, whereupon they were able to use equations from Refs. 14 and 15 to determine the instantaneous dimensions of the plasma region $\Delta Y(t)$.

Figure 2 shows experimental data for the Mach wave velocity[19] and the results of calculations of the dynamics of expansion of the plasma region $\Delta Y(t)$. It is evident that the

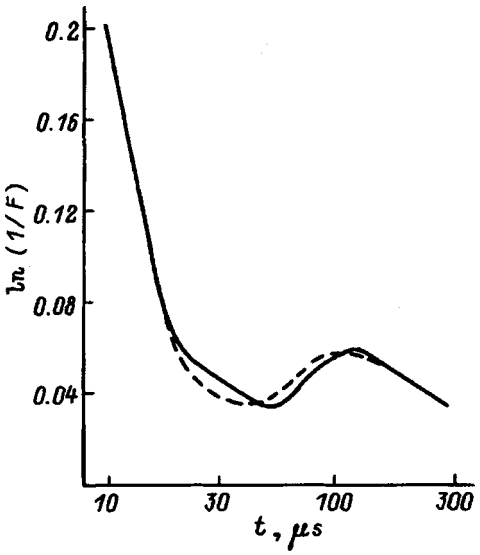


FIG. 3. Dynamics of the absorption coefficient of diagnostic microwave radiation in an annular discharge plasma under the conditions of Ref. 5: nitrogen, $P=225$ Torr, $\lambda=2$ cm. The solid curve gives the results of calculations, and the dashed line represents experimental data.[5]

dimensions of the plasmoid increase more than fivefold within $140-150\ \mu$s, exerting an appreciable influence on the microwave transmission coefficient.

The following experiments have been carried out[7] to determine the longitudinal (along the Y axis) dimensions of the plasma region. Radio-transparent plates were mounted in the working chamber, parallel to the plane of the ring and at equal distances from the ring, to limit the size of the plasma. The distance L between the plates was varied from 4 cm to 20 cm (for a ring of width 1 cm). It was shown that for $L=10-12$ cm the presence of the plates did not affect the dynamics of the microwave absorption coefficient, i.e., the dimensions of the plasma region did not exceed 10–20 cm. As the spacing L was decreased, the duration of anomalous absorption of radiation was observed to decrease considerably until the effect all but totally disappeared at $L\simeq 4$ cm. In this case the behavior of the absorption coefficient corresponded to recombination-type decay of the sensing plasma. According to these data, the maximum length of the plasma region is approximately 10–12 cm, consistent with the results of our calculations in Fig. 2.

To explain the experimentally observed anomalies of the dynamics of the microwave absorption coefficient dynamics $F(t)$, we have calculated the evolution of the parameter

$$\Theta(t)=\alpha_0(N_e^{\mathrm{mid}}/N)\Delta Y(t),$$

which, according to Eq. (3), governs the quantity $\ln[1/F(t)]$. To determine $N_3^{\mathrm{mid}}(t)$, we have solved the system of balance equations for the densities of electrons and the principal species of positive and negative ions: N_4^+, O_2^+, O_4^+, $O_2^+\cdot N_2$, and O_2^-. The system of ion-molecular reactions[17] was adopted as the basis. The calculations were carried out for a $N_2+0.1\%O_2$ mixture at a pressure $P=225$ Torr. Equations (14) and (15) were used to calculate $\Delta Y(t)$.

The results of the calculations of $\Theta(t)$ and experimental data[5] are shown in Fig. 3. It is evident that the proposed model can be used to account for the nonmonotonic dynamics of the absorption coefficient of the microwave signal, attributing it to the expansion of the plasma region as a result of gasdynamic processes. It should be emphasized that the phenomenon described here is a consequence of the annular shape of the source of energy release. As mentioned, this source has the effect of driving the plasma from the peripheral regions of the ring toward its axis, into the microwave sensing region, thereby accounting for the variation in the dynamics of the microwave absorption coefficient.

It follows from this model, in particular, that the amplitude of the indicated nonmonotonicity of the absorption coefficient should increase when the energy input during the discharge stage increases (causing the intensity of the shock waves and the rate of expansion of the plasma region to increase). This effect has been observed previously[7] in nitrogen and in argon with oxygen additives when the voltage was increased from 14 kV to 21 kV. Also noted in Ref. 7 is an increase in the lifetime of the photoplasma of an annular discharge when the gas pressure is lowered (the lifetime τ was determined as the time at which the absorption coefficient of the sensing radiation attained a prescribed value). According to Ref. 7, $\tau\cdot P\simeq$ const at $P\geqslant 50$ Torr. The microwave absorption coefficient F for $\nu_m\gg\omega$ and a fixed value of ΔY depends only on the ratio N_e/N [see Eq. (3)]. Consequently, when the pressure is reduced, lower densities N_e are required in order to attain a given value of F. If we assume that the plasma decays according to the recombination law, the characteristic decay time is $\tau=(N_e\beta)^{-1}=\Delta Y/[N\beta\alpha_0\ln(1/F)]$. According to Ref. 7, ΔY changes only slightly at pressures $P=70-760$ Torr, so that $\tau N\simeq$ const at a given value of the coefficient F, consistent with the data in Ref. 7. This result affords further confirmation of the recombination character of the decay of the investigated plasma.

5. This investigation can be summarized in the following conclusions.

The decay of the plasma generated by hard ultraviolet radiation from an annular sliding discharge obeys the recombination law. The experimentally observed monotonicities of the microwave absorption coefficient (which are treated as resulting from secondary ionization reactions and are counted among the important advantages of this preionization technique[5–7]) can be attributed to expansion of the plasma region due to intense gasdynamic processes. This phenomenon is a consequence of the annular shape of the energy-release source and is not observed in a planar geometry.

Gasdynamic processes are also responsible for the formation of nonuniform distributions of the temperature and gas density in the plane of the ring. Zones of hot, rarefied gas are observed near the surface of the ring and particularly in the zone around the axis; these zones are clearly distinguishable on shadowgrams.[9,18] The application of an annular sliding discharge for preionization of the gas in the live zone of an externally sustained discharge can lead to breakdown of the gas in heated regions (owing to an increase in the normalized electric field E/N and a rise in the electron temperature). This phenomenon has been observed[9] in a study of an

externally sustained microwave discharge in nitrogen.

Gasdynamic processes thus have a significant and usually deleterious influence on the properties of an annular discharge as a device for the preionization of gas mixtures. This influence can be avoided by placing the emitting ring outside the gas-discharge chamber in which the externally sustained discharge is ignited. By maintaining the axial symmetry, this should also preserve such an important advantage of annular discharges as the uniform distribution of the intensity of the ionizing radiation flux in the plane of the ring. Moreover, this arrangement makes it possible to ignite an externally sustained discharge and an annular sliding discharge in different gas mixtures, thereby creating further opportunities for optimizing the parameters of the overall system.

[1)]It is possible for N_e to decrease by virtue of the more rapid recombinative decay of the plasma in the postshock compression zone, where N_e is higher than the corresponding background values. However, the compression zones of explosive shock waves have relatively small dimensions. Under the stated conditions the residence time of the gas in the zone where the particle concentration is two or more times the background level is $\tau_0 \simeq 2-3\ \mu$s. In times $t \gg \tau_0$ the background value of N_e is $N_e^f \simeq (\beta \cdot t)^{-1}$ and, hence, the variation of the ratio N_e/N (due to the recombination of electrons) as the gas passes through the postshock compression zone is relatively small.

[1] S. I. Andreev, I. M. Belousova, P. N. Dashuk *et al.*, Kvantovaya Élektron. (Moscow) **3**, 1721 (1976) [Sov. J. Quantum Electron. **6**, 931 (1976)].
[2] Yu. I. Bychkov, D. Yu. Zaroslov, N. V. Karlov *et al.*, Kvantovaya Elektron. (Moscow) **9**, 1718 (1982) [Sov. J. Quantum Electron. **12**, 1105 (1982)].
[3] N. K. Berezhetskaya, E. F. Bol'shakov, A. A. Dorofeyuk *et al.*, FIAN Preprint No. 224 [in Russian] (P. N. Lebedev Physics Institute, Russian Academy of Sciences, Moscow, 1983), 28 pp.
[4] É. M. Barkhudarov, N. K. Berezhetskaya, E. F. Bol'shakov *et al.*, Zh. Tekh. Fiz. **54**, 1219 (1984) [Sov. Phys. Tech. Phys. **29**, 701 (1984)].
[5] S. I. Gritsinin, I. A. Kossyĭ, V. P. Silakov *et al.*, Teplofiz. Vys. Temp. **24**, 662 (1986).
[6] L. S. Bogdan, Yu. V. Zadiraka, S. M. Levitskiĭ *et al.*, Pis'ma Zh. Tekh. Fiz. **15**(9), 6 (1989) [Sov. Tech. Phys. Lett. **15**, 331 (1989)].
[7] L. S. Bogdan, *Candidate's Dissertation* [in Russian] (Kiev, 1990).
[8] V. E. Golant, *Microwave Methods in Plasma Research* [in Russian] (Nauka, Moscow, 1968), 327 pp.
[9] S. I. Gritsinin, I. A. Kossyĭ, V. P. Silakov *et al.*, Zh. Tekh. Fiz. **57**, 681 (1987) [Sov. Phys. Tech. Phys. **32**, 414 (1987)].
[10] G. A. Askar'yan, M. S. Rabinovich, M. M. Savchenko, and A. D. Smirnova, JETP Lett. **1**, 162 (1965).
[11] G. A. Askar'yan, M. S. Rabinovich, M. M. Savchenko, and V. K. Stepanov, JETP Lett. **3**, 303 (1966).
[12] A. Kh. Mnatsakanyan and G. V. Naĭdis, in *Plasma Chemistry*, No. 14 (Énergoatomizdat, Moscow, 1987), pp. 227–255.
[13] V. P. Silakov, Zh. Tekh. Fiz. **57**, 361 (1987) [Sov. Phys. Tech. Phys. **32**, 217 (1987)].
[14] A. M. Pravilov, *Photoprocesses in Molecular Gases* [in Russian] (Énergoatomizdat, Moscow, 1992).
[15] D. Yu. Zaroslov, N. V. Karlov, G. P. Kuz'min, and S. M. Nikiforov, Kvantovaya Élektron. (Moscow) **5**, 1221 (1978) [Sov. J. Quantum Electron. **8**, 695 (1978)].
[16] R. V. Babcock, J. Liberman, and W. D. Parthow, Plasma Sources Sci. Technol. **QE-12**, 29 (1976).
[17] I. A. Kossyi, A. Yu. Kostinsky, A. A. Matveyev, and V. P. Silakov, Plasma Sources Sci. Technol. **1**, 207 (1992).
[18] N. K. Berezhetskaya, E. F. Bol'shakov, S. K. Golubev *et al.*, Zh. Éksp. Teor. Fiz. **87**, 1926 (1984) [Sov. Phys. JETP **60**, 1108 (1984)].
[19] A. P. Bedin, A. B. Safonov, and M. N. Troitskiĭ, Zh. Tekh. Fiz. **60**(12), 142 (1990) [Sov. Phys. Tech. Phys. **35**, 1447 (1990)].
[20] É. M. Barkhudarov, M. O. Mdivnishvili, I. V. Sokolov *et al.*, JETP Lett. **52**, 379 (1990).
[21] V. M. Khudyakov, Dokl. Akad. Nauk SSSR **287**, 802 (1986) [Sov. Phys. Dokl. **31**, 298 (1986)].
[22] A. V. Baranov, A. V. Vasil'ev, and V. V. Krasnobaev, Mat. Model. **4**(12), 3 (1992).
[23] P. A. Voĭnovich, E. L. Satunina, and E. V. Timofeev, Zh. Tekh. Fiz. **66**(2), 12 (1996) [Tech. Phys. **41**, 117 (1996)].
[24] V. P. Korobeĭnikov, *Theory of Point Detonation* [in Russian] (Nauka, Moscow, 1985).

Translated by James S. Wood

Conditions for the existence of a positively charged structure in a Penning discharge

S. P. Nikulin

Institute of Electrophysics, Ural Branch of the Russian Academy of Sciences, 620049 Ekaterinburg, Russia
(Submitted April 14, 1997)
Zh. Tekh. Fiz. **68**, 56–63 (July 1998)

An analytical model is developed for the high-current form of a low-pressure glow discharge in a magnetic field. Expressions are derived for the critical magnetic induction and critical pressure, below which it becomes impossible for this form of discharge to exist. It is shown that the transition from the high-voltage form to the high-current form of discharge with increasing pressure is not attributable to an increase in the ionization rate, but to an increase in the drift velocity of plasma electrons across the magnetic field. Estimates based on the expressions derived in the article agree in order of magnitude with the experimental data. It is shown that the region in which discharge exists can change considerably in the presence of electron emission. © *1998 American Institute of Physics.* [S1063-7842(98)01007-1]

The burning of a glow discharge with oscillating electrons (Penning discharge) can take place in two forms: 1) high-voltage discharge characterized by the predominance of a negative charge in the gap; 2) high-current discharge, where essentially the entire gap is filled with a plasma having a low potential drop, and the discharge voltage is localized almost entirely in the ionic cathode sheath.[1] In the second case the structure formed in the discharge has a positive charge on the whole. Zharinov and Nikonov[2] have investigated a discharge in a magnetic field, noting the potential benefit of using discharges with a positively charged structure for the design of efficient gas-discharge devices, electron sources in particular. However, the analysis in Ref. 2 treats a situation where the unneutralized positive charge, which is proportional to the difference between the densities of ions and electrons, is not localized in the cathode sheath, but is distributed uniformly in the discharge gap. It is also assumed in the paper that ionization in the discharge is implemented by plasma electrons, whereas experimental results[3] have shown that in glow discharges with electron oscillations, in addition to the group of slow plasma electrons having an almost-Maxwellian distribution, there is also a group of so-called fast particles. These particles are the result of γ-ray processes on the cathode, and when they are accelerated in the cathode sheath, they acquire an energy that corresponds to the cathode drop and is then spent in elastic and inelastic collisions with neutral gas atoms. The fraction of these particles is not very large, but they provide the main contribution to ionization, while the contribution of secondary plasma electrons resulting from ionization by fast particles is not significant, because they do not acquire sufficient energy from the weak electric field present in the plasma. The separation of the gas-discharge gap into the cathode sheath and the plasma region, together with allowance for ionization by fast particles, has made it possible to calculate the characteristics of a glow discharge with electron oscillations in a hollow cathode in satisfactory agreement with the experimental results.[4] The objective of the present study is to develop a simplified analytical model of the high-current form of a discharge with electron oscillations and to estimate the conditions under which this form occurs. We are particularly interested in determining the lower limit of the working pressure range, because one of the main applications of the investigated discharge is the development of charged-particle sources utilizing it, where low pressure is necessary to ensure the electric strength of the accelerating gap. Moreover, the influence of electron emission on the discharge characteristics is investigated within the framework of the model developed here.

MODEL OF DISCHARGE IN A MAGNETIC FIELD

We consider the problem in planar geometry. We assume that the anode of the gas discharge is situated in the plane $x=0$, and the cathode is situated in the plane $x=d$. In the gap, which is filled with gas to a pressure p, there is a magnetic field with induction B perpendicular to the electric field. The influence of the magnetic field on the motion of ions can be disregarded, and it can be assumed that at low pressures in the Coulomb regime ions leave the gap without collisions. As for electrons, they are magnetized and, because of their oscillations in the magnetic field, traverse a path considerably longer than the gap; hence, their collisions must be taken into account, even though the conditions of the Coulomb regime are formally satisfied.

Physical justification for the separation of electrons into two groups can be found in the observation that the transport cross section of Coulomb interaction drops abruptly as the energy increases. Calculations based on an approximate equation in Ref. 5 give this cross section as 4×10^{-14} cm^2 for electrons with a thermal energy of 5 eV, which is characteristic of plasma electrons in the investigated discharges, and 4×10^{-18} cm^2 for electrons with an energy of 500 eV, which is characteristic of fast particles. On the other hand, the cross sections of the different types of interaction between electrons and atoms for the majority of gases is of the order of magnitude of 10^{-16} cm^2. It is evident from these numbers that slow particles interact fairly vigorously, prob-

ably accounting for the nearly Maxwellian distribution in the low-energy range, whereas the main interaction for fast particles, even when the gas is highly ionized, is interaction with neutral particles.

As a fast particle slows down in the gas, its energy is spent in the excitation and ionization of atoms and also in elastic collisions. Over a wide range of initial energies in this case the formation of one pair of charged particles requires, on the average, the expenditure of a certain energy W, which is constant for each type of gas. Also allowing for the fact that the ionization cross section in the characteristic energy range for fast particles of a glow discharge, 100–1000 eV, can be very accurately approximated by a dependence of the type $1/v$, where v is the particle velocity, and that the ionization frequency ν_i can therefore be regarded as a constant, we write the following relation for the characteristic fast-particle relaxation time:

$$\tau_r=\frac{eU_c}{\nu_i W}, \tag{1}$$

where e is the electron charge, U_c is the cathode drop, which is essentially equal to the discharge voltage, and eU_c is the energy acquired by a fast particle after traversing the cathode sheath.

During the time τ_r a fast particle loses its ability to ionize and transfers to a group of slow particles. For fast particles we can then write the equation of continuity in the form

$$\frac{d(n_f v_f)}{dx}=\frac{n_f}{\tau_r}, \tag{2}$$

where n_f is the density of fast particles, and v_f is the average velocity of their directional motion across the magnetic field.

There is no minus sign on the right side of Eq. (2), because the cathode-to-anode direction is adopted as the positive direction for the fluxes of both fast and slow electrons. The fast-particle flux is governed by diffusion across the magnetic field:

$$n_f v_f=D_f\frac{dn_f}{dx}, \tag{3}$$

where D_f is the fast-particle diffusion coefficient, and the drift component can be disregarded, because the weak electric field in the plasma does not have any significant influence on the motion of fast particles.

To verify the latter assertion, we estimate the ratio of the characteristic length l traversed by a fast electron as a result of drift to the characteristic diffusion length $l_D=\sqrt{D_f\tau_r}$. The following relation can be written for the fast-electron mobility coefficient:

$$\mu_f=\frac{e\nu_f}{m\omega^2}, \tag{4}$$

where ν_f is the effective collision frequency for fast electrons, and m and ω are the mass and Larmor frequency of the electron.

The relation $\nu_f \ll \omega$ for magnetized electrons is taken into account in Eq. (4). Also allowing for the fact that the square of the fast-particle velocity varies from $2eU_c/m$ essentially to zero and maintains a value $\sim eU_c/m$ on the average, we write the expression for the fast-particle diffusion coefficient

$$D_f=\frac{eU_c}{3m}\frac{\nu_f}{\omega^2}. \tag{5}$$

The potential drop across the plasma region of the investigated discharges is of the order of kT_e/e, where k is the Boltzmann constant, and T_e is the plasma electron temperature. Since the thickness of the cathode sheath is generally much smaller than the length of the plasma region, the average electric field in the plasma can be estimated as

$$E=\frac{kT_e}{ed}. \tag{6}$$

Using Eqs. (4)–(6), we obtain

$$\frac{l}{l_D}=\frac{\mu_f E\tau_r}{\sqrt{D_f\tau_r}}\approx\sqrt{\frac{\nu_f}{\nu_i}}\frac{R_L}{d}\frac{kT_e}{\sqrt{eU_c W}}, \tag{7}$$

where R_L is a characteristic Larmor radius of fast electrons.

In Eq. (7) the ratio l/l_D is represented by a product of three factors: The first, $\sqrt{\nu_f/\nu_i}$, is a quantity of the order of unity; the second, $R_L d$, must be smaller than unity for the investigated discharge, otherwise electrons would immediately leave the anode and oscillations could not take place; finally, the third factor, $kT_e/\sqrt{eU_c W}$, is much smaller than unity, because the thermal energy of the plasma electrons is substantially lower than both the fast-particle energy and the available ion energy. On the whole, therefore, the given ratio is much smaller than unity and, hence, the drift component of the fast-particle flux can be ignored in comparison with the diffusion component.

Both the diffusion and the drift component are significant for the flux of plasma electrons:

$$n_e v_e=D_e\frac{dn_e}{dx}-\mu_e n_e\frac{d\varphi}{dx}, \tag{8}$$

where n_e, v_e, D_e, and μ_e are the density, average directional velocity, diffusion coefficient, and mobility coefficient of slow electrons in a transverse magnetic field, and φ is the potential.

We write the equation of continuity for slow particles in the form

$$\frac{d(n_e v_e)}{dx}=-v_e n_f. \tag{9}$$

A term corresponding to the transition of electrons from the group of fast particles to the group of slow particles could be added to the right side of Eq. (9), but it is much smaller than the ionization term. The equation of continuity for ions has a similar form:

$$\frac{d(n_i v_i)}{dx}=v_i n_f, \tag{10}$$

where n_i and v_i are the density and average velocity of the ions, but now the right side has a plus sign, because the anode-to-cathode direction is chosen as the positive direction for the ion flux.

We use a simplified form of the equation of motion for ions,[6] but without the collision term:

$$\frac{d(n_i v_i^2)}{dx} = -\frac{en_i}{M}\frac{d\varphi}{dx}, \tag{11}$$

where M is the ion mass.

Combining Eqs. (2) and (3), we obtain a second-order differential equation for n_f:

$$n_f - l_D^2 \frac{d^2 n_f}{dx^2} = 0. \tag{12}$$

One of the two boundary conditions needed to solve this equation is dictated by cathode processes:

$$n_f(d) v_f(d) = D_f dn_f/dx_{x=d} = \gamma j_c/e, \tag{13}$$

where γ is the effective ion-electron emission coefficient, and j_c is the ion current density at the cathode.

The second condition can be formulated by setting n_f equal to zero at an absorbing wall, i.e., at the anode:

$$n_f(0) = 0. \tag{14}$$

A solution that satisfies both these boundary conditions has the form

$$n_f = \frac{\gamma j_c l_D \sinh(x/l_D)}{eD_f \cosh(d/l_D)}. \tag{15}$$

Assuming that all the ions formed in the gap enter the cathode, we can describe the ion current density at the cathode by the expression

$$j_c = \int_0^d e\nu_i n_f dx = \frac{\gamma j_c \nu_i l_D^2}{D_f \cosh(d/l_D)}(\cosh(d/l_D) - 1). \tag{16}$$

It follows from Eq. (16) that the following condition must hold for the charge to be self-sustained:

$$\gamma \nu_i \tau_r (1 - 1/\cosh(d/l_D)) = 1. \tag{17}$$

Introducing the parameters

$$U_0 = \frac{W}{e\gamma}, \quad B_0 = 1.5\,\frac{\sqrt{mW\nu_f/\nu_i}}{\gamma ed}, \tag{18}$$

and the dimensionless variables

$$u = U_c/U_0, \quad b = B/B_0, \tag{19}$$

we transform Eq. (17) as follows:

$$u(1 - 1/\cosh(2.61b/u)) = 1. \tag{20}$$

The physical significance of the new parameter U_0 is perfectly clear. It is the minimum possible voltage at which the discharge can burn; it is attained when all the fast electrons are able to expend their energy in the gap, and only then can they escape to the anode. In this case the number of ions formed by a single fast electron is $N_i = eU_0/W = 1/\gamma$. When the discharge burns at a higher voltage, each electron again produces $1/\gamma$ ions on the average, and the excess energy is transferred with the electron flux to the anode.

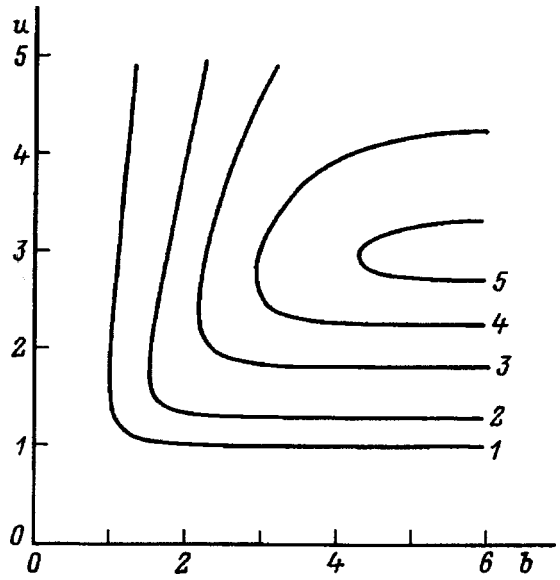


FIG. 1. Voltage versus induction of the magnetic field. *1)* $\beta = 0$; *2)* 0.2; *3)* 0.33; *4)* 0.37; *5)* 0.384.

Further transformations are needed to understand the physical significance of the parameter B_0. From Eq. (20) we obtain an explicit expression for the function $b(u)$:

$$b = \frac{u}{2.61}\,\mathrm{arccosh}\left(\frac{u}{u-1}\right). \tag{21}$$

The function (21) has a minimum equal to unity at the point $u = 1.73$. Accordingly, the inverse function $u(b)$, which characterizes the dependence of the discharge voltage on the induction of the magnetic field, is defined in the domain $b \geqslant 1$, in which it is two-valued (curve *1* in Fig. 1); for $b < 1$ or, equivalently, for $B < B_0$ it is impossible for the self-sustainment condition to be satisfied at any voltage. Consequently, B_0 represents the minimum magnetic induction at which it is still possible for the given form of discharge to burn.

Estimates of B_0 from the above expression are in good agreement with the results of discharge experiments[7,8] in a so-called inverted magnetron. It should be noted, however, that the experimental results show that the lower limit of the magnetic field range varies with the pressure in the discharge, although the actual dependence is not very strong: For example, in Ref. 7 it was found possible to ignite a high-current discharge with $B = 14$ mT at a gas pressure of 0.3 Pa, but when the pressure was doubled, the magnetic induction could be lowered to 13 mT. Because of the underlying assumptions ($\nu_i \sim$ const, $\nu_f \sim$ const) the proposed model does not describe this weak effect. A more accurate model must be developed to describe it.

The physical significance of the two-valuedness of the function $u(b)$ is most likely that electrons move comparatively slowly across the magnetic field under the conditions corresponding to the lower branch and manage to expend all their energy in ionization, whereas under the conditions of the upper branch fast particles have a high energy and can

effect a large number of ionizations, but they also diffuse more rapidly across the magnetic field and reach the anode, expending only an insignificant fraction of their energy. It is indeed correct to say that the qualitative form of the experimental $U(B)$ curves correspond to the lower branch of the given dependence, and for the upper branch such ascending curves have been observed in discharge ignition experiments,[9] but the present author does not know of any papers in which dependences of this kind have been observed for the discharge voltage. The states corresponding to the upper branch are probably unstable.

The following considerations are tendered in corroboration of the latter conjecture. The self-sustainment condition $\gamma N_i=1$ is satisfied exactly on the reduced $u(b)$ curve; to the left of this curve the ionization is weaker, and $\gamma N_i<1$, while to the right of the curve is a region of enhanced multiplication, $\gamma N_i>1$. We postulate that the discharge is fed by an emf source through a ballast resistor. Accordingly, if the discharge is in a state corresponding to the upper branch, for random attenuation of the intensity of the ionization processes and the discharge current the voltage drop across the ballast resistor decreases, and the voltage across the discharge gap increases. As a result, the discharge enters a region corresponding to reduced multiplication, ionization continues to diminish, and the discharge is extinguished. But if the discharge is in a state corresponding to the lower branch of the $u(b)$ curve, then in the analogous situation an increase in the discharge voltage causes the discharge to enter a region of enhanced multiplication, and a randomly occurring decrease in the ionization intensity is compensated. Thus, from the foregoing considerations and published experimental data we can only conclude that although the self-sustainment condition is satisfied for two different discharge voltages, a state corresponding to the lower branch of the $u(b)$ curve is stable and is the one that will occur experimentally.

The self-sustainment condition is a necessary but not a sufficient condition for the given form of discharge to occur. It is also necessary to determine the conditions under which a quasineutral plasma will fill up the entire gap from the cathode sheath to the anode. Ignoring the contribution of fast particles to the total negative charge, we write the quasineutrality condition in the form

$$n_e=n_i=n. \tag{22}$$

When the function $n_f(x)$ has the form determined here, the solution of the system of equations (8)–(11), (22) does not present any special difficulties. For the electron and ion flux densities we obtain, respectively,

$$n\upsilon_e=\frac{j_c}{e}\,\frac{\cosh(d/l_D)-\cosh(x/l_d)}{\cosh(d/l_D)-1}, \tag{23}$$

$$n\upsilon_i=\frac{j_c}{e}\,\frac{\cosh(x/l_D)-1}{\cosh(d/l_D)-1}. \tag{24}$$

To determine the concentration of the plasma, we combine Eqs. (10) and (11) and, rejecting terms that contain $d\varphi/dx$, after suitable transformations we obtain the equation

$$\frac{d}{dx}\left(\frac{(n\nu_i)^2}{n}+n\upsilon_b^2\right)=\frac{en\upsilon_e}{M\mu_e}, \tag{25}$$

where $\upsilon_b=\sqrt{kT_e/M}$ is the Bohm velocity.

We integrate this equation from a certain point x to the interface between the plasma and the ionic cathode sheath, disregarding the thickness of the cathode region and assuming that this interface is situated at the point $x\approx d$,

$$\left(\frac{(j_c/e)^2}{n(d)}+n(d)\upsilon_b^2\right)-\left(\frac{(n\nu_i)^2}{n}+n\upsilon_b^2\right)=\int_x^d\frac{en\upsilon_e dx}{M\mu_e}\equiv f(x). \tag{26}$$

Solving this equation for n, we obtain

$$n=\frac{g(x)+\sqrt{g^2(x)-4(n\nu_i)^2\upsilon_b^2}}{2\upsilon_b^2}, \tag{27}$$

where

$$g(x)\equiv\left(\frac{(j_c/e)^2}{n(d)}+n(d)\upsilon_b^2\right)-f(x). \tag{28}$$

The final equation (27) is physical only when the square-root expression is positive and the points at which this expression vanishes are boundaries of the quasineutral plasma. One such point is the interface between the plasma and the cathode sheath. Making use of the fact that $f(d)=0$ and $n\nu_i(d)=j_c/e$, we have

$$\left(\frac{(j_c/e)^2}{n(d)}+n(d)\upsilon_b^2\right)-\frac{2j_c\upsilon_b}{e}=0, \tag{29}$$

from which we find an equation for the plasma concentration at the interface with the sheath:

$$n(d)=\frac{j_c}{e\upsilon_b}. \tag{30}$$

Inasmuch as we are interested in the case where the plasma fills up the entire gas-discharge gap from the cathode sheath to the anode, the square-root expression must be positive throughout the rest of the gap, and only in an extreme case can it vanish at the anode. Allowing for the fact that the ion current is equal to zero at the anode, we find that the condition $g(0)\geqslant 0$ must hold in order for the given discharge form to occur; the satisfaction of this condition is ensured if

$$f(0)=\int_0^d\frac{en\nu_e dx}{M\mu_e}$$

$$=\frac{j_c}{M\mu_e}\left(\frac{d\cosh(d/l_D)-l_D\sinh(d/l_D)}{\cosh(d/l_D)-1}\right)<\frac{2j_c\nu_b}{e}. \tag{31}$$

Making use of the fact that the mobility coefficient of electrons across the magnetic field is proportional to the neutral gas concentration N, we can transform relation (31) as follows:

$$Nd\geqslant\frac{m}{2M}\,\frac{(\omega d)^2}{\upsilon_b k_e}F(b), \tag{32}$$

where $k_e = v_b/N$ is the effective collision frequency of plasma electrons, normalized to unit concentration of the neutral gas, and the function $F(b)$, which is defined by the relation

$$F(b) = u(b)\left(1 - \frac{\tanh(2.61b/u(b))}{2.61b/u(b)}\right), \tag{33}$$

varies only slightly as b increases (it goes from 0.7 at $b=1$ to 1 in the limit $b\to\infty$) and can be replaced by unity for estimates. Relation (32) gives the lower limit of the working pressure range for the given discharge form. It is interesting to note that relation (32) can be transformed to the following, which has a simple physical interpretation:

$$\overline{v}_e \approx \frac{\mu_e kT_e}{ed} \geqslant \frac{v_b}{2} \approx \overline{\nu}_i. \tag{34}$$

Consequently, the transition from the high-voltage to the high-current form of discharge at elevated pressure becomes possible, not because of stronger ionization as suggested in Refs. 2 and 7, but because the electron velocity across the magnetic field is comparable to or higher than that for ions. This conclusion is consistent with the published results of experiments on the conditions for the burning of a penning discharge in a hollow cathode,[10] where it has been shown that the discharge voltage rises sharply when the ratio of the anode area to the cathode area decreases below the value $\sim\sqrt{m/M}$. When this condition holds, the escape velocity of electrons from the discharge gap in zero field is lower than the ion escape velocity.

The lowest working pressure p_{cr} is attained for the minimum possible magnetic induction and can be determined from the equations

$$p_{cr} = \frac{0.35kTm}{k_e v_b Md}(\omega_0 d)^2 = \frac{0.78kTW\nu_f}{\gamma^2 k_e v_b Md\nu_i}, \tag{35}$$

where T is the temperature of the gas, and $\omega_0 = eB_0/m$.

At $p = p_{cr}$ the investigated discharge can burn only for one value of the magnetic induction. When the pressure is increased above p_{cr}, the range of magnetic fields broadens, whereupon different plasma concentration profiles can be obtained by varying B (Fig. 2), including the possibility of obtaining a distribution with a wide interval in which the plasma is nearly homogeneous (curve *2* in Fig. 2). This feature of the given type of discharge means that it can be used to generate beams of large cross section.

As the magnetic field decreases, the minimum working pressure increases approximately as the square of the magnetic induction. The region of pressures and magnetic fields in which the high-current form discharge is possible is shown in Fig. 3. To the left of curve OA it cannot be sustained in weak magnetic fields, owing to insufficient ionization and failure of the self-sustainment condition, nor can it exist below curve OC, because the electron velocity is not high enough, and the quasineutral state cannot be maintained in the interval from the cathode sheath to the anode. In measurements of the experimental dependence of the discharge voltage on the magnetic induction when a constant pressure is maintained, the operating point can be made to intersect

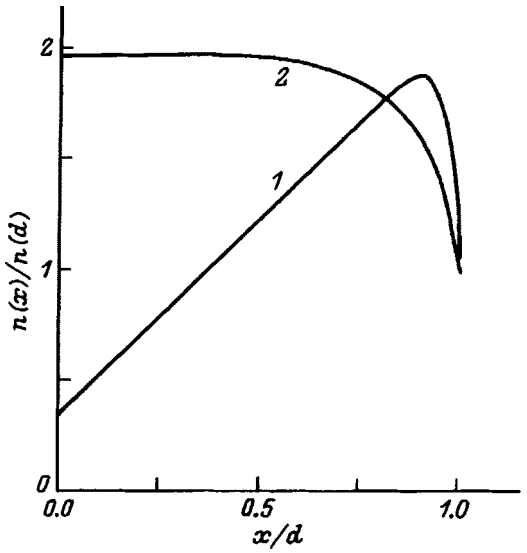


FIG. 2. Typical distributions of the plasma concentration in strong and weak magnetic fields. *1)* $b=6$; *2)* $b=1$.

curve OC by gradually increasing B, in which case the quasineutral state in the anode sheath is violated, and an electron layer begins to form in the vicinity of the anode. The measured dependence becomes monotonic in this case. The descending part, corresponding to the lower branch of the $u(b)$ curve, gives way to an ascending part, because the voltage drop across the anode sheath begins to contribute significantly to the total discharge voltage.[8]

It is important to note that the critical neutral gas concentration can be lowered considerably if its degree of ionization is sufficiently high, because the transfer of electrons across the magnetic field is initiated not only by collisions with neutral atoms, but also as a result of electron-ion collisions ($\mu_e \sim v_e + v_{ei}$). An analytical solution of the problem has not been obtainable in this case, but the results of numerical calculations lead to remarkably obvious conclusions. For example, when the frequency of electron-ion collisions is compared with the effective frequency of electron-atom col-

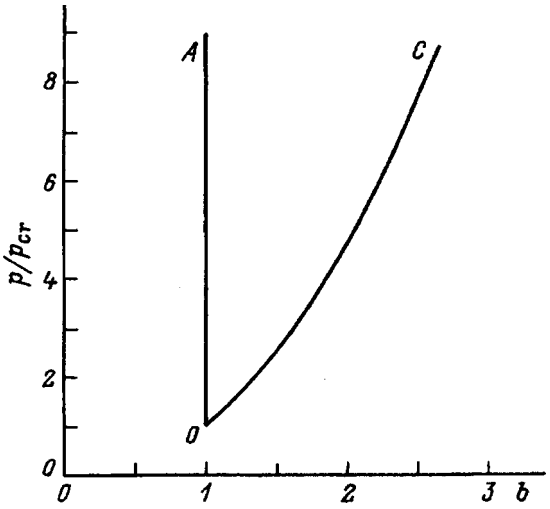


FIG. 3. Region in which the high-current form of glow discharge can occur in a magnetic field.

lisions, it is found that the critical concentration can be reduced by approximately one half. However, since a high degree of ionization is attained for high discharge currents and is accompanied by a significant increase in the temperature of the gas, one cannot look for a significant gain in the pressure of the gas. On the other hand, the results of experiments show that an increase in the discharge current is accompanied by a measurable increase in the critical pressure. The detailed analysis of the thermal processes involved in the investigated discharges and the determination of the gas temperature and the electron temperature are beyond the scope of the present study and could be the object of future research.

Another factor that could significantly influence the critical pressure is the possible onset of noise and instabilities in the plasma in strong magnetic fields, as they can lead to so-called anomalous diffusion and a sharp increase in the velocity of electrons across the magnetic field. These effects will obviously be conducive to broadening the working pressure range, but this regime, which is characterized by abrupt spatial and temporal irregularities in the plasma, should scarcely be recommended for the development of plasma sources of charged particles, at least not for those designed to generate beams of large cross section. But should there be some application for which the presence of a stable, homogeneous plasma is not a prerequisite and for which, at the same time, lowering of the pressure is to be desired, the indicated regime could be used.

INFLUENCE OF ELECTRON EMISSION ON THE DISCHARGE CHARACTERISTICS

Let the gas-discharge gap be bounded in the direction of the magnetic field and have a length L in this direction. A cathode potential is applied to the end electrodes to prevent the escape of electrons, but one of these electrodes is in the form of a grid with a transmittance α, through which electrons are emitted from the discharge plasma when an accelerating voltage is applied between it and an accelerating electrode. Skipping over distracting details associated with specific aspects of emission in the presence of electrode sheaths,[11] we assume that $\alpha n\bar{v}dt/4$ particles are emitted through a section of the grid of unit area during a time dt and, accordingly, that the number of particles nL in the fictitious column resting on this section decreases by the same amount:

$$d(nL) = -\alpha n\bar{v}dt/4. \tag{36}$$

From this relation we obtain an expression for the emission-induced variation of the concentration:

$$\frac{dn}{dt} = -\frac{\alpha\bar{v}n}{4L} = -\frac{n}{\tau}, \tag{37}$$

where we have introduced the characteristic time of the loss of particles through emission

$$\tau = \frac{4L}{\alpha\bar{v}}, \tag{38}$$

For plasma electrons we have $\bar{v} = \sqrt{8kT_e/\pi m}$ and for fast particles, whose velocities vary from $\sqrt{2eU_c/m}$ essentially to zero, the average velocity can be set equal to $\sqrt{eU_c/2m}$, whereupon we obtain the following expression for the characteristic time t_{eff} of fast-particle emission losses:

$$\tau_{\text{ef}} = \frac{4L}{\alpha}\sqrt{\frac{2m}{eU_c}}. \tag{39}$$

In the presence of emission the equation of continuity for fast particles acquires the form

$$\frac{d(n_f v_f)}{dx} = \frac{n_f}{\tau_r} + \frac{n_f}{\tau_{\text{ef}}} = \frac{n_f}{\tau_f}. \tag{2′}$$

where we have introduced the characteristic time of total losses of fast particles

$$\tau_f = \frac{\tau_r\tau_{\text{ef}}}{\tau_r + \tau_{\text{ef}}}. \tag{40}$$

The equation of motion (3) is unchanged, and the simultaneous solution of Eqs. (2′) and (3) yields the density of fast particles

$$n_f = \frac{\gamma j_c l_f \sinh(x/l_f)}{eD_f \cosh(d/l_f)}, \tag{41}$$

where $l_f = \sqrt{D_f\tau_f}$.

Disregarding the loss of ions at the ends, we assume as before that all the ions formed in the gap reach the cathode; introducing similar transformations, we obtain the following condition for the discharge to be self-sustained:

$$\frac{u}{1+\beta u^{3/2}}\left(1 - \frac{1}{\cosh(2.61b\sqrt{1+\beta u^{3/2}}/u)}\right) = 1, \tag{42}$$

where the parameter β is given by the equation

$$\beta = \frac{\alpha\sqrt{2W/m}}{8\gamma^{3/2}\nu_i L}. \tag{43}$$

It is evident at once that without emission ($\alpha = 0$) Eq. (42) goes over to (20). Equation (42) describes in implicit form the $u(b)$ curves shown in Fig. 1 for various values of the parameter β. Clearly, as β increases, the curves shift to the right, and the lower limit of the magnetic field range increases. An alternative interpretation is that every value of b has a corresponding critical value of the parameter β_{cr}, which cannot be exceeded without violating the condition for self-sustainment of the discharge. This property imposes a lower bound on the pressure required to sustain the discharge in the electron emission regime. Introducing the ionization constant $k_i = \nu_i/N$, we obtain the relation

$$p > \frac{\alpha kT\sqrt{2W/m}}{\beta_{\text{cr}}(b)8\gamma^{3/2}k_i L}. \tag{44}$$

The form of the function $\beta_{\text{cr}}(b)$ can be determined by deducing an explicit expression for the function $b(u)$ from Eq. (43) and analyzing it for the minimum. As a result, we obtain the function $b_{\min}(\beta)$, whose inverse is then the function $\beta_{\text{cr}}(b)$. The function $\beta_{\text{cr}}(b)$ is equal to 0 for $b = 1$, then

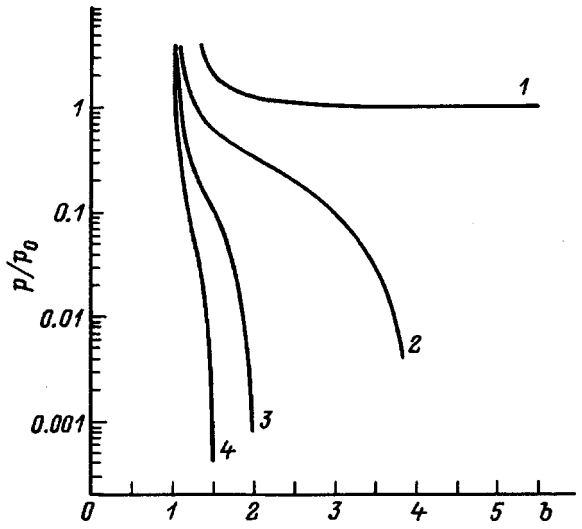


FIG. 4. Critical pressure versus magnetic induction. *1)* $d_e/d=1$; *2)* 0.75; *3)* 0.5; *4)* 0.3.

gradually increases, tending to 0.4 in the limit $b\to\infty$, and for $\beta>0.4$ the function $u(b)$ is no longer real. Consequently, the minimum pressure p_0 at which the given form of discharge can still occur with strong magnetic fields ($b\gg1$) applied to the gap is given by the relation

$$p_0=\frac{\alpha kT\sqrt{2W/m}}{3.2\gamma^{3/2}k_iL}. \tag{45}$$

In weak magnetic fields ($b\geqslant1$) the pressure must be even higher (curve *1* in Fig. 4), because now $\beta_{cr}<0.4$).

Thus, the situation of establishing a self-sustained discharge changes significantly upon transition to the electron emission regime. For self-sustainment without emission it is sufficient for the magnetic field to exceed a certain level, whereas in the presence of emission a definite pressure level must be established, where estimates based on Eq. (45) show that for not too great lengths L (of the order of a few times d) this level is substantially higher than the critical pressure determined from Eq. (35). This means that the transition to electron emission can have the effect of destabilizing the discharge, as is indeed confirmed by experimental results.[12] In Ref. 12 the attempt to extract electrons through the entire end electrode in a system of the inverted magnetron type made it necessary to significantly increase the admission of gas into the system in order to maintain a stable discharge, and this operation, in turn, created difficulties in the operation of the accelerating system of the source. On the other hand, the extraction of electrons only through the anode sheath region rather than through the entire end face did not produce any negative consequences.

An analysis of the situation with electrons extracted only from a selected part of the gap (through a section of width $d_e<d$) within the framework of the given model could help to explain this difference. For these calculations on the section $[d_e,d]$ we use the equation of continuity in the form (2), and on the section $[0,d_e]$ we use the form (2′). We match the solutions on the basis of the condition of continuity of the function $n_f(x)$ and its first derivative at the point $x=d_e$. An analysis of the resulting self-sustainment condition (which is too cumbersome to write out here) has shown that its satisfaction also requires a certain pressure level, which changes as the magnetic field is varied, but this level is considerably lower (by several orders of magnitude) than in the situation where electrons are extracted through the entire end face. It is evident from Fig. 4, which shows the critical pressure as a function of the magnetic field for several values of d_e/d, that a large difference is attained not only when $d_e/d\ll1$, but even at values $d_e/d\sim0.5$. This result is attributable to the fact that the density of fast particles drops abruptly from the cathode to the anode, and in the part of the discharge near the anode they are so few in number that they are capable of being ionized.

The drastic pressure reduction required to ensure the self-sustainment condition with electrons extracted only from the anode sheath region means that the lower limit of the working pressure range for the investigated discharge again does not depend on the ionization conditions, but on the conditions of motion of the plasma electrons, as was the case without emission. An analytical solution could not be obtained for the plasma concentration in the case of electron emission from a part of the gas-discharge gap. The results of a numerical analysis show that the range in which the high-current discharge form is possible broadens toward the low-pressure end in this case. This effect is attributed to an increase in the rate at which electrons leave the discharge gap, because now electrons escape not only across the magnetic field to the anode, but also along the magnetic field in the accelerating gap.[2] The expansion of the range in which discharge can occur sets the stage for the situation where a discharge burning in the high-voltage form switches to the high-current form when an accelerating voltage is applied and electrons are taken off. This event lowers the discharge voltage, as has also been observed experimentally.[12] It should also be noted that limiting the emission region does not rule out the attainment of a high emission efficiency. The results of calculations and experimental data attest to the correctness of this assertion.

CONCLUSION

To achieve the high-current form of a glow discharge with electron oscillations in a magnetic field, it is necessary that the induction of the magnetic field be equal to or greater than a critical value B_0, otherwise fast electrons escape to the anode before they can undergo a sufficient number of ionizations to satisfy the self-sustainment condition. In addition, it is necessary to maintain the pressure in the discharge at a level such that the velocity of plasma electrons across the magnetic field will be comparable with or greater than the ion velocity. The minimum working pressure is attained at $B=B_0$. As the magnetic induction increases, the minimum pressure increases approximately as the square of B under ''classical'' diffusion conditions.

The discharge conditions can change significantly in the presence of electron emission. The voltage can increase and the discharge can even die out when fast particles escape it in large numbers. This negative influence can be reduced con-

siderably if particles are taken off from the anode sheath region of the discharge, in which case limiting the current-takeoff region does not prevent emission from occurring with a high efficiency. Both before and after transition to the electron emission regime, if the operating point (B,p) falls within the limits of the region where discharge can exist, the voltage does not change significantly. Finally, it is possible for the situation to arise where electron emission can have the effect of creating a positively charged structure in a discharge burning in the high-voltage form with a predominance of negative charge, i.e., the discharge switches to the high-current form, and the voltage is lowered.

[1] M. D. Gabovich, *Physics and Technology of Plasma Ion Sources* [in Russian] (Atomizdat, Moscow, 1972), 304 pp.
[2] A. V. Zharinov and S. V. Nikonov, Zh. Tekh. Fiz. **58**, 1268 (1988) [Sov. Phys. Tech. Phys. **33**, 754 (1988)].
[3] *Electron Sources with a Plasma Emitter*, edited by Yu. E. Kreĭndel' [in Russian] (Nauka, Novosibirsk, 1976), 130 pp.
[4] S. P. Nikulin, Zh. Tekh. Fiz. **62**(12), 21 (1992) [Sov. Phys. Tech. Phys. **37**, 1142 (1992)].
[5] É. D. Lozanskiĭ and O. B. Firsov, *Spark Theory* [in Russian] (Atomizdat, Moscow, 1975), 271 pp.
[6] L. Eromberg and L. D. Smullin, J. Appl. Phys. **50**, 2634 (1979).
[7] E. M. Oks, A. A. Chagin, and P. M. Shchanin, in *Proceedings of the First All-Union Conference on Plasma Emission Electronics* [in Russian] (Ulan-Ude, 1991), pp. 18–23.
[8] V. V. Bersenev, N. V. Gavrilov, and S. P. Nikulin, in *Proceedings of the Conference on Low-Temperature Plasma Physics* [in Russian], Part 1, (Petrozavodsk, 1995), pp. 251–253.
[9] E. E. Kunhardt, in *The Behavior of Systems in the Space Environment*, edited by R. N. DeWitt *et al.* [in Russian] (Kluwer Acad. Publ., Dordrecht-Boston, 1993), pp. 669–711.
[10] A. A. Metel', Zh. Tekh. Fiz. **54**, 241 (1984) [Sov. Phys. Tech. Phys. **29**, 141 (1984)].
[11] Yu. E. Kreĭndel', S. P. Nikulin, and A. V. Ponomarev, Zh. Tekh. Fiz. **59**(6), 196 (1989) [Sov. Phys. Tech. Phys. **34**, 710 (1989)].
[12] N. V. Gavrilov, S. P. Nikulin, and A. V. Ponomarev, in *Proceedings of the Tenth International Conference on High Power Particle Beams* (USA, 1994), pp. 507–510.

Translated by James S. Wood

Changes in the fine structure of grain boundaries, induced by the absorption of helium, and helium embrittlement

V. I. Gerasimenko, I. M. Mikhaĭlovskiĭ, I. M. Neklyudov, A. A. Parkhomenko, and O. A. Velikodnaya

National Science Center, Kharkov Physicotechnical Institute, 310108 Kharkov, Ukraine
(Submitted February 5, 1996)
Zh. Tekh. Fiz. **68**, 64–69 (July 1998)

The changes in the structure of grain boundaries in tungsten due to the absorption of helium atoms are investigated experimentally and theoretically. Intergranular dilatation localized in a plane layer of subatomic thickness is observed. It is established that dilatation is accompanied by splitting of the cores of grain-boundary dislocations and a decrease in the grain-boundary stacking fault energy. The relationship of intergranular damage to the changes induced in the parameters of grain-boundary dislocations by the absorption of helium is discussed. [S1063-7842(98)01107-6]

INTRODUCTION

Grain boundaries are a major factor in determining the mechanical properties of irradiated materials. For example, an abrupt and irreversible loss of plasticity of irradiated metals and alloys at temperatures above $0.5T_m$ — high-temperature radiation embrittlement (HTRE) — can substantially shorten their trouble-free period of operation when exposed to radiation. An investigation of the nature of HTRE of reactor materials has revealed the significant role in this phenomenon of helium formed in nuclear reactions or of irradiation by alpha particles.[1] However, the mechanism underlying the radiation damage of materials as a result of the accumulation of helium in them remains elusive, especially for lack of information at the atomic level about the structural changes that take place in the early stages of helium buildup at the grain boundaries.

We have used field-ion microscopy methods to investigate the influence of helium absorption on the fine structure of high-angle grain boundaries with radiation-induced changes in the mechanical properties of polycrystalline materials. To exclude effects produced by peculiarities of the behavior of ensembles of disoriented grains, the investigations have been carried out on individual, crystallographically certified grain boundaries in bicrystalline samples.

EXPERIMENTAL PROCEDURE AND RESULTS

The investigations were carried out in a field-ion microscope with liquid-nitrogen cooling of the samples. The imaging gas was helium at a pressure of $(1-2)\times10^{-2}$ Pa, and the residual gas pressure did not exceed 10^{-6} Pa. We investigated bicrystalline tungsten samples (99.98% pure) in the form of points with a radius of curvature equal to 10–80 nm at the tip. The samples were irradiated *in situ* by low-energy helium ions formed in the collision of helium atoms with electrons emitted by the sample when a pulsating negative potential was applied to it. The value of this potential was chosen in correspondence with the radius of curvature of the sample at the tip so as to be sufficient for attaining an electric field of $(3-5)\times10^7$ V/cm and, accordingly, an electron emission current density of 10^6-10^7 A/cm^2. The energy distribution of the helium ions formed during passage of the electron stream was close to a Maxwellian distribution and matched the distribution of ions in the low-energy plasma of Tokamak-type equipment.[2] The fluence ϕ and the energy spectrum of the ions were calculated by a method developed in Refs. 2 and 3; the value of ϕ in the experiment was varied in the range $1\times10^{13}-5\times10^{17}$ ion/cm^2. The average energy of the helium ions was 200–350 eV, so that the energy transferred in primary collisions was below the threshold for the displacement of tungsten atoms (50 eV) and was just high enough for the injection of helium atoms. The temperature of the sample was maintained in the interval 100–400 K during irradiation, thereby affording the possibility of intragranular migration of helium by the interstitial mechanism while preventing migration by the vacancy mechanism.[4] Bearing in mind that in the irradiated hemispherical part of a bicrystal its surface and grain boundaries are sinks for interstitial atoms and that the vacancy sink is eliminated, we can assume in the first approximation that the total flux of helium atoms onto the parts of the grain boundaries situated in the investigated zone near the surface is equal to the fluence of bombarding ions.

After irradiation at an elevated temperature the sample was cooled to 78 K, its surface was cleaned by low-temperature field desorption, and successive field-ion images of the sample were recorded during controlled layer-by-layer field evaporation.

Bicrystals containing high-angle, strongly bound grain boundaries were chosen for the investigation. The field-ion images of such crystals usually lack any traces of preferential field etching of parts of the grain boundaries, so that displacements of crystal atoms during irradiation can be recorded with a resolution to 1 Å in the image plane and to 0.1 Å in the direction perpendicular to the image plane, owing to the indirect magnification effect.[5]

Figure 1 shows field-ion images of a coherent inter-

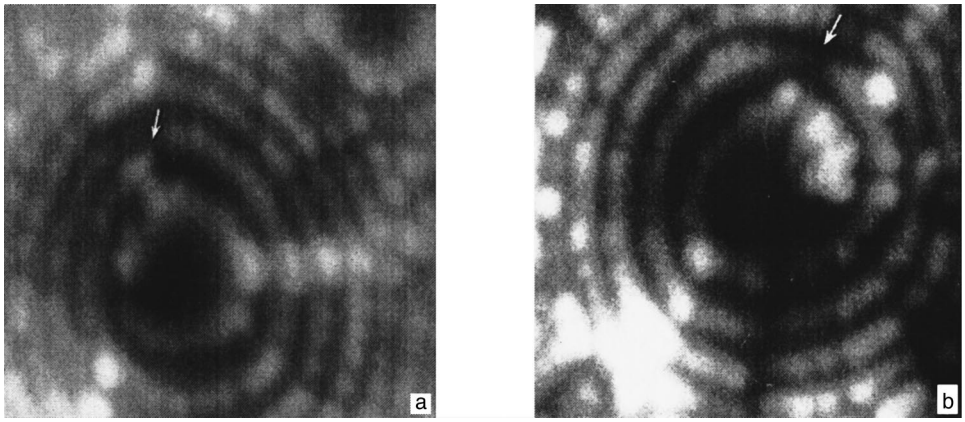


FIG. 1. Field-ion micrographs of a high-angle coherent grain boundary before (a) and after (b) irradiation with helium ions.

granular boundary with a 52° angle of misorientation relative to the [110] axis. The arrows indicate the boundary positions corresponding to the bonding $(115)_{\mathrm{I}}$ and $(1941\ 35)_{\mathrm{II}}$ planes of crystals I and II. One observes bonding of the $(110)_{\mathrm{I}}$ and $(110)_{\mathrm{II}}$ planes, which is broken only in the zone where the 1/2 [110] grain-boundary dislocation (indicated by the white arrow in Fig. 1a) approaches the surface. After the bicrystal is irradiated with helium ions having an average energy $W=300$ eV at a fluence $\phi=9\times10^{14}$ ion/cm^2, a narrow dark fringe is observed (Fig. 1b) in the field-ion image, running along the initial trace of the grain boundary (this fringe is indicated by arrows). The average width of this fringe is 2.2 Å, which is smaller than the minimum interatomic spacing in tungsten (2.7 Å) and is much smaller than the average interatomic spacing (3.5 Å) on the steps of the (110) face. The narrow width of the dark fringe (less than the interatomic spacing) indicates that it is not the result of preferential field evaporation of the boundary material. It is reasonable to conclude that a ''gap'' of subatomic width is formed between the grains at the boundary when the sample is irradiated with helium ions. An analysis of the micrograph in Fig. 1b shows that in addition to the above-described increase in the distance between grains in the direction perpendicular to the surface of the boundary, i.e., linear intergranular dilatation, a rigid shift of the lattices of adjacent grains along the boundary is also observed. The shift is detected from the relative displacement of the steps of the $(110)_{\mathrm{I}}$ and $(110)_{\mathrm{II}}$ faces. A calculation of the shift in the [110] direction by the indirect magnification method[5] gives an average value of 0.5 Å. The appearance of the shift along the grain boundary can be described in terms of the formation of a grain-boundary stacking fault.[6]

Figure 2 shows the dependence of the linear intergranular dilatation of a grain boundary on the fluence of helium ions at an ion energy $W=300\pm50$ eV. As the fluence is increased in the interval $10^{14}-10^{15}$ ion/cm^2, a dark fringe is observed to emerge along the boundary, i.e., a linear intergranular dilatation $\overline{l}$ of width up to 2.3 Å occurs. The quantity $\overline{l}$ represents the dilatation averaged over the results of measurements of the width of the dark region along the grain boundary in different sections of it.

When the fluence is increased further from 10^{15} ion/cm^2 to 2×10^{17} ion/cm^2, the average dilatation $\overline{l}$ scarcely increases at all for the indicated energy $W=300$ eV. Figure 3 shows a micrograph of a bicrystal, obtained after its irradiation with ions at a higher average energy $W=350$ eV with a fluence of 2×10^{17} ion/cm^2. It follows from Fig. 3 that adjacent grains are misoriented by a 33° angle relative to the [110] axis, and the width of the dark fringe along the trace of the boundary (indicated by dark arrows) is 2.0 ∓ 0.6 Å. The surface emergence of a 1/2 [110] grain-boundary dislocation, indicated by a white arrow in Fig. 3, was observed in this experiment. The half-width of the core of this dislocation is

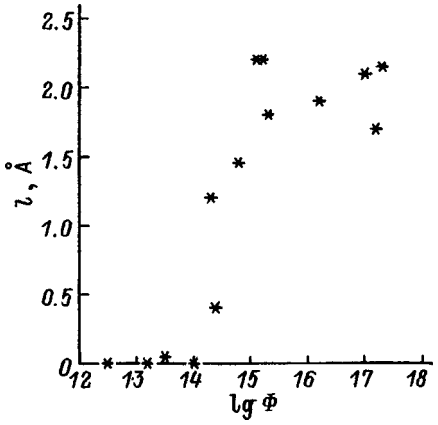


FIG. 2. Linear intergranular dilatation versus fluence of helium ions.

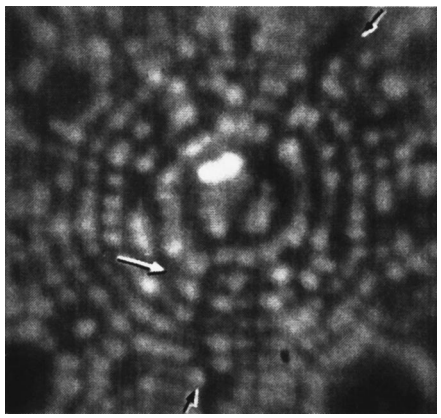


FIG. 3. Tungsten bicrystal after irradiation with helium ions and field evaporation to a depth of 20 atomic layers.

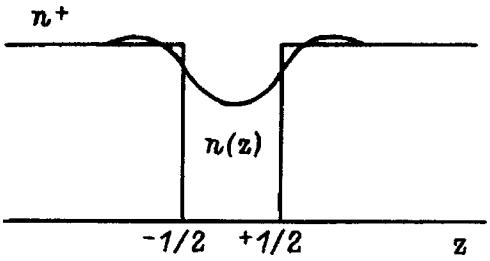


FIG. 4. Ion density n^+ and electron density $n(z)$ in the vicinity of a grain boundary.

determined from the distance between the dislocation emergence site and the step of the (110) plane farthest from this site, where a kink is also observed; the half-width of the core is 9 ∓ 2 Å, which is approximately twice the half-width of the dislocation core in nonirradiated tungsten.[6] An estimate of the grain-boundary stacking fault energy from the width of the dislocation core, based on the theory of elasticity, gives a value of 0.16 J/m^2 for the recorded displacement of the lattices of adjacent grains along the direction of the [110] misorientation axis. By way of comparison the grain-boundary stacking fault energy in nonirradiated tungsten is 0.33 J/m^2.

JELLIUM MODEL OF INTERGRANULAR HELIUM ABSORPTION

The cohesive strength of the grain boundaries of materials doped with harmful impurities depends largely on the redistribution of the electron density induced by them. To account for the atomic mechanisms underlying the influence of impurities, we draw on notions of the formation of highly polarized, directional chemical bonds.[7] The situation with helium present at the boundaries cannot be described by these notions, because helium is a neutral element.

Let us suppose, for example, that helium has appeared at a grain boundary as a result of diffusion transport. Relying on the model of a homogeneous positive background (the jellium model), we consider the disjoining effect of helium atoms adsorbed at the intercrystallite boundaries. We introduce a plane interface between two grains in the form of a gap in a homogeneous positive background (Fig. 4). The helium atoms are uniformly distributed in the plane $Z=0$ and repel the grains in the direction of the normal to this plane. The helium atoms interact with the metal mainly by way of the electron gas; the direct interaction of helium atoms with the screened metal ions is slight and decays rapidly with distance. We can assume in this regard that the energy variation associated with the introduction of helium at the boundary depends only on the local electron density. The energy shift of the electron ground state of the metal with the introduction of a helium atom is equal to[8]

$$\Delta E_{am}=\alpha N(Z,l), \tag{1}$$

where $\alpha=5.51$ in atomic units ($h=l=m=1$).

The total energy variation of the metal due to the grain-boundary absorption of helium atoms is equal to

$$\Delta E(l)=N\Delta E_{am}-V_b(l), \tag{2}$$

where $V_b(l)$ is the binding energy of the two grains, and N is the number of helium atoms on unit surface of the grain boundaries.

The binding energy of two identical metals separated by a gap of width l is satisfactorily described by the semiempirical expression[9]

$$V_b(l)=V_b^0\exp(-\gamma l)(1+\gamma l), \tag{3}$$

where $\gamma=0.9\cdot\lambda_{\mathrm{TF}}^{-1}$, $\lambda_{\mathrm{TF}}^{-1}=1/3\cdot(9/4\pi)^{1/3}r_s^{1/2}$ is the Thomas–Fermi screening parameter, $4/3\pi\cdot r_s^3=1/n_0$, and n_0 is the density of the homogeneous electron gas in the grain interior.

By definition $V_b=2E_s^0$ for $l=0$, where E_s is the surface energy.

The determination of the electron density $n(Z)$ in the vicinity of the grain boundaries reduces to the problem of the response of the system of electrons to the effective perturbing potential $V(Z)$ associated with the formation and broadening of the gap in the homogeneous positive background during the absorption of helium. This potential has the form

$$V(Z)=2\pi\int_{\infty}^{\infty}|Z-Z'|[n_+(Z')-n_-(Z')]dZ' + V_{\text{ex-cor}}(n_0)[n(Z)-n_0]. \tag{4}$$

Here $V_{\text{ex-cor}}$ is the sum of the local exchange and local correlation energies. Based on linear response theory, the perturbation of the electron density $\delta n(Z,l)=n_0-n(Z,l)$ in Fourier representation is equal to

$$\delta n(\lambda_{\mathrm{TF}},\, l)=-F(q)V(q), \tag{5}$$

where

$$F(q)=K_{\mathrm{F}}/2\pi^2((1-q^2/2q)\ln|1-q/1+q|-1), \tag{6}$$

$$V(q)=4\pi/q^2\,\delta n_0(q)-(4\pi/q^2+V'_{\text{ex-cor}})\,\delta n(q), \tag{7}$$

$q=q/2K_{\mathrm{F}}$, K_{F} is the Fermi wave vector, $\delta n_0=n_0-n_+$, and n_+ is the density of the positive background.

In the long-wavelength approximation ($q\to 0$) we have $F(q)=-K_{\mathrm{F}}/\pi^2$, and

$$\delta n(q,l)=(q_0^2/q^2)\,\delta n_0(q)/[1+q_0^2/q^2-\lambda], \tag{8}$$

$$\lambda=|V'_{\text{ex-cor}}|\,\nu(K_{\mathrm{F}})=1/\pi K_{\mathrm{F}},\quad q_0=\sqrt{4K_{\mathrm{F}}/\pi}, \tag{9}$$

$$\delta n_0(q)=\int_{-\infty}^{\infty}\delta n_0(Z)\exp(iqz)=2n_0\,\sin(ql/2)/q. \tag{10}$$

The variation of the electron density is expressed in the form

$$\delta n(z,l)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{q_0^2 n_0(\sin(ql)/2)/q}{q_0^2+q^2(1-\lambda)}\exp(iqz)dq, \tag{11}$$

so that the total electron density at the center of the gap is equal to

$$n(0,\, l)=n_0\,\exp(-q_0 l/2\sqrt{1-\lambda}). \tag{12}$$

At high electron densities, as are characteristic of transition metals, relation (11) corresponds to the quasiclassical case, and the inhomogeneity of the electron density depends on the Thomas–Fermi screening parameter. This expression, strictly speaking, is valid only for a narrow gap,

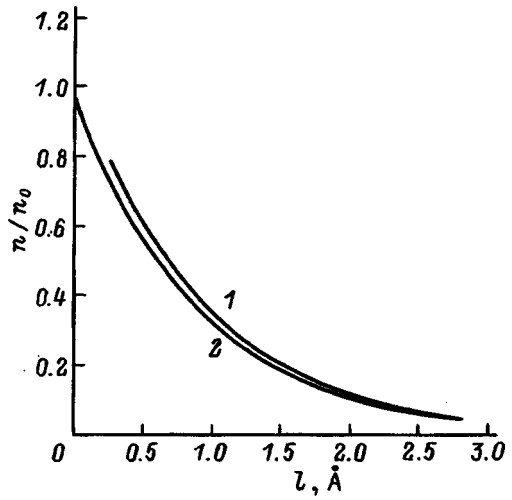


FIG. 5. Profile of the electron density at the center of the gap between crystallites.

$l/2\lambda_{TF}\sqrt{1-\lambda} \ll 1$ (small perturbation). On the other hand, relation (1) is satisfied if the variation of the electron density over a distance equal to the diameter of the helium atom is small, but this condition is possible only for a gap of finite dimensions. In this light, to test the validity of expression (11) for $l \sim \lambda_{TF}$, we have carried out quantum-mechanical calculations of the electron density distribution by the density functional method[10] near the edge of the homogeneous positive background. The electron density at the center of the gap was determined as twice the electron density at a distance $l/2$ from the edge of the homogeneous background. In the density functional formalism[10] the problem of determining the electron density at the boundary of the homogeneous positive background reduces to the self-consistent solution of the equation

$$-\frac{1}{2}\Psi_k''(Z)+V(Z)\Psi_k(Z)=k^2/2\Psi_k(Z), \tag{13}$$

where $V(Z)$ is given by Eq. (4).

The electron density is

$$n(Z)=\frac{1}{2\pi}\int_0^{k_F}(k_F^2-k^2)\Psi_k^2(Z)dk. \tag{14}$$

The solution of this system of equations must satisfy the electroneutrality condition

$$4\pi\int_{-\infty}^{\infty}(n_0-n(z))dz=0 \tag{15}$$

and the boundary conditions $\psi(z)(z\to\infty)=0$ and $\psi(z\to-\infty)=\sin(kz+\delta)$. The calculations are carried out for the case $K_F=1$, which approximately corresponds to the electron density of aluminum. It follows from Fig. 5 that the results of the electron density calculations (curve *1*) agree satisfactorily with the dependence of the electron density at the center of the gap on the width of the gap (curve *2*) as determined within the framework of linear response theory. In the calculations that follow, therefore, we use the analytical equation (11) for the local electron density.

TABLE I. Parameters of the jellium model of a high-angle grain boundary in tungsten.

λ_{FT}	n_0	K_F	$\varkappa$	α	γ	V_b^0
0.814	0.056	1.23	1.44	5.51	1.105	3.6×10^{15} eV/cm^2

Solving Eqs. (1), (3), and (11) simultaneously and setting the derivative $d\Delta E(l)/dl$ equal to zero, we obtain an equation for the gap width l as a function of the helium concentration at the grain boundary:

$$\varkappa/2N\alpha n_0 \exp(-\varkappa l/2)=\gamma^2 V_{vb}^0\cdot l\exp(-\gamma l), \tag{16}$$

where $\varkappa=(1-\lambda)^{-1/2}/\lambda_{TF}$.

If $\gamma-\varkappa/2>0$ (as is observed for all metals), the function on the right side of Eq. (16) has a maximum at the point

$$l_c=1/(\gamma-\varkappa/2), \tag{17}$$

the critical helium density at the grain boundaries, i.e., the density at which the intergranular cohesive force becomes equal to the repulsive forces induced by helium adsorption, is equal to

$$N_c=(2\gamma^2\cdot l_c V_b^0)/(e\varkappa\alpha n_0), \tag{18}$$

and for $l \ll l_c$ it follows from relation (12) that

$$l=\xi\cdot N, \tag{19}$$

where $\xi=\alpha n_0\varkappa/2\gamma^2 l_c V_b^0$.

Table I gives the parameters of the jellium model of a tungsten bicrystal in atomic units, calculated from data in Ref. 11.

The calculations show that the dilatation increases sharply upon an order-of-magnitude increase of the surface density of helium N and attains $l_c=1.4$ Å for $N_c=6\times10^{14}$ atoms/cm^2, for which dl/dN tends to infinity. These results agree satisfactorily with the experimental field-ion microscope data. The values of l determined from the field-ion microscope experiments correspond to the critical state of the boundaries. As a rule, however, the measured values of l are greater than the theoretical value l_c.

DISCUSSION OF THE RESULTS

The intergranular dilatation, dislocation density, and stacking fault energy play an important role in the evolution of the strength properties of materials. The presence of helium in materials, where it segregates predominantly at the grain boundaries, tends to alter the strength and plastic properties of the material and is responsible, in particular, for the high-temperature radiation embrittlement of materials.[1] In this regard, the theoretical and experimental results obtained in the first and second sections can be used to explain a number of phenomena observed in materials with different types of crystal lattices.

Despite the qualitative agreement of the experimental and theoretical results, the average width of the observed darker grain-boundary fringes exceeds the maximum width l_c of the intergranular gap in the jellium model. This discrepancy is probably attributable to the fact that the emission

contrast reflects a perturbation of the electron density near the boundary. In the vicinity of the subatomic ''gap'' in the crystal the perturbation of the electron density propagates into the bulk of the two halves of the bicrystal to a depth of the order of λ_{TF} (Ref. 10). Consequently, the grain-boundary dilatation corresponds to an electron-density perturbation of width $l+2\lambda_{TF}$, which is in satisfactory agreement with the field-ion microscope data.

A region in which $l(\phi)$ is independent of the fluence (at $\phi > 10^{15}$ ion/cm^2) has been observed experimentally, probably indicating that the helium concentration is no longer increasing at these fluences. This result can be identified with the emergence of atoms to the free surface as a result of the helium acquiring grain-boundary mobility at the irradiation temperature (100 K) when the intergranular dilatation increases to a value close to l_c. The onset of mobility in bulk polycrystalline samples will necessarily cause helium to build up in the micropores and cause growth of the latter at the grain boundaries.

It has been shown[12] that the energy of the grain boundaries can be represented by Fourier series. For grain boundaries oriented along close-packed planes the factors to be taken into account can be limited to interaction between nearest and next-nearest neighbor atomic planes. For a symmetric tilt boundary oriented at 70.5° relative to the [110] axis, corresponding to a lattice of coincident nodes with an inverse density of coincidences $\Sigma = 3$, the expansion can be limited to the first term for estimating the influence of helium adsorption on the grain-boundary stacking fault energy ε. Assuming that a grain-boundary stacking fault is formed when the lattice of adjacent grains shifts by $\alpha/12\langle 111\rangle$ (Ref. 13), we infer from the expression for the energy of the boundaries[12] that

$$\varepsilon(N) \sim \exp[-4\pi/\sqrt{3}(d_{211}+l(N))/a], \tag{20}$$

where d_{211} is the (211) interplanar distance.

For not too large values of the relative variation of the grain-boundary stacking fault energy

$$\varepsilon(N)/\varepsilon_0 = \exp[-4\pi/\sqrt{3}\xi N/a], \tag{21}$$

where ε_0 is the grain-boundary stacking fault energy for $N=0$ (in the absence of helium).

As shown above, the maximum intergranular dilatation preceding damage is equal to λ_{TF} (0.42 Å for W). The substitution of this value into relation (21) gives $\varepsilon(N_c)/\varepsilon_0 = 0.38$, which agrees satisfactorily with the results of field-ion microscope studies of the influence of helium on the width of the cores of grain-boundary dislocations. The results can be used to explain the decrease in the strength of the grain boundaries of nickel irradiated and strained at high temperatures (see, e.g., Ref. 14).

Indeed, the normal (adhesive) strength has a very steep dependence on the intergranular dilatation ($\sigma \sim 1/l^4$), and the resistance to intergranular damage diminishes abruptly as the internal (disjoining) pressure[4] in the gas-filled grain-boundary voids increases. The introduction of helium into the grain boundaries, producing grain-boundary dilatation and increasing the disjoining pressure, is also accompanied by a change in the properties of grain-boundary dislocations, i.e., the stacking fault energy (an increase in the width of the core) (Fig. 3).

It is interesting that this result can be obtained from first principles of the theory of elasticity with definite but very insignificant approximations, and the numerical value of the relative variation of the grain-boundary stacking fault energy (at maximum helium concentration) agrees satisfactorily with the results of microscope observations: The theory gives $\varepsilon(N_c)/\varepsilon_0 = 0.38$, and the experimental value is 0.45.

The variation of the grain-boundary stacking fault energy can also serve as a basis for explaining phenomena observed in our work: the decrease in the grain-boundary sliding voltage[14] and, as a consequence, the intensification (acceleration) of grain-boundary damage to materials deformed at high temperatures.[15] According to the Peierls model, the voltage required to move a dislocation is equal to[16]

$$\sigma = 2\mu/(1-\nu)\exp(-4\pi r_{cor}/b_{gb}), \tag{22}$$

where μ is an effective grain-boundary modulus, ν is the Poisson ratio, r_{cor} is the radius of the core of the grain-boundary dislocation, and b_{gb} is the Burgers vector of grain-boundary dislocations.

Clearly, the required voltage depends on the width of the core of the grain-boundary dislocation. According to the theoretical and experimental results obtained in the present study, the presence of helium at the grain boundaries must lower the resistance of grain-boundary dislocations to motion and stimulate grain-boundary sliding, as is indeed observed in samples of various materials prone to high-temperature radiation (helium) embrittlement.

CONCLUSIONS

1. The results of theoretical and experimental investigations indicate an increase in the intergranular dilatation and enlargement of the cores of grain-boundary dislocations when helium is adsorbed at the boundaries. The latter result can be interpreted as a reduction in the energy of grain-boundary dislocation stacking faults. An equation has been derived, interrelating the helium concentration and the grain-boundary stacking fault energy.

2. A phenomenological dislocation model of helium embrittlement has been presented, relating the loss of strength of grain boundaries to the stimulation of grain-boundary sliding as the parameters of grain-boundary dislocations change in a helium-saturated material.

This work has received partial support from the International Science Foundation and the State Committee of Ukraine on Science and Technology Issues: the Funds for Fundamental Research (Project No. 2.3/933) and State Scientific-Technical Programs (Project Nos. 5.42.06/040 and 7.02.05/-93).

The authors are indebted to A. S. Bakaĭ and P. A. Bereznyak for helpful discussions.

[1] I. M. Neklyudov, L. S. Ozhigov, N. M. Kiryukhin, and A. A. Parkhomenko, KhFTI AN Ukr. SSR Preprint No. 83-43 [in Russian] (Kharkov Physicotechnical Institute, Academy of Sciences of the Ukrainian SSR, Kharkov, 1983), 47 pp.
[2] Zh. I. Dranova and I. M. Mikhaĭlovskiĭ, in *Problems of Nuclear Science and Engineering (Series on the Physics of Radiation Damage and Radiation Materials Research)* [in Russian], No. 2(13) (1980), pp. 31–34.
[3] P. A. Bereznyak and V. V. Slezov, Radiotekh. Elektron. **17**, 354 (1972).
[4] A. G. Zaluzhnyĭ, Yu. N. Ṣokurskiĭ, and V. N. Tebus, *Helium in Reactor Materials* [in Russian] (Énergoatomizdat, Moscow, 1988), pp. 3210–3215.
[5] I. M. Mikhaĭlovskiĭ, Fiz. Tverd. Tela (Leningrad) **24**, 3210 (1982) [Sov. Phys. Solid State **24**, 1822 (1982)].
[6] I. M. Mikhaĭlovskiĭ, Zh. I. Dranova, and V. A. Ksenofontov, in *Problems of Nuclear Science and Engineering (Series on the Physics of Radiation Damage and Radiation Materials Research)* [in Russian], No. 2(25) (1983), pp. 14–19.
[7] M. E. Eberhart, R. M. Latanison, and K. H. Johnson, Acta Metall. **3**, 1769 (1985).
[8] V. I. Gerasimenko, in *Problems of Nuclear Science and Engineering (Series on the Physics of Radiation Damage and Radiation Materials Research)* [in Russian], No. 4(37) (1985), pp. 10–14.
[9] J. H. Rose, J. Ferrante, and J. R. Smith, Phys. Rev. Lett. **47**, 675 (1981).
[10] N. D. Lang, Solid State Phys. **28**, 225 (1973).
[11] A. N. Zelikman and L. S. Nikitina, *Tungsten* [in Russian] (Metallurgiya, Moscow, 1978), 272 pp.
[12] V. I. Gerasimenko, I. M. Mikhaĭlovskiĭ, P. Ya. Poltinin, and L. I. Fedorova, Fiz. Met. Metalloved. **68**, 806 (1989).
[13] A. I. Orlov, V. N. Perevezentsev, and V. V. Rybin, *Grain Boundaries in Metals* [in Russian] (Metallurgiya, Moscow, 1980), 156 pp.
[14] V. F. Zelensky, I. M. Neklyudov, L. S. Ozhigov, and A. A. Parkhomenko, in *Effects of Radiation Materials, Vol. 1* (*American Society of Testing Materials Special Technical Publication 1046*) (ASTM, Philadelphia, 1989), pp. 295–300.
[15] R. Z. Valiev, V. G. Khaĭrullin, and A. D. Sheĭkh-Ali, Izv. Vyssh. Uchebn. Zaved. Fiz. , No. **3**, 93 (1991).
[16] J. P. Hirth and J. Lothe, *Theory of Dislocations* [McGraw–Hill, New York, 1968; Atomizdat, Moscow, 1972, 599 pp.].

Translated by James S. Wood

Mechanical properties of materials and the object of description in gauge theories

Yu. V. Grinyaev and N. V. Chertova

Institute of Physics of Strength and Materials Science, Siberian Branch of the Russian Academy of Sciences, 634821 Tomsk, Russia

(Submitted January 9, 1997)
Zh. Tekh. Fiz. **68**, 70–74 (July 1998)

In the continuum theory of defects the total strain of an inelastic material can be represented by the sum of the reversible elastic strain associated with external loads, the compatible elastoplastic strain due to defects of the material, and the compatible plastic strain responsible for irreversible deformation of the material. The proposed scheme of separation of the strain into distinct components can be used to determine the physical significance of the inelastic properties of materials and the object of description of a gauge model representing a dynamical generalization of the continuum theory of defects. © *1998 American Institute of Physics.* [S1063-7842(98)01207-0]

INTRODUCTION

In the many papers[1–6] devoted to the gauge description of the deformation of solids containing defects the most detailed attention has been given to the mathematical formalism of gauge theories, the relationship to the continuum theory of defects, phenomenological generalizations of the theory, and plane-wave solutions. The feasibility of describing the basic mechanical properties of solids by the given approach is not discussed in the cited papers. Real materials exhibit three basic mechanical properties in varying degrees: elasticity, plasticity, and viscosity. The elastic properties of materials are well described by the classical theory of elasticity,[7] which treats reversible equilibrium processes. In the elastic deformation domain, when the stresses (independent parameter) change, the elastic strain (conjugate parameter) ''instantaneously'' adjusts to the equilibrium value, and during unloading the system returns to the initial state by the same path as in loading. As parameters characterizing the state of a system, the stresses and elastic strains are related one-to-one by the constitutive equation, i.e., by Hooke's law. The stress-strain diagram, including subsequent unloading (Fig. 1), can be used to establish which part of the total strain is reversible, i.e., elastic. In phenomenological theories the residual strain in the material after elastic unloading is assumed to be plastic. Theories of plasticity deal with problems related to the part of the $\sigma-\varepsilon$ diagram where it departs from the linear relation between strains and stresses. The most common approach in the solution of plasticity problems is to approximate the real $\sigma-\varepsilon$ diagrams by a set of piecewise-linear curves representing an ideal elastoplastic solid or an elastic solid with linear hardening. Phenomenological models of viscous media in the elastic, plastic, and elastoplastic deformation domains (this separation is customary in continuum mechanics[8]) express the dependence of the stressed state at a particular time t on the loading history in the time from zero to t. In the opinion of Pazhina,[8] the total strain in a viscoelastic–viscoplastic solid can be represented by the sum of three terms

$$\varepsilon^{\mathrm{tot}}=\varepsilon^{\mathrm{el}}+\varepsilon^{\mathrm{v}}+\varepsilon^{\mathrm{pl}}, \tag{1}$$

where $\varepsilon^{\mathrm{el}}$, ε^{v}, and $\varepsilon^{\mathrm{pl}}$ are the elastic, viscous, and plastic strain components.

The physical nature of the strain components (1) is not discussed in the cited paper. To ascertain the meaning of Eq. (1) and to determine the object described in the gauge model, we propose to analyze the total strain on the basis of the continuum theory of defects; the motivation for the latter choice will become clear below in the description of the mathematical algorithm used to construct the gauge theory.

ANALYSIS OF THE TOTAL STRAIN IN THE CONTINUUM THEORY OF DEFECTS

The total strain in a material with defects can be written as a sum of three terms, which are well known in the continuum theory of defects:

$$\varepsilon^{\mathrm{tot}}=\varepsilon^{\mathrm{el}}+\varepsilon^{\mathrm{el-plD}}+\varepsilon^{\mathrm{pl}}, \tag{2}$$

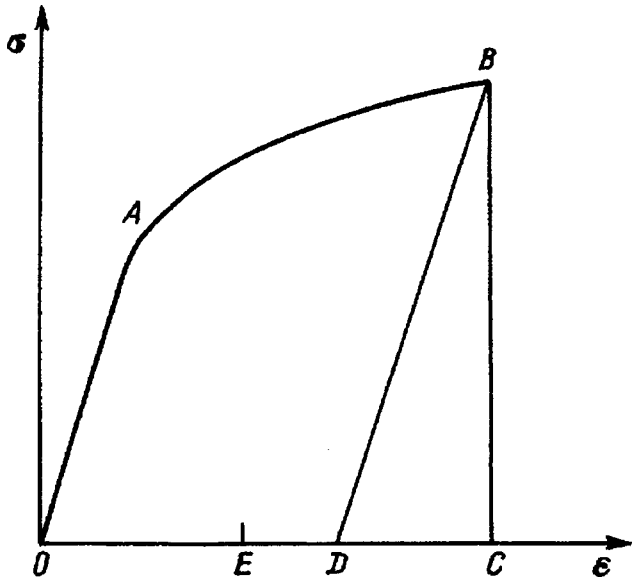


FIG. 1.

each of which represents a symmetric part of the gradient of the continuous displacement vector

$$u^{\mathrm{tot}}_{(i,j)} = u^{\mathrm{el}}_{(i,j)} + u^{\mathrm{el-plD}}_{(i,j)} + u^{\mathrm{pl}}_{(i,j)}. \tag{3}$$

In this equation and below, the subscripts in parentheses indicate symmetrization, and the comma signifies differentiation with respect to the coordinate. The first terms in Eqs. (2) and (3) correspond to elastic deformation associated with external loads, and when the latter are lifted, it vanishes at the speed of sound. The strain $\varepsilon^{\mathrm{el}}$ satisfies the compatibility condition

$$e_{ikn} e_{jlm} \varepsilon^{\mathrm{el}}_{nm,kl} = 0 \tag{4}$$

and is aptly called the external field, since it is attributable to external loads. The second term in Eq. (2) characterizes the compatible elastoplastic strain associated with defects of the material:

$$e_{ikn} e_{jlm} \varepsilon^{\mathrm{el-plD}}_{nm,kl} = 0. \tag{5}$$

As is customary in the continuum theory of dislocations, the gradient $u^{\mathrm{el-plD}}$ represents the sum of the elastic and plastic distortions

$$u^{\mathrm{el-plD}}_{i,j} = \beta^{\mathrm{elD}}_{ji} + \beta^{\mathrm{plD}}_{ji}, \tag{6}$$

neither of which is a gradient of the continuous displacement vector. By definition, an arbitrarily specified plastic distortion β^{plD} corresponds to a dislocation density

$$\alpha_{ij} = -e_{ikl} \beta^{plD}_{lj,k}, \tag{7}$$

and the elastic distortion β^{elD} characterizes distortions of the solid that are responsible for its continuity at a given density of dislocations:

$$e_{ikl} u^{\mathrm{el-plD}}_{j,lk} = e_{ikl} (\beta^{\mathrm{elD}}_{lj,k} + \beta^{\mathrm{plD}}_{lj,k}) = 0. \tag{8}$$

Inasmuch as β^{elD} and β^{plD} taken separately do not satisfy the compatibility condition, we designate them by the terms ''incompatible elastic'' and ''incompatible plastic'' distortion. For the symmetric part of β^{elD} and β^{plD} the compatibility conditions have the form

$$e_{ikn} e_{jlm} \beta^{\mathrm{elD}}_{(nm),kl} = -e_{ikn} e_{jlm} \beta^{\mathrm{plD}}_{(nm),kl} = -\eta_{ij}, \tag{9}$$

where η is the incompatibility tensor.

The quantities β^{elD}, which are associated with internal sources, i.e., with material defects, can be regarded as the internal elastic field. The incompatibility tensor η separates the elastic strain fields (4) and (9) into external and internal fields. The last term in Eq. (2), $\varepsilon^{\mathrm{pl}}$, represents compatible plastic deformation unrelated to stresses and characterizes the irreversible distortion of the material due to the annihilation of defects or their emergence to the surface. The components $\varepsilon^{\mathrm{pl}}$ satisfy the compatibility condition

$$e_{ikn} e_{jlm} \varepsilon^{\mathrm{pl}}_{nm,kl} = 0 \tag{10}$$

and characterize the deformation of the defect-free material. From the standpoint of the structure of the crystal lattice the nature of the strain components $\varepsilon^{\mathrm{el}}$, $\varepsilon^{\mathrm{el-plD}}$, and $\varepsilon^{\mathrm{pl}}$ is illustrated in Figs. 2, 3a, 3b, and 4.

On the $\sigma-\varepsilon$ diagram (Fig. 1) the elastic strain $\varepsilon^{\mathrm{el}}$ corresponds to the segment DC. The strain OD remaining in the material after elastic unloading does not vary with time in the mechanics of plasticity. Fracture processes occurring in the material for a certain period of time after removal of the external load and relaxation of the external elastic fields $\varepsilon^{\mathrm{el}}$

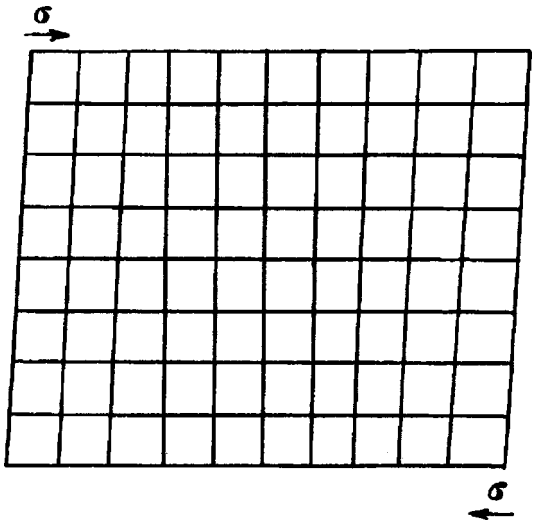


FIG. 2.

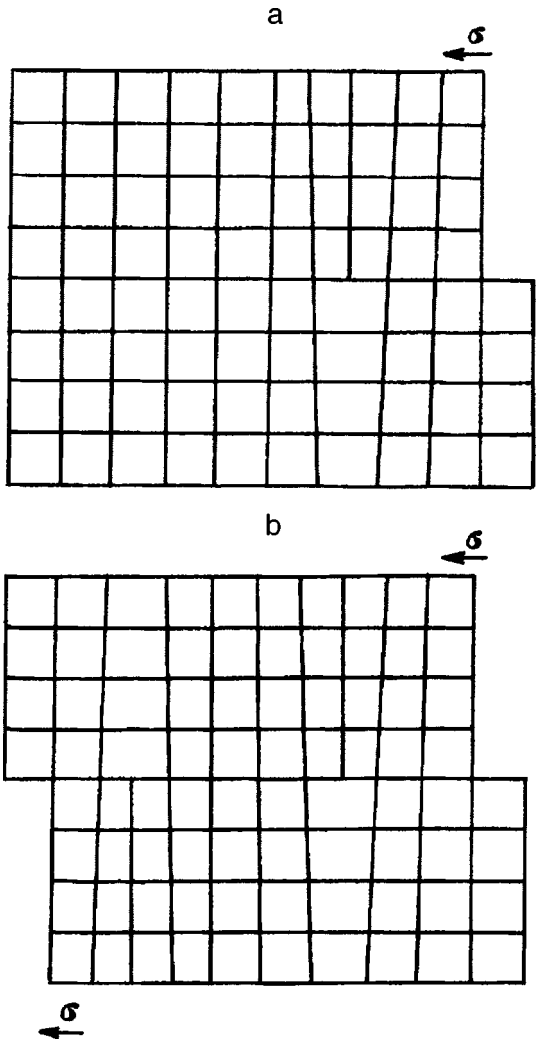


FIG. 3.

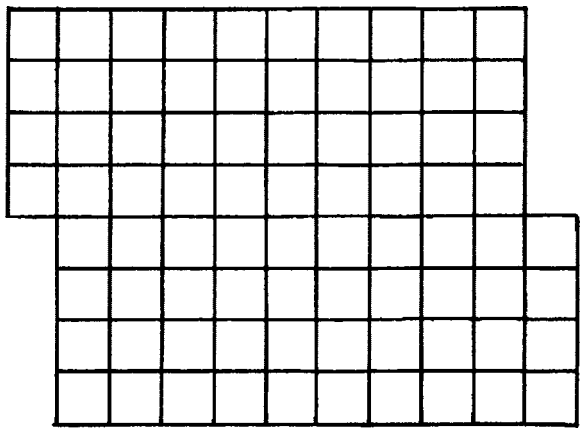


FIG. 4.

are left out of the picture in this case. Various softening mechanisms are known. The most thoroughly investigated processes are those associated with the reduction in the number of defects and will be the ones discussed below. The softening effect determines how the material makes the transition from the nonequilibrium state to the equilibrium state. The driving force behind this process is the decrease in the energy of deformation defects, as characterized by the quantity β^{elD}, owing to the restructuring of defects into low-energy configurations, annihilation, and departure into sinks (pores, free surfaces, etc.). In Ref. 9 softening is treated as a phenomenon of reverse mechanical aftereffect (RMA), and estimates of the strain due to RMA are given. Typical values of the RMA strain in torsion or bending are $10^{-4}-10^{-2}$. RMA effects are less pronounced in other loading techniques. For comparison the maximum elastic strain is tenths of one percent, since the elastic limit of pure metals is $(10-10^{2})$MPa and the Young's modulus is of the order of $(10^{4}-10^{5}$MPa.[10] It follows, therefore, that the residual strain OD contains an irreversible plastic part OE ($\varepsilon^{\mathrm{pl}}$) and another part ED that is potentially reversible in softening processes ($\varepsilon^{\mathrm{el-plD}}$). The material at point D after loading to point B and subsequent unloading exists in a nonequilibrium state, because the elastic energy of the internal fields is stored in it. The transition to RMA equilibrium takes place through relaxation of the incompatible elastic distortion ($\beta^{\mathrm{elD}}\to 0$ completely or partially). It follows from Eq. (8) that the incompatible plastic distortion β^{plD} becomes compatible by virtue of the relaxation of β^{elD}, i.e., is determined in the form of the gradient of the continuous displacement vector and belongs to the segment OE. The strain of the segment OE does not determine the state of the material, because the initial state (Fig. 2) and the final state (Fig. 4) are indistinguishable from the energy standpoint. The proposed scheme of separating the strain into components is idealized, because all the components (2) are in fact interrelated and would certainly not be amenable to separate experimental investigation. However, such a representation can be used to establish the interrelationship of $\varepsilon^{\mathrm{el}}$, $\varepsilon^{\mathrm{el-plD}}$, and $\varepsilon^{\mathrm{pl}}$, to interpret the Lagrangian of the gauge model or to attempt to write the Lagrangian of a medium with defects on the basis of physical considerations, and to obtain dynamical equations establishing the interaction between the external and internal elastic fields.

ALGORITHM FOR CONSTRUCTING THE GAUGE MODEL

It has been shown[1,4] that dynamical models of solids with dislocations, disclinations, and both types of defects together can be constructed on the basis of the nonlinear Lagrangian in the theory of elasticity. In this section we consider a linear model of an elastic solid with dislocations as a first approximation. Schematically the procedure for constructing the gauge model entails writing the Lagrangian for a homogeneous, isotropic elastic body

$$L=\int dv\left\{\frac{\rho}{2}\frac{\partial u_i}{\partial t}\frac{\partial u_i}{\partial t}-\frac{\mu}{2}\left(\frac{\partial u_i}{\partial x_j}\frac{\partial u_i}{\partial x_j}-\frac{\partial u_i}{\partial x_j}\frac{\partial u_j}{\partial x_i}\right)-\frac{\lambda}{2}\frac{\partial u_j}{\partial x_j}\frac{\partial u_i}{\partial x_i}\right\} \tag{11}$$

and determining its gauge group. In Eq. (11) u_i is the vector of elastic displacements, λ and μ are the Lamé constants, and ρ is the density of the material. The basic Lagrangian is invariant under homogeneous translations:

$$u_i(x,t)=u_i(x,t)+a_i; \tag{12}$$

this property corresponds to transport of the elastic body as a unit whole. The localization of the translation group

$$u_i(x,t)=u_i(x,t)+a_i(x,t) \tag{13}$$

breaks the invariance of (11) by the emergence of increments associated with differentiation of the parameters of the group $a_i(x,t)$. The minimal replacement procedure, which replaces the ordinary derivatives with covariant derivatives:

$$\frac{\partial u_i}{\partial x_j}\to D_j u_i=\frac{\partial u_i}{\partial x_j}+\beta_{ji},\qquad \frac{\partial u_i}{\partial t}\to D_0 u_i=\frac{\partial u_i}{\partial t}+v_i, \tag{14}$$

restores the invariance of (11) under the inhomogeneous transformations (13):

$$L_1=\int dv\left\{\frac{\rho}{2}D_0u_iD_0u_i-\frac{\mu}{2}(D_ju_iD_ju_i-D_ju_iD_iu_j)-\frac{\lambda}{2}D_ju_jD_iu_i\right\}. \tag{15}$$

The substitution (14) introduces new gauge or compensating fields β_{ij} and v_i; according to Ref. 1, they are related to the additional kinetic and potential energy governing the Lagrangian of the gauge fields

$$L_2=\int dv\left\{\frac{B}{2}I_{ij}I_{ij}-\frac{S}{2}\alpha_{ij}\alpha_{ij}\right\} \tag{16}$$

as a function of the quantities

$$I_{ij}=\frac{\partial v_j}{\partial x_i}-\frac{\partial \beta_{ij}}{\partial t} \tag{17}$$

$$\alpha_{ij}=e_{ikl}\frac{\partial}{\partial x_k}\beta_{lj}, \tag{18}$$

where B and S are new constants of the theory.

The procedure for constructing the gauge model (11)–(18), which is described more in detail in Refs. 1 and 4, does not elucidate the physical significance of the model potentials β, v, and u. To solve this problem, we rely on a scheme of representation of the total strain within the framework of the continuum theory of defects (2), since the localization of the translation group at a point (13) corresponds to the introduction of a single Volterra dislocation, and the functional dependence on the coordinates determines the averaged distribution of defects.[11,12]

PHYSICAL MEANING OF THE GAUGE FIELDS

According to Eqs. (2) and (6), the elastic deformations in a material with defects (14) are determined by the reversible elastic distortion associated with external loads and by the incompatible elastic distortion due to defects of the material:

$$D_j u_i = u^{\mathrm{el}}_{i,j} + \beta^{\mathrm{elD}}_{ji}. \tag{19}$$

The total displacement velocity, which determines the kinetic energy L_1, can be represented by the sum of the velocities of the elastic displacements and displacements due to material defects:

$$D_0 u_i = \frac{\partial u^{\mathrm{el}}_i}{\partial t} + v_i, \quad \text{where} \quad v_i = \frac{\partial}{\partial t} u^{\mathrm{el-plD}}_i. \tag{20}$$

The velocity v^{pl}, calculated as the time derivative of the compatible plastic displacements u^{pl}, does not contribute to the kinetic energy L_1, because the increment of this quantity depends on relaxation processes in the system of defects associated with the segment ED in Fig. 1. If we assume that v^{pl} is the term $D_0 u_i$, we must then assume that under the influence of an external load all atoms in the glide plane over the length of the macrosample simultaneously undergo a shift by a vector that is a multiple of the lattice period. In Cottrell's opinion,[13] this assumption is inadmissible in solids, even though such processes can certainly take place in liquids and in media of the Plasticine (modeling clay) type.

For the above-determined values of the potentials u^{el}, β^{elD}, and v the Lagrangian L_2 of the gauge fields is described by the dislocation density tensor

$$\alpha_{ij} = e_{ikl} \beta^{\mathrm{elD}}_{lj,k} \tag{21}$$

and the dislocation flux density I_{ij}. By definition,

$$I = -\frac{\partial \beta^{\mathrm{plD}}}{\partial t}. \tag{22}$$

The dislocation flux density tensor can be expressed in terms of the incompatible elastic distortion tensor and the velocity associated with material defects by straightforward transformations based on the proposed scheme (2):

$$I_{ij} = -\frac{\partial}{\partial t}(u^{\mathrm{el-plD}}_{j,i} - \beta^{\mathrm{elD}}_{ij}) = \frac{\partial \beta^{\mathrm{elD}}_{ij}}{\partial t} - \frac{\partial v_j}{\partial x_i}. \tag{23}$$

We now attempt to justify the Lagrangian of the gauge model from the physical point of view. It is clear from an analysis of the components of the total deformation (Fig. 1) that the state of the material at point B can be described on the basis of the model of a mixture of two continua: an elastic continuum and a continuum of defects. In the elastic continuum the effective field (19) is formed as the superposition of the external and internal fields. The potential energy of the effective fields corresponds to the second term L_1 (15). The kinetic energy of material particles depends on the velocity (20). The state of the continuum of defects is characterized by the dislocation density tensor (21) and the dislocation flux density (23), whose homogeneous quadratic functions determine the Lagrangian L_2 (16). Since the increments of the compatible plastic displacements do not contribute to the potential and kinetic energies, the total Lagrangian of the gauge model ($L = L_1 + L_2$) can be expressed in terms of the displacement vector $u^t = u^{\mathrm{el}} + u^{\mathrm{el-}plD}$ and the distortion tensor β^{plD}:

$$D_0 u_i = \frac{\partial u^t_i}{\partial t}, \quad D_j u_i = u^t_{i,j} - \beta^{\mathrm{plD}}_{ji},$$

$$\alpha_{ij} = -e_{ikl} \beta^{\mathrm{plD}}_{lj,k}, \quad I_{ij} = -\frac{\partial \beta^{\mathrm{plD}}_{ij}}{\partial t}. \tag{24}$$

A standard technique is used to deduce the dynamical equations of the model from the condition of time invariance of the action integral:

$$\rho \frac{\partial^2 u_i}{\partial t^2} - \lambda \frac{\partial^2 u_j}{\partial x_i \partial x_j} - \mu \left(\frac{\partial^2 u_i}{\partial x_j \partial x_j} + \frac{\partial^2 u_j}{\partial x_i \partial x_j} \right)$$

$$+ \mu \left(\frac{\partial \beta_{ij}}{\partial x_j} + \frac{\partial \beta_{ji}}{\partial x_j} \right) + \lambda \frac{\partial \beta_{kk}}{\partial x_i} = 0,$$

$$B \frac{\partial^2 \beta_{ij}}{\partial t^2} - S \left(\frac{\partial^2 \beta_{ij}}{\partial x_k \partial x_k} - \frac{\partial^2 \beta_{kj}}{\partial x_k \partial x_i} \right) - \lambda \frac{\partial u_k}{\partial x_k} \delta_{ij}$$

$$- \mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) + \mu (\beta_{ij} + \beta_{ji}) - \lambda \beta_{kk} \delta_{ij} = 0,$$

in which the superscripts are dropped from u^t and β^{plD}.

CONCLUSION

For any choice of independent variables the system of equations of motion of the gauge model describes the dynamics of an elastic solid with internal stresses. The object of the gauge description is best exemplified in relations (19)–(23), where the elastic quantities u^{el} and β^{elD} are treated as independent variables. If the energy dissipation associated with the relaxation of external and internal stresses is taken into account phenomenological as proposed, for example, in Refs. 5 and 6, we obtain a model that describes the viscoelastic properties of materials. Processes of accumulation of compatible plastic deformation due to the annihilation of defects or departure into sinks are not included in the given theory, despite the need to determine these relations in constructing a physical theory of plasticity. The dynamical equations of the gauge model can be used to analyze the inelastic behavior of materials on the premise that in general the three strain components (2) exist only if the compatible plastic deformation is insignificant. The necessary conditions pre-

vail in shock-loading processes, which do not allow enough time for many defects to reach the surface, so that compatible plastic deformation can set in.

[1] A. Kadic and D. G. B. Edelen, *A Gauge Theory of Dislocations and Disclinations* (*Lectures Notes in Physics, Vol. 174*) (Springer-Verlag, Heidelberg, 1983), 168 pp.
[2] D. G. B. Edelen and D. C. Lagoudas, *A Gauge Theory and Defects in Solids* (North-Holland, Amsterdam, 1988), 189 pp.
[3] D. G. B. Edelen and D. C. Lagoudas, Int. J. Eng. Sci. **26**, 837 (1988).
[4] Yu. V. Grinyaev and N. V. Chertova, Izv. Vyssh. Uchebn. Zaved. Fiz. No. **2**, 34 (1990).
[5] V. L. Popov and N. V. Chertova, Izv. Vyssh. Uchebn. Zaved. Fiz. No. **4**, 81 (1992).
[6] N. V. Chertova, Int. J. Eng. Sci. **33**, 1315 (1988).
[7] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity*, 4th ed. (Cambridge University Press, Cambridge; Macmillan, London, 1927; Reprint, Dover, New York, 1944) [ONTI, Moscow-Leningrad, 1935].
[8] P. Perzyna, *Fundamental Problems of Viscoplasticity* [Mir, Moscow, 1968, 176 pp.].
[9] A. B. Volyntsev, *Hereditary Mechanics of Dislocation Ensembles* [in Russian] (Izd. Irkutsk. Univ., Irkutsk, 1984), 288 pp.
[10] V. A. Zaĭmovskiĭ and T. L. Kolupaeva, *Extraordinary Properties of Ordinary Metals* [in Russian] (Nauka, Moscow, 1984), 191 pp.
[11] J. D. Eshelby, ''Continuum theory of lattice defects'' in *Solid State Physics: Advances in Research and Applications*, Vol. 3 [Academic Press, New York, 1956; IL, Moscow, 1963, 268 pp.].
[12] R. De Wit, *Continuum Theory of Dislocations* [Mir, Moscow, 1977, 208 pp.].
[13] A. H. Cottrell, *Theory of Crystal Dislocations* [Gordon and Breach, New York, 1964; Mir, Moscow, 1969].

Translated by James S. Wood

Piezoelectric properties of oriented Z' cuts of PZT-type ferroelectric ceramics

M. G. Minchina and V. P. Dudkevich†

Rostov State University, 344091 Rostov-on-Don, Russia
(Submitted January 15, 1997)
Zh. Tekh. Fiz. **68**, 75–79 (July 1998)

The piezoelectric and dielectric properties of oriented $\beta°Z'$ cuts ($\beta=0$, 15, 30, 45, 60°) of piezoelectric crystals of TsTS-83G (lead zirconate titanate) composite are studied. A static model is proposed for the case of a maximally polarized ceramic based on the conditions of complete and partial stability of the polar axes **c** with allowance for their nonuniform distribution for 180° and 90° domain reorientations. It is found that the TsTS-83G piezoelectric composite does not exhibit anisotropy in the piezoelectric coefficient d'_{33} when the axis of the Z' cut is rotated in the ZY plane relative to the XYZ coordinate system. © *1998 American Institute of Physics.* [S1063-7842(98)01307-5]

The piezoelectric properties of ferroelectric ceramics are known to depend on their orientation relative to the principal crystallographic axes.[1] There are a number of piezoelectric crystals and textures of the $\infty \cdot m$ group whose piezoelectric coefficient d_{33}, which relates the polarization vector directed along the polar axis to the longitudinal strain along the same axis in the principal crystallographic coordinate system XYZ, does not have the maximum value.[2] A new coordinate system $X'Y'Z'$ (related to the old XYZ system in a certain way) can be found in which the piezoelectric coefficient d'_{33} is maximized. This effect, referred to as the anisotropy of the piezoelectric coefficient d'_{33}, is observed in piezoelectric ceramics of the $4mm$ group ($Pb(Ti_{0.48}Zr_{0.52})O_3$, $Na_{0.5}K_{0.5}NbO_3$, PZT-2, etc.) and can be determined experimentally with the help of oriented cuts. The solid solution with the composition $Pb(Ti_{0.48}Zr_{0.52})O_3$ is of special interest, since, in the piezoelectric state, it possesses anisotropy of the stiffness coefficient C_{11} (Ref. 3) as well as anisotropy of the piezoelectric coefficient d_{33}.

This paper is devoted to an analysis of the behavior of the piezoelectric coefficient d'_{33} of oriented Z' cuts with respect to the principal crystallographic coordinate system XYZ of the solid solution TsTS-83G (lead zirconate titanate), which is close in composition to the solid solution $Pb(Ti_{0.48}Zr_{0.52})O_3$, taking into account the distribution of the polar axes of the domains for the case in which the ceramic is maximally polarized.

EXPERIMENTAL TECHNIQUE

A ferroelectric ceramic with the TsTS-83G composition was obtained by hot pressing with a load of 90 kg/cm^2 at $T=1100$ °C for 5 h, in the form of a block of diameter 100 mm and height 12 mm. Electrodes were deposited by brazing a silver-containing paste. The block was poled in silicone oil at $T=120$ °C for 1 h at $E=25$ kV/cm. Oriented $\beta°Z'$-cuts ($\beta=0$, 15, 30, 45, and 60°) were cut with a diamond wheel from the poled block with strict maintenance of the orientations of the XYZ and $X'Y'Z'$ coordinate systems and the direction of the poling field **E** (Fig. 1). Samples of size $4\times4\times4$ mm were cut from each oriented $\beta°Z'$-cut for measuring the piezoelectric coefficients d_{33} in the quasistatic regime. Samples of sizes $12\times2\times2$, $12\times6\times6$, and $6\times6\times0.4$ mm, respectively, were cut from the $0Z'$ cut for the dynamic measurements of the piezoelectric coefficients d_{33}, d_{31}, and d_{15}. Electrodes were deposited on the oriented poled samples by cathode sputtering from Al+Cr at $T=80$ °C for 30 min. Dielectric measurements at a frequency of 1 kHz were made on a MOST E8-2 system.

For the x-ray structural analysis we used unpoled samples in the form of $1\times1\times0.1$ cm slabs that had been initially polished and annealed at $T=600$ °C for 3 h.

RESULTS AND DISCUSSION

Let us consider a ferroelectric ceramic of class $4mm$ in a principal crystallographic coordinate system XYZ where the polarization vector **P** coincides in direction with the poling

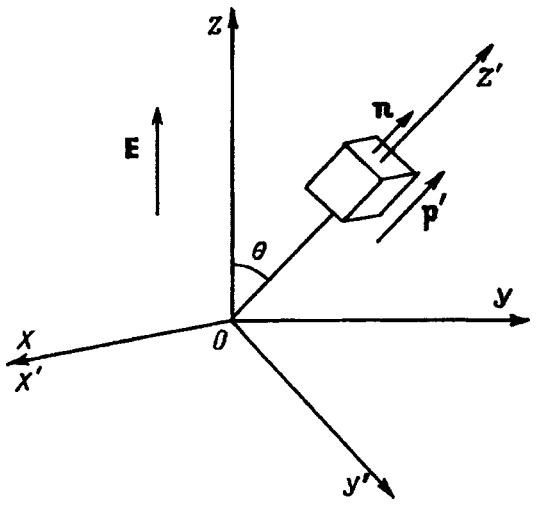


FIG. 1. The transformation from the principal crystallographic coordinate system XYZ to the coordinate system $X'Y'Z'$. Θ is the angle between the Z and Z' axes.

TABLE I. Theoretical values of the piezoelectric coefficients d'_{33} for different angles Θ in $Pb(Ti_{0.48}Zr_{0.52})O_3$ and TsTS-83G ceramics.

Piezoceramic composition	0	10°	20°	30°	40°	50°	60°	70°	80°	90°
$Pb(Ti_{0.48}Zr_{0.52})O_3$	223	225	229	232	227	210	178	129	69	0
TsTS-83G	395	387	364	327	281	229	172	115	57	0

field **E**. We cut from this ceramic a slab whose edges are parallel to the $X'Y'Z'$ axes and apply to this slab a uniaxial mechanical stress of the form (Fig. 1)

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix}.$$

The equation for the direct piezoelectric effect has the form

$$P_i = d_{ijk}\sigma_{jk}, \tag{1}$$

where P_i is the polarization vector, d_{ijk} are the piezoelectric coefficients, which form a tensor of the third rank, and σ_{jk} are the mechanical stresses.

On changing from one coordinate system to another, the piezoelectric coefficient d'_{33} transforms according to

$$d'_{ijk} = a_{im}a_{jn}a_{kl}d_{mnl}, \tag{2}$$

where a_{im}, a_{jn}, and a_{kl} are the direction cosines relating the coordinate systems XYZ and $X'Y'Z'$.

The matrix of piezoelectric coefficients for a ceramic in the $4mm$ group has the form

$$\begin{pmatrix} 0 & 0 & 0 & 0 & d_{15} & 0 \\ 0 & 0 & 0 & d_{15} & 0 & 0 \\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 \end{pmatrix}. \tag{3}$$

On substituting Eq. (3) into Eq. (2), we obtain an equation for the piezoelectric coefficient d'_{33},

$$d'_{33} = (d_{15}+d_{31})\cos\Theta \sin^2\Theta + d_{33}\cos^3\Theta, \tag{4}$$

where Θ is the angle between the Z and Z' axes.

An examination of Eq. (4) at its extremal values shows that there is a maximum value of d'_{33} which differs from d_{33} and is given by

$$d'_{33} = \frac{2}{3}(d_{15}+d_{31})\cos\Theta, \tag{5}$$

where

$$\cos\Theta = \sqrt{\frac{d_{15}+d_{31}}{3d_{15}+3d_{31}-3d_{33}}}.$$

Table I lists values of $d'_{33} = d'_{33}(\Theta)$ calculated according to Eq. (4) for $Pb(Ti_{0.48}Zr_{0.52})O_3$ ($d_{33}=223$, $d_{31}=-93.5$, $d_{15}=494$) and TsTS-83G ($d_{33}=394.8$, $d_{31}=-180.9$, $d_{15}=508.2$) solid solutions. The values in parentheses are given in pC/N.

In accordance with Table I, for the $Pb(Ti_{0.48}Zr_{0.52})O_3$ solid solution the piezoelectric coefficient d'_{33} reaches its highest value $d'_{33\,\max}=231.6$ pC/N at an angle of $\Theta = 29.86°$. For the TsTS-83G solid solution the piezoelectric coefficient d'_{33} exhibits no anisotropy as the Z' axis is rotated in the ZY plane relative to the XYZ coordinate system.

In order to study the behavior of the piezoelectric coefficient d'_{33} of the TsTS-83G ceramic, we cut oriented $\beta°Z'$-cut ($\beta=0$, 15, 30, 45, and 60°) samples in which the normals to the faces of the cut cube are parallel to the $X'Y'Z'$ axes (Fig. 1), so that the angle between the Z' and Z axes equals Θ. The oriented cuts were cut from a single poled block in a way such that when the axis of the Z' cut is rotated by an angle Θ in the ZY plane, no changes occur in the distribution of the polar angles of the domains (in the block and in the cuts).

The composition of TsTS-83G ferroelectric ceramic belongs to the tetragonal (T) boundary of the morphotropic transition region. According to x-ray structural data, TsTS-83G ferroelectric ceramic contains about 70% of the T phase (unit cell parameters $a_T=4.0402$ Å, $c_T=4.1311$ Å) and 30% of the rhombohedral (R) phase ($a_R=4.0722$ Å and $\alpha_R=89.783°$).

The proposed static model of the ceramic includes all the possible domain reorientations that correspond only to the T phase for the case of the maximally polarized ceramic.

Some theoretical and experimental values of the piezoelectric coefficient d'_{33} of the $\beta°Z'$ cuts are listed in Table II.

The theoretical values $d'^{\,\text{theor I}}_{33}$ were obtained by transforming the piezoelectric coefficient d_{33} from the principal crystallographic coordinate system XYZ to an arbitrary sys-

TABLE II. Piezoelectric coefficient d'_{33} and relative permittivity $\varepsilon'^{\,T}_{33}/\varepsilon_0$ of $\beta°Z'$ cuts of TsTS-83G piezoelectric ceramic

$\beta°Z'$ cuts	0	15°	30°	45°	60°
$d'^{\,\text{exp}}_{33}$	395	389	350	296	213
$d'^{\,\text{theor 1}}_{33}$	395	377	327	255	172
$d'^{\,\text{theor II}}_{33}$	395	379	341	266	191
$\varepsilon'^{\,T\,\text{exp}}_{33}/\varepsilon_0$	1874	1854	1771	1639	1479
$\varepsilon'^{\,T\text{theor}}_{33}/\varepsilon_0$	1790	1770	1714	1639	1563

tem $X'Y'Z'$ using Eq. (4) without taking into account the internal structure of the ceramic. The coefficients d_{33}, d_{31}, and d_{15} are the basic experimentally determined piezoelectric coefficients of the TsTS-83G ceramic in the XYZ coordinate system. Equation (4) can also be used to examine the relation between the parameters of the poled ceramic and those of the crystallites (assuming they are made up of single domains) on the basis of a statistical averaging of the physical constants.[4,5]

In the second case, the $d'^{\text{theor II}}_{33}$ were obtained on the basis of a static model of the ceramic by averaging the physical constants of individual crystallites (domains) with the distribution of the polar axes of the domains for 180° and 90° switchings in the case of a maximally polarized ceramic. In describing this model, we shall ignore the interaction among the crystallites (domains), domain wall movement, and the growth and nucleation of new domains. We shall assume that the internal mechanical stresses which develop in the ceramic during sintering do not change under external forces and that only the field applied to the ceramic acts on each crystallite.

In this model the poled ceramic was represented as a system of N saturation-polarized single-domain crystallites, whose polar axes have a predominant orientation and lie along those allowed directions of the spontaneous polarization vector **P** in the crystallites which are closest to the direction of the poling field **E**.[6] In a geometric-statistical fashion, we shall consider the poled ceramic to be a sphere of unit radius, whose surface is covered with varying densities by the ends of the polar axes of the domains, which start from the center of this sphere. An examination of the case of a maximally polarized ceramic shows that all the polar axes of the domains are distributed in the upper half of the orientation sphere, within a solid angle $2\pi(1-\cos\Theta)$ about the direction of the field. Consideration of all the possible 180° and 90° reorientations shows that the domains whose polar axes are the farthest from the direction of the poling field **E** are polarized at an angle $\Theta = 54°44'$ to the field direction. The conditions for a reorientation of the polar axes by 180° and 90° were obtained from the conditions for complete and partial stability of the **c** axes relative to the direction of the poling field **E** (Fig. 2) with allowance for the geometric relation among the angles γ_c, γ_a, and ψ,[7]

$$\cos\gamma_a = \sin\gamma_c \cos\psi. \tag{6}$$

For the region in which the polar axes of the crystallites (domains) are completely stable, the following condition holds:

$$\cos\gamma_c - \cos\gamma_a \geqslant 0. \tag{7}$$

Substituting condition (6) into Eq. (7), we obtain

$$\cos\gamma - \sin\gamma_c \cos\psi \geqslant 0. \tag{8}$$

The **a** axes of all the crystallites with given γ_c in the poled state of the ceramic are uniformly distributed over the circle P. The ones closer to the field **E** are concentrated in a quarter of this circle ($\psi = -45°$ to $\psi = 45°$). Let us consider those axes **a** which are closest to the direction of the poling field ($\psi = 0$), in which case, according to condition (8), the zone of complete stability of the **c** axes will lie in the region

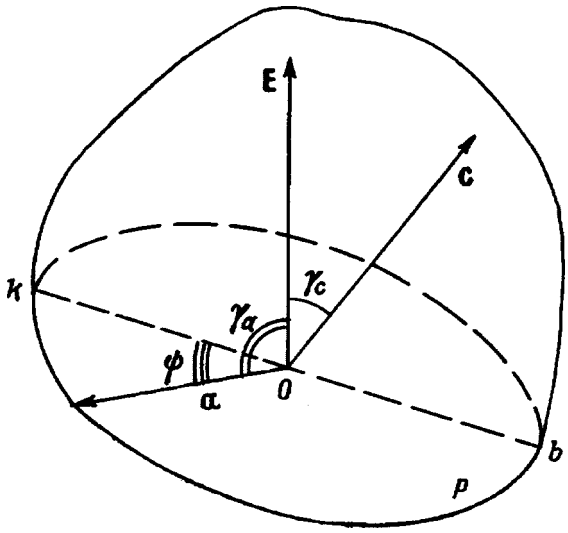


FIG. 2. Notation for the angles characterizing the position of a domain relative to the poling field **E**.

$$0 \leqslant \gamma_c \leqslant 45°. \tag{9}$$

Including the 180° reversals of the **c** axes ($\rho' = N/2\pi$) and the 90° switchings of the **a** axes into **c** axes ($\rho'' = N/\pi$), one finds that the density of **c** axes in this zone will be

$$\rho_c^1 = \frac{3N}{2\pi}. \tag{10}$$

We shall refer to the region in which, consistent with condition (7), the **a** axes of the crystallites are farthest from the direction of the poling field **E** ($\psi = 45°$) as the zone of partial stability of the **c** axes for the case of a maximally polarized ceramic. Then, according to Eq. (8), the zone of partial stability of the polar axes **c** lies within the range of angles

$$45 \leqslant \gamma_c \leqslant 54°44'. \tag{11}$$

In this region we shall also take into account the fraction of **c** axes that are converted to **a** axes after a 90° switching, i.e.,

$$\cos\gamma_a - \cos\gamma_c \geqslant 0. \tag{12}$$

Substituting Eq. (6) into Eq. (12), we find the angle $|\psi|$ within which lie the **c** axes that are closer to the field **E**,

$$|\psi| \leqslant \arccos(\cot\gamma_c). \tag{13}$$

The fraction crystallites whose **c** axes move out of the region upon a 90° switching will be given by β_c, where

$$\beta_c = \frac{2|\psi|}{\pi/2} = \frac{4\arccos(\cot\gamma_c)}{\pi}. \tag{14}$$

The fraction of crystallites whose **c** axes remain after the switching will be given by

$$\beta_c' = 1 - \frac{4\arccos(\cot\gamma_c)}{\pi}. \tag{15}$$

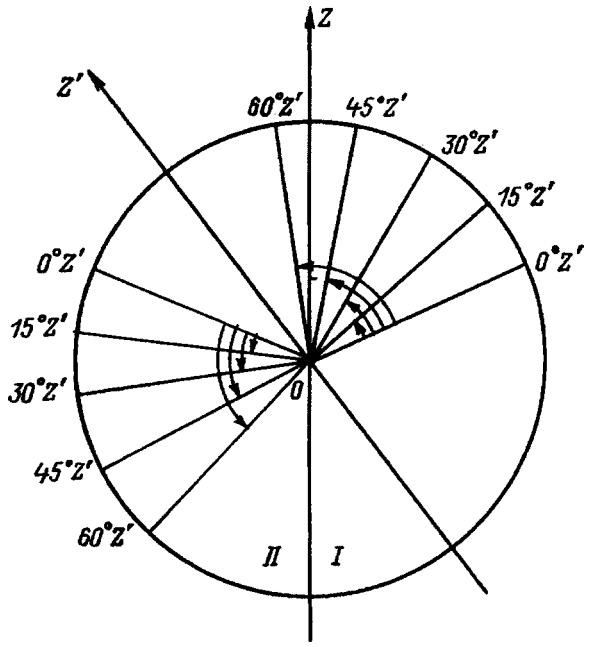


FIG. 3. Angular boundaries of the polar **c** axes for $\beta°Z'$ cuts. First hemisphere: $0°Z'$, $54°44'$; $15°Z'$, $39°44'$; $30°Z'$, $24°44'$; $45°Z'$, $9°44'$. Second hemisphere: $60°Z'$, $6°44'$; $0°Z'$, $54°44'$; $15°Z'$, $69°44'$; $30°Z'$, $84°44'$; $45°Z'$, $99°44'$; $60°Z'$, $114°44'$.

We shall assume that in the zone of partial stability of the **c** axes, for the **a** axes which lie within the circle P and are closer to the field **E**, a fraction β_c (Eq. (14)) will leave during a 90° switching. After all the possible switchings, the density of **c** axes in this zone will be equal to

$$\rho_c^{\mathrm{II}}=\frac{3N}{2\pi}\left(1-\frac{4\arccos(\cot\gamma_c)}{\pi}\right). \tag{16}$$

The distribution of the polar **c** axes examined here using this model refers to a $0°Z'$ cut.

On going to $\beta°Z'$ cuts ($\beta=0$, 15, 30, 45, 60°), according to the static model, for the case of a maximally polarized ceramic there is a region containing only **a** axes and in which only 90° switchings of **a** axes to **c** axes can occur (Fig. 3). In calculating the piezoelectric coefficient d'_{33}, we shall include only that fraction of the reorientations for which the **a** axes, after a 90° switching, become **c** axes. This fraction is given by

$$\beta''_c=\frac{4\arccos(\cot\gamma_c)}{\pi}. \tag{17}$$

Here we have calculated the piezoelectric coefficients d'_{33} of the $\beta°Z'$ cuts in the principal crystallographic coordinate system XYZ. The angular boundaries of the polar **c** axes for $\beta°Z'$ cuts are shown in Fig. 3. The piezoelectric coefficient $d'^{\,\mathrm{theor\ II}}_{33}$ was calculated by integrating over the orientation sphere with allowance for the distribution of polar **c** axes and averaging over all the crystallites (domains). The equation used to calculate the piezoelectric coefficient $d'^{\,\mathrm{theor\ II}}_{33}$ for the $0°Z'$ cut has the form

$$d'^{\,\mathrm{theor\ II}}_{33}=\frac{1}{N}\int_{\Theta=0°}^{45°}\int_{\phi=0°}^{2\pi} d_{33}(\Theta)\,\frac{3N}{2\pi}\sin\Theta\,d\Theta\,d\phi + \frac{1}{N}\int_{\Theta=45°}^{54°44'}\int_{\phi=0°}^{2\pi} d_{33}(\Theta)\,\frac{3N}{2\pi} \times\left(1-\frac{4\arccos(\cot\gamma_c)}{\pi}\right)\sin\Theta\,d\Theta\,d\phi. \tag{18}$$

For the $\beta°Z'$ cut, the piezoelectric coefficient $d'^{\,\mathrm{theor\ II}}_{33}$ is calculated as follows:

$$d'^{\,\mathrm{theor\ II}}_{33}=d'^{\,\beta}_{33}+\frac{1}{N}\int_{\Theta=54°44'}^{54°44'+\beta°}\int_{\phi=\pi}^{2\pi} d_{33}(\Theta)\,\frac{N}{\pi} \times\frac{4\arccos(\cot\gamma_c)}{\pi}\sin\Theta\,d\Theta\,d\phi, \tag{19}$$

where $d'^{\,\beta}_{33}$ is calculating using Eq. (18) with the angular boundaries corresponding to the $\beta°Z'$ cuts (Fig. 3).

The discrepancy between $d'^{\,\mathrm{exp}}_{33}$ and $d'^{\,\mathrm{theor\ II}}_{33}$ can be explained by the fact that we have neglected the 55° (and 109°) reorientations of the polar **c** axes of the R phase.

The experimental data on the relative dielectric permittivity ($\varepsilon'^{\,T}_{33}/\varepsilon_0$) of the oriented $\beta°Z'$ cuts are also in good agreement with the theoretical results calculated according to the formula

$$\varepsilon'^{\,T}_{33}=\varepsilon^T_{11}\sin^2\Theta+\varepsilon^T_{33}\cos^2\Theta, \tag{20}$$

where the relative permittivities $\varepsilon^T_{11}/\varepsilon_0$ and $\varepsilon^T_{33}/\varepsilon_0$ of the TsTS-83G ceramic are 1487 and 1790, respectively.

CONCLUSIONS

1. We have proposed a static model for the case of a maximally polarized ceramic that is based on the conditions for complete and partial stability of the polar **c** axes, including their nonuniform distribution for 180° and 90° domain reorientations.

2. It has been shown that the reduction in the piezoelectric coefficient d'_{33} of oriented $\beta°Z'$ cuts of TsTS-83G ceramic is primarily caused by 90° domain reorientations.

3. TsTS-83G piezoelectric ceramic does not exhibit anisotropy of the piezoelectric coefficient d'_{33} as the axis of the Z' cut is rotated in the ZY plane relative to the XYZ coordinate system.

†Deceased

[1] W. Cady, *Piezoelectricity* [McGraw-Hill, New York (1946); IL, Moscow (1946)].
[2] A. E. Mattiat (ed), *Ultrasonic Transducer Materials*, Plenum Press, New York (1971), 103 p.
[3] Zhiwu Wang and Rishi Raj, Am. Ceram. Soc. Bull. **74**, 3363 (1995).
[4] S. V. Bogdanov, B. M. Vul, and A. M. Timonin, Izv. AN SSSR Ser. Fiz. **21**, 374 (1957).
[5] Yu. S. Dudek, M. G. Radchenko, A. V. Turik *et al.*, ''Orientational polarization of ferro(piezo)ceramics'' [in Russian], RGU, Rostov-on-Don (1983), Deposited in VINITI, No. 3478-83, 57 pp.
[6] B. Yaffe, W. Cooke, and H. Yaffe, *Piezoelectric Ceramics* [Academic Press, New York (1971); Mir, Moscow (1974)].
[7] E. G. Fesenko (ed), *Polarization of Piezoelectric Ceramics* [in Russian], Izd. RGU, Rostov-on-Don (1962).

Translated by D. H. McNeill

Nuclear pumping of a carbon monoxide laser

V. A. Zherebtsov

A. I. Leĭpunskiĭ Institute of Physics and Power Engineering, State Scientific Center of the Russian Federation, 249020 Obninsk, Russia

(Submitted March 13, 1997)
Zh. Tekh. Fiz. **68**, 80–85 (July 1998)

It is shown that plasma-chemical processes involving ionized and excited particles can make the main contribution to the pumping of energy into vibrations of carbon monoxide molecules. It is noted that the use of helium as a buffer gas in the active laser medium is not optimal. The employment of argon instead of helium permits a 1.5-fold increase in the efficiency of the pumping of energy into carbon monoxide molecules and an order-of-magnitude decrease in the threshold energy for pumping the active medium. © *1998 American Institute of Physics.* [S1063-7842(98)01407-X]

1. The nuclear pumping of lasers is one of the promising ways to directly convert nuclear energy into laser radiant energy. Nuclear-pumped lasers on several tens of atomic, molecular, and ionic transitions have been created. Fairly high (3–5%) efficiencies have been achieved in converting the energy imparted to the active medium into laser radiant energy. However, a considerable increase in the efficiency of nuclear-pumped lasers is needed to expand the areas of their possible application in technology and power production. For this reason, one of the main problems in the physics of nuclear-pumped lasers is to find and study efficient active media for such lasers.

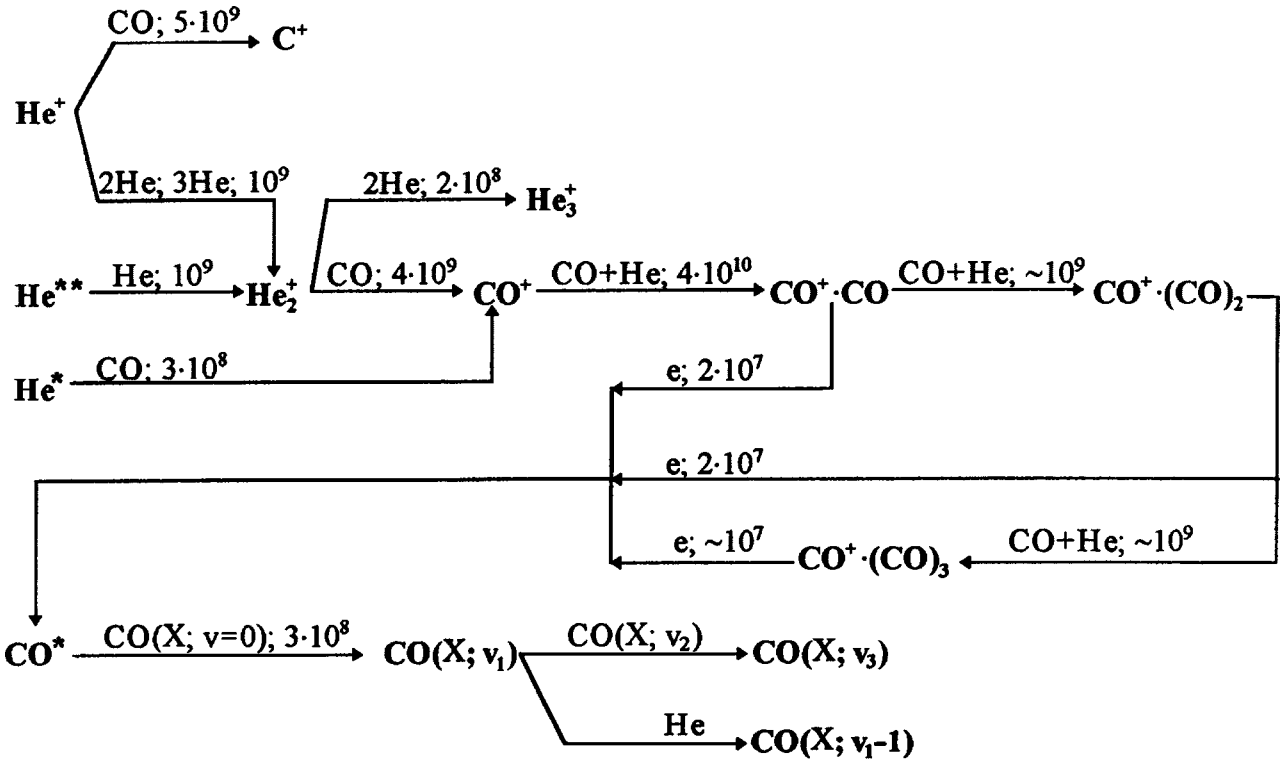
As an analysis of the mechanism underlying the operation of CO lasers has shown,[1] one of the promising active media for efficient nuclear pumping is a medium based on carbon monoxide. Its promising properties are due primarily to the fact that, unlike the active media for electronic transition lasers, the active medium of a CO laser does not require highly selective pumping. It is only important that the energy fall within the broad band of vibrational levels of the ground electronic state. Furthermore, because of the autonomous nature of the vibrational subsystem and the anharmonicity of the vibrations of carbon monoxide molecules, this energy is redistributed at a fairly low translational temperature through the exchange of vibrational quanta between molecules so that a state with complete or partial population inversion is achieved. This property of the active medium of a CO laser makes it possible, in principle, to direct a considerable portion of the energy released in the medium by fast ions produced in nuclear reactions into laser levels. The experimental results on the volume nuclear pumping of a CO laser obtained in Ref. 2 indicate that it is realistic to expect a solution for this problem. However, we cannot say that a concrete solution has been found.

The pumping of a pure carbon monoxide laser by exciting it with hard sources (fast ions, electrons) was investigated theoretically in Ref. 3. It was assumed that only electrons formed as a result of the ionization of carbon monoxide molecules participate in the excitation of their vibrations. The larger remaining portion of the energy imparted to the medium by the fast particles, which was concentrated in excited electronic states of the molecules and in the potential energy of electron–ion pairs, was assumed to be unavailable for exciting vibrations in the molecules. However, under certain conditions part of this energy can be directed into vibrational degrees of freedom of the molecules through plasma-chemical processes involving ionized and excited particles. This possibility was pointed out in Refs. 2 and 4. However, these processes have not been examined in detail in the active medium of a nuclear-pumped CO laser.

2. The process of the nuclear pumping of a CO laser can be divided into three stages: 1) ionization and excitation of the buffer gas by fast ions; 2) plasma-chemical relaxation, including ion-molecule reactions, ion–electron recombination, quenching of electronically excited particles by molecules with the transfer of part of the excitation energy into vibrational degrees of freedom of the molecules, and the excitation of molecular vibrations by electrons; 3) vibrational relaxation, which leads to redistribution of the energy over the vibrational spectrum and the formation of a distribution of carbon monoxide molecules with complete or partial population inversion.

Let us examine these stages and concentrate our attention on the plasma-chemical relaxation stage. We first consider an active medium containing helium as the buffer gas and carbon monoxide. Such a medium was investigated experimentally in Ref. 2. To estimate the rates of the processes occurring in the active medium we choose the following parameters for the medium, which are close to the optimum values found in Ref. 2: a gas temperature (translational) $T_g = 200$ K, a helium density $n_{\mathrm{He}} = 5 \times 10^{19}$ cm^{-3}, a carbon monoxide density $n_{\mathrm{CO}} = 2.8 \times 10^{18}$ cm^{-3}, and a specific power deposition in the medium $W = 400$ W/cm^3.

When a fast ion passes through a He–CO medium, its energy is expended mainly on the ionization and excitation of helium atoms. During this process, an energy $E_p \approx 45$ eV is expended to form an electron–ion pair. We neglect the energy expended on the direct ionization and excitation of carbon monoxide in view of its small concentration in comparison to the concentration of helium. The helium

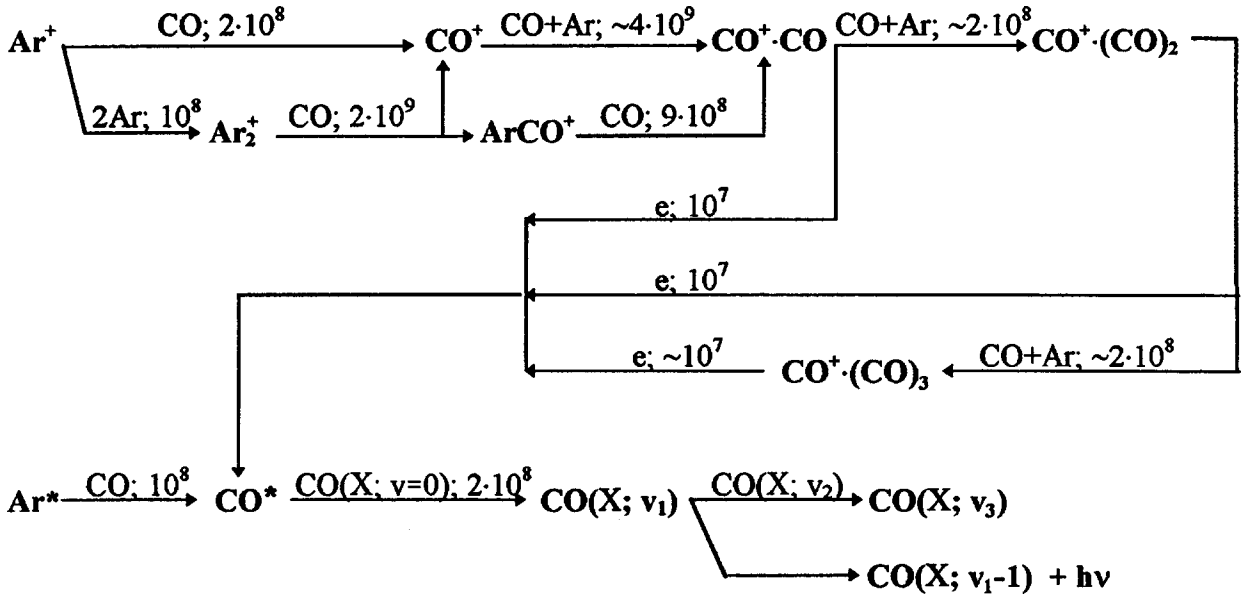


(1)

ions formed as a result of the passage of a fast ion, excited helium atoms, and electrons react with components of the active medium. The principal series of such reactions are shown in formula (1).

The reaction partners and the reaction rates in s^{-1} are shown over the arrows for the conditions considered. The reaction rate constants from Refs. 5–12 were used for the calculations.

The main process involving He^+ ions is charge exchange with CO molecules, which leads to dissociation of the latter into C^+ and O. The C^+ ions do not participate in any further chemical reactions in the medium under consideration and recombine with electrons in the volume or on the wall of the cell. As a result, the bulk of the energy of the fast ions expended on the ionization of helium atoms is not used to excite vibrations of carbon monoxide molecules. Three- and four-particle reactions leading to the formation of He_2^+ molecular ions take place with a small rate. These ions subsequently undergo charge exchange with CO molecules to form a CO^+ ion, part of whose energy is transferred to vibrational degrees of freedom of CO molecules during further



(1a)

relaxation. The role of this channel in the utilization of the energy of He^+ ions increases as the fraction of carbon monoxide in the mixture decreases.

Excited helium atoms can be divided into two types: He** atoms, which are excited to high levels, and He* atoms, which are in metastable states. The main process involving He** atoms is associative ionization with the formation of He_2^+ molecular ions, which subsequently undergo charge exchange with CO molecules to form CO^+ ions. Metastable He* atoms mainly generate CO^+ ions as a result of the Penning effect. We note that since the main products of the reactions involving either He** or He* are CO^+ ions under the conditions considered here, we can refrain from separating the excited atoms into two types and consider them together, neglecting the less intense reactions (which are not shown in Fig. 1).

Thus, taking into account that the passage of a fast ion results in the generation of approximately 0.5 of an excited helium atom per He^+ ion, we find that the series of plasma-chemical processes produces about 0.5 of a CO^+ ion per He^+ ion. The bulk of the He^+ ions generate C^+ ions with a resultant loss of the energy expended by the fast ions on the ionization of helium from the energy available for pumping the laser.

The CO^+ ions formed are then rapidly converted as a result of three-particle reactions into $CO^+(CO)_n$ cluster ions, which subsequently recombine with electrons. Under the conditions considered here the main recombination process of the ions is dissociative recombination. Taking this into account, for the electron density we have $\approx 1.3 \times 10^{13}$ cm^{-3}. The electron temperature was set equal to 1000 K for estimates of the ion recombination rates.

Recombining dissociatively, the cluster ions form electronically excited molecules, which pass into metastable states as a result of radiative transitions. Such metastable molecules will be denoted by CO*. These CO* molecules are effectively quenched by carbon monoxide molecules in the ground electronic state.[8,12] As was shown in Ref. 13, 30–40% of the energy of the CO* molecules is transferred into vibrational degrees of freedom of carbon monoxide molecules, and the 5–13th vibrational levels are excited. Thus, the energy supplied to the vibrational degrees of freedom of CO molecules as a result of the plasma-chemical processes per He^+ ion formed by a fast ion is

$$K_{CO*}K_v E_{CO*} \approx 1.2 \text{ eV}, \tag{2}$$

where E_{CO*} is the electronic excitation energy of a CO* molecule, K_{CO*} is the number of CO* molecules formed per He^+ ion, and K_v is the fraction of E_{CO*} transferred to vibrational degrees of freedom when a CO* molecule is quenched.

Both the electrons generated by fast ions and the electrons formed in the Penning process have fairly high energies (~ 7 eV). During thermalization as a result of collisions with helium atoms, they reach the range for the effective excitation of molecular vibrations (1.2–2.5 eV) and transfer part of their energy to vibrational degrees of freedom of the molecules with a probability practically equal to unity, exciting the 1–8th vibrational levels (the *VE* process). Assuming that the fourth level is excited on the average, we find that the energy supplied to the vibrational degrees of freedom of CO molecules per He^+ ion as a consequence of the *VE* process is $1.5 \times 4E_v \approx 1.6$ eV, where E_v is the energy of a single vibrational quantum of a carbon monoxide molecule in the ground electronic state. As a result, the efficiency of the transfer of the energy imparted to the active medium by fast ions into vibrational degrees of freedom of the molecules

$$\eta_v = \frac{6E_v + K_{CO*}K_v E_{CO*}}{E_p} \tag{3}$$

is about 6%. There are two reasons for the not very high value of η_v: first, the loss of energy as a consequence of the conversion of He^+ ions into C^+ ions and, second, the large difference between the energy expended by a fast ion to form an electron–He^+ pair and the energy of the metastable level of a carbon monoxide molecule. Elimination of the first reason raises K_{CO*} to 1.5 and, accordingly, increases η_v to 12%. The value of K_{CO*} can be increased by adding a small quantity of carbon dioxide ($\sim 10^{-2} n_{CO}$) to the active medium. In this case the conversion of C^+ ions into CO^+ ions takes place according to the reaction

$$C^+ + CO_2 \rightarrow CO^+ + CO. \tag{4}$$

The addition of such a quantity of carbon dioxide converts C^+ ions into CO^+ ions, which are useful for pumping energy into vibrations of carbon monoxide molecules without creating other channels for significant energy losses. A small addition of oxygen can also provide for the conversion of C^+ into CO^+.

A decrease in the role of the conversion of He^+ into C^+ is also produced by an addition of nitrogen. When the nitrogen density $n_{N_2} = n_{CO} = 2.8 \times 10^{18}$ cm^{-3}, He^+ ions are converted into N_2^+ and N^+ ions with a rate comparable to the rate of their conversion into C^+ ions. Then the N_2^+ and N^+ ions are converted partially into CO^+ ions and partially into N_4^+ and N_3^+ clusters, which (N_4^+), in turn, are converted into CO^+ ions. The CO^+ ions formed are then converted into metastable CO* molecules, which are quenched by both CO and N_2 molecules with the transfer of part of the energy E_{CO*} to vibrational degrees of freedom of the CO and N_2 molecules.

The energy supplied to the vibrational degrees of freedom of the carbon monoxide molecules is subsequently redistributed over the vibrational spectrum as a consequence of the quenching of the metastable states of CO* molecules, the quasiresonant transfer of energy from $N_2(X, v')$, and the *VE* process. Finding the distribution function of the molecules among the vibrational levels calls for solving the vibrational kinetic equation, which is a problem in itself. Here we present the characteristic times of the principal processes in the vibrational subsystem of carbon monoxide under the conditions considered using the rate constants of the processes from Ref. 14. The characteristic time for the exchange of vibrational quanta between CO molecules amounts to $\approx 10^{-6}$ s for low levels (0,1; 1,0) and $\approx 10^{-5}$ s for intermediate levels (10,1; 11,0). The values in parentheses are the numbers of the initial and final vibrational levels. The time

for deactivation of the molecular vibrations as a result of collisions with helium atoms (the *VT* process) is $\approx 6 \times 10^{-3}$ s for the low levels and $\approx 10^{-4}$ s for the intermediate levels. The time for deactivation of the molecular vibrations as a result of spontaneous emission is $\approx 3 \times 10^{-2}$ s for the low levels and $\approx 4 \times 10^{-3}$ s for the intermediate levels.

Thus, under the conditions of a nuclear-pumped laser considered, as well as under the conditions of a gas-discharge laser, the transitions between the low and intermediate levels are determined by the exchange of vibrational quanta between molecules. The losses of energy from the vibrational subsystem are determined mainly by the *VT* process in helium atoms. The role of spontaneous emission is small.

As a consequence of their small mass, helium atoms effectively deactivate vibrationally excited molecules. Since the helium density in a nuclear-pumped laser is large and is two to three orders of magnitude greater than the helium density in a gas-discharge laser, the losses of vibrational energy as a result of the *VT* process increase by the same factor. This leads to an increase in the lasing threshold and a decrease in the efficiency of the laser. For example, the lasing of a nuclear-pumped He–CO laser was observed in Ref. 2 when the specific power deposition exceeds 200 W/cm^2.

Thus the plasma-chemical processes in the active medium of a He–CO nuclear-pumped laser can make the main contribution to the pumping of energy into the vibrational degrees of freedom of carbon monoxide molecules, permitting an increase in the pumping efficiency to $\eta_v \approx 10\%$ in the case of the use of additions of carbon dioxide, nitrogen, etc. However, the use of helium as a buffer gas in a nuclear-pumped CO laser is not optimal, since, first, the charge-exchange reaction of He^+ ions with carbon monoxide takes place with a large yield of C^+ ions, whose use for pumping energy into vibrations of CO molecules requires complication of the composition of the active medium, second, the generation of an electron/helium-ion pair by a fast ion requires large expenditures of energy ($\approx$45 eV), which significantly exceed the energy transferred as a result of this process to vibrational degrees of freedom of the carbon monoxide molecules ($\approx 3-5$ eV), and, third, helium atoms rapidly deactivate the vibrations of the carbon monoxide molecules. These deficiencies are significantly reduced when heavier inert gases are employed as the buffer gas.

3. Among the heavy inert gases, argon appears to be one of the promising substances for use as a buffer gas in a nuclear-pumped CO laser. The energy expended on the formation of an electron–ion pair by a fast ion is almost two times lower in argon than in helium. The ionization energy of argon is sufficient for the occurrence of fast charge exchange between argon ions and carbon monoxide molecules. The excited states of argon are effectively quenched by carbon monoxide molecules. Argon atoms have a mass which is sufficiently large for the probability of the *VT* process to be small.[15]

Let us examine the processes determining the pumping of energy into the vibrational degrees of freedom of carbon monoxide molecules in the active medium of a nuclear-pumped Ar–CO laser in greater detail. For estimates we choose the following parameters of the active medium:

$$\begin{aligned} &\text{gas temperature (translational)} && T_g = 200\ K, \\ &\text{argon density} && n_{Ar} = 2 \times 10^{19}\ \text{cm}^{-3}, \\ &\text{carbon monoxide density} && n_{CO} = 2 \times 10^{18}\ \text{cm}^{-3}, \\ &\text{specific power deposition} && \\ &\quad \text{in the medium} && W = 100\ \text{W/cm}^3. \end{aligned} \tag{5}$$

The argon pressure was chosen such that the mean free path of a U^{235} fission fragment would be approximately two times smaller than the diameter of the laser-active element used in the power model of a pulsed-reactor-pumped laser system.[16]

Formula (1a) shows the principal series of plasma-chemical reactions in the active medium of a nuclear-pumped Ar–CO laser under conditions (5).

The reaction partners and the reaction rates in s^{-1} are shown over the arrows, as in formula (1). The constants from the reference cited above were used to calculate the reaction rates. To estimate the recombination rates of the ions, the electron density was set equal to 10^{13} cm^{-3}, and the temperature of the thermal electrons was set equal to 1000 K.

The energy imparted to the Ar–CO medium by a fast ion is concentrated mainly in the form of the potential energy of Ar^+–electron pairs and in electronically excited argon atoms after the first stage of relaxation. As above, the electronically excited atoms, which are found mainly in metastable or nearby resonant states, are denoted by Ar*. Further conversion of the Ar^+ ions occurs along two channels. When Ar^+ ions interact with carbon monoxide molecules, rapid charge exchange takes place with the formation of CO^+ ions, which are subsequently converted into $CO^+ \cdot CO$ cluster ions. In the second channel Ar^+ ions are converted into Ar_2^+ ions, which, in turn, are converted as a result of interactions with carbon monoxide molecules into CO^+ or $ArCO^+$ ions, and these ions are subsequently converted into $CO^+ \cdot CO$ cluster ions.[17] The subsequent fate of the $CO^+ \cdot CO$ clusters in an Ar–CO active medium is similar to their fate in a He–CO medium, and the final step is the formation of CO* molecules.

The excited Ar* atoms are effectively quenched by carbon monoxide molecules. We do not know of any detailed investigations of the products of this reaction. However, since carbon monoxide molecules have several electronic levels with excitation energies close to the excitation energy of the Ar* atoms, it can be assumed that upon quenching the excitation energy of the argon atoms is used to electronically excite carbon monoxide molecules, which then form CO* molecules.

Thus, as a result of the series of plasma-chemical processes, the energy stored in argon ions and excited atoms is transformed into the electronic excitation energy of carbon monoxide molecules, and 1.5 CO* molecules form per Ar^+ ion. As in a He–CO medium, these molecules are quenched by CO molecules with the passage of $30-40\%$ of the electronic excitation energy into vibrational degrees of freedom of the carbon monoxide molecules. As a result, taking into account that there is no Penning effect accompanying the interaction of Ar* with CO molecules, we find in analogy to

(3) that the efficiency of the transfer of the energy imparted to the Ar–CO medium by fast ions into vibrational degrees of freedom of carbon monoxide molecules amounts to 18%, which is 1.5 times greater than in a He–CO medium with an addition of carbon dioxide.

Several important details have been omitted in the mechanisms for the plasma-chemical pumping of nuclear-pumped CO lasers considered above. It is known that the products of several reactions taking place in the active medium of a nuclear-pumped CO laser are in electronically and vibrationally excited states with a high probability. This occurs, in particular, in the charge-exchange reactions of He_2^+ with nitrogen, the Penning ionization of carbon monoxide and nitrogen molecules by He* atoms, etc. The quenching of these excited states by CO or N_2 molecules can increase the flow of energy into the vibrational degrees of freedom of carbon monoxide molecules and thus increase η_v. The characteristic cross sections for the quenching of electronically excited states of the particles by these molecules are close to the gas-kinetic value, and the fraction of energy entering the vibrational degrees of freedom of the molecules amounts to a considerable fraction of the electronic transition energy. However, additional detailing of the mechanism of the plasma-chemical pumping of a nuclear-pumped CO laser with the inclusion of these processes would hardly be productive at the present time because of the lack of reliable data on the probabilities of the various reaction channels for deactivating the excited particles by these molecules and, in some cases, on the overall rate constants of the quenching of the excited states.

In the vibrational subsystem of carbon monoxide molecules in an Ar–CO active medium under conditions (5), just as in a He–CO medium under conditions (1), the transitions between low and intermediate vibrational levels are determined by the exchange of vibrational quanta. However, the role of the *VT* process in the energy losses from the vibrational subsystem is considerably smaller in an Ar–CO medium than in a He–CO medium. This fact is no less important for the efficiency of laser-pumped Ar–CO lasers than is the smaller expenditure of energy for the formation of an Ar^+–electron pair by a fast ion in comparison to a He^+–electron pair.

At the present time the literature does not offer any information on reliable measurements of the rate constant of the *VT* process for carbon monoxide molecules resulting from their collisions with argon atoms at low gas temperatures.[15] To estimate it, we use the formula given in Ref. 18, from which it follows that

$$k_{VT;\text{CO-Ar}}^{(1;0)}=4.0\times10^{-22}\text{cm}^3/\text{s}. \tag{6}$$

Using (6) we find that the deactivation time of the first vibrational level of carbon monoxide molecules resulting from their collisions with argon atoms is equal to 120 s. Assuming for an estimate that the dependence of $k_{VT;\text{CO-Ar}}^{(v;v-1)}$ on the number of the level is similar to the dependence of $k_{VT;\text{CO-CO}}^{(v;v-1)}$ and using $k_{VT;\text{CO-CO}}^{(v;v-1)}$ from Ref. 14, for intermediate vibrational levels we have $k_{VT;\text{CO-Ar}}^{(10;9)}=3\times10^{-20}$ cm^3 and a deactivation time equal to 1.7 s.

Since the concentration of carbon monoxide molecules is an order of magnitude lower than the concentration of argon atoms under the conditions considered and since $k_{VT;\text{CO-CO}}^{(v;v-1)}$ is an order of magnitude greater than $k_{VT;\text{CO-Ar}}^{(v;v-1)}$, the rate of the *VT* process resulting from collisions of the molecules with one another is comparable to the rate of the *VT* process resulting from collisions of the molecules with argon atoms.

It is seen from the estimates presented that the rate of the deactivation of the low and intermediate vibrational levels of carbon monoxide molecules as a consequence of their collisions with one another and with argon atoms is considerably lower than the rate of their deactivation as a consequence of spontaneous emission, so that the energy loss from the vibrational subsystem in the absence of lasing is determined by spontaneous emission. Let us estimate this loss. At the low and intermediate vibrational levels the probability of spontaneous radiative de-excitation can be approximated by a linear dependence on the level number v:

$$(\tau_{\text{rad}}^{(v;v-1)})^{-1}=(\tau_{\text{rad}}^{(1;0)})^{-1}v. \tag{7}$$

Since the probability of radiative de-excitation does not increase rapidly with the level number, the main contribution to the energy loss is made by the lower levels due to their larger populations, and in estimating the energy loss we can confine ourselves to the harmonic approximation, assuming that the populating of the vibrational levels obeys a Boltzmann law with a vibrational temperature T_v. As a result, for the radiative energy loss we have

$$Q_{\text{rad}}=\frac{E_v\varepsilon_v n_{\text{CO}}}{\tau_{\text{rad}}^{(1;0)}}, \tag{8}$$

where

$$\varepsilon_v=[\exp(E_v/T_v)-1]^{-1} \tag{1}$$

is the mean number of quanta per molecule at the vibrational temperature T_v.

Let us use (8) to estimate the pump power density W_{th} corresponding to the lasing threshold of a nuclear-pumped Ar–CO laser. Since the lasing of a CO laser is observed at a vibrational temperature $T_v^{(0)}\approx3000$ K,[19], at the lasing threshold the supply of energy to the vibrational subsystem of the carbon monoxide molecules should be equal to Q_{rad} at $T_v=T_v^{(0)}$, and

$$W_{\text{th}}=\frac{1}{\eta_v}Q_{\text{rad}}(T_v=T_v^{(0)}). \tag{10}$$

When conditions (5) are satisfied and $\eta_v=18\%$, we have $W_{\text{th}}\approx10$ W/cm^3. This value is an order of magnitude lower than the value of W_{th} for a nuclear-pumped He–CO laser.

4. Thus, the plasma-chemical processes in the active medium of a nuclear-pumped CO laser can make a significant contribution to the pumping of energy into the vibrational degrees of freedom of carbon monoxide molecules, permitting the achievement of a pumping efficiency of 18% for them. The use of helium as the buffer gas in the active medium of a nuclear-pumped CO laser is not optimal because of the large expenditures of the energy of the fast ions pro-

duced by nuclear reactions on the generation of electron/helium-ion pairs and the high rate of the *VT* process in the helium atoms. The use of argon as the buffer gas instead of helium permits a 1.5-fold increase in the efficiency of the pumping of energy into vibrational degrees of freedom of the carbon monoxide molecules and an order-of-magnitude decrease in the threshold energy for pumping the active medium

In conclusion, we sincerely thank P. P. D'yachenko and A. P. Budnik for their interest in this work and for discussing its results.

[1] V. A. Zherebtsov, ''A nuclear-pumped CO laser. 1. Plasma-chemical mechanism for the nuclear pumping of a He–CO–N_2 laser'' [in Russian], Preprint No. 2548, Institute of Physics and Power Engineering, Obninsk (1996), 22 pp.
[2] N. W. Jalufka and F. Hohl, Appl. Phys. Lett. **39**, 139 (1981).
[3] L. I. Gudzenko, V. S. Malyshevskiĭ, and S. I. Yakovlenko, Zh. Tekh. Fiz. **48**, 2150 (1978) [Sov. Phys. Tech. Phys. **23**, 1228 (1978)].
[4] R. T. Schneider and F. Hohl, in *Advances in Nuclear Science and Technology*, Vol. 16, New York (1984), pp. 123–287.
[5] L. I. Virin, R. V. Dzhagatspanyan, G. V. Karachentsev *et al.*, *Ion-Molecule Reactions in Gases* [in Russian] (Nauka, Moscow, 1979), 548 pp.
[6] B. M. Smirnov, *Atomic Collisions and Elementary Processes in Plasmas*, [in Russian] (Atomizdat, Moscow, 1968), 363 pp.
[7] B. M. Smirnov, *Ions and Excited Atoms in Plasmas* [in Russian] (Atomizdat, Moscow, 1974), 456 pp.
[8] B. M. Smirnov, *Excited Atoms* [in Russian] (Énergoizdat, Moscow, 1982), 232 pp.
[9] B. M. Smirnov, *Cluster Ions and Van der Waals Molecules* [Gordon and Breach, Philadelphia (1992); Nauka, Moscow (1983), 150 pp.].
[10] A. V. Eletskiĭ and B. M. Smirnov, Usp. Fiz. Nauk **136**, 25 (1982) [Sov. Phys. Usp. **25**, 13 (1982)].
[11] W. J. Wiegand Jr., in *Applied Atomic Collision Physics, Vol. 3: Gas Lasers*, edited by H. S. W. Massey, E. W. McDaniel, and B. Bederson, [Academic Press, New York (1982), Chap. 3; Mir, Moscow (1986), pp. 95–123].
[12] R. A. Young and G. V. Volkenburgh, J. Chem. Phys. **55**, 2990 (1971).
[13] Yu. Z. Ionikh, A. L. Kuranov, A. N. Lobanov *et al.*, Opt. Spektrosk. **60**, 727 (1986) [Opt. Spectrosc. **60**, 444 (1986)].
[14] G. D. Billing, in *Nonequilibrium Vibrational Kinetics*, edited by M. Capitelli [Springer-Verlag, Berlin–New York (1986), pp. 85–112; Mir, Moscow (1989), pp. 104–136].
[15] E. E. Nikitin, A. I. Osipov, and S. Ya. Umanskiĭ, in *Reviews of Plasma Chemistry, Vol. 2*, edited by B. M. Smirnov [Consultants Bureau, New York (1994); Énergoatomizdat, Moscow (1989), pp. 3–44].
[16] A. V. Gulevich, P. P. D'yachenko, A. V. Zrodnikov *et al.*, At. Energy **80**, 361 (1996).
[17] D. K. Bohme, N. G. Adams, and M. Mosesman *et al.*, J. Chem. Phys. **52**, 5094 (1970).
[18] V. K. Ablekov, Yu. N. Denisov, and F. N. Lyubchenko, *Handbook of Gas-Dynamic Lasers* [in Russian] (Mashinostroenie, Moscow, 1982), 162 pp.
[19] N. N. Sobolev and V. V. Sokovikov, Usp. Fiz. Nauk **110**, 191 (1973) [Sov. Phys. Usp. **16**, 350 (1973)].

Translated by P. Shelnitz

Theory of scanning near-field magnetooptical microscopy

V. A. Kosobukin

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia
(Submitted March 13, 1997)
Zh. Tekh. Fiz. **68**, 86–91 (July 1998)

An analytical theory of scanning near-field magnetooptical microscopy is developed. The theory is based on the elastic scattering of light by small, resonantly polarizable particles, which are used to scan the plane surface of a nonuniformly magnetized medium. The effective polarizability of the particles is calculated with the effect of dynamic ''image forces'' taken into account in all orders of perturbation theory with respect to the interaction of the particle with a demagnetized ferromagnet, and the magnetooptical perturbation is calculated to first order in the magnetization. The major contributions to the magnetooptical light scattering for a ferromagnetic structure magnetized perpendicular to the surface are found, including a quasistatic approximation for the near-field particle–magnet interaction. The optical size resolution of a magnetic (dielectric) inhomogeneity is estimated. © *1998 American Institute of Physics.* [S1063-7842(98)01507-4]

The prospects for magnetooptical studies have expanded significantly as a result of the overall progress in near-field optics.[1] Thus, recently developed near-field magnetooptical microscopy techniques[1–5] make it possible to observe contrast in the magnetooptical response over scale lengths considerably shorter than an optical wavelength, $\sim c/\omega$, where c is the wavelength and ω is the frequency of the light. A theory of magnetooptical Kerr effects in near-field light has been proposed[6,7] for application to optical experiments[3,4] in the case of uniform magnetization in the plane of the surface. The near-field response of nonmagnetic surface inhomogeneities has been studied theoretically in several papers,[8,9] primarily by numerical methods for near-field probing using an optical fiber with a sharpened tip. The possibility of determining the size and shape of small surface dielectric inhomogeneities based on their near-field response in the absence of magnetooptical effects was also discussed.

The purpose of this paper is to construct an analytical theory of the near-field microscopy of magnetic structures during scanning of the surface by small metallic particles. The model is illustrated in Fig. 1. It includes a plane interface boundary $z=0$ between a transparent nonmagnetic dielectric (vacuum) ($z<0$) and a medium ($z>0$) with a magnetic structure that is nonuniform in the planes $z=\text{const}$. A small (of size much smaller than c/ω) particle or dipole-polarizable inhomogeneity of another type exists in the first medium. A linearly polarized light wave incident on the nonmagnetic medium can be scattered by the particle or magnetic inhomogeneity or can participate in combination processes involving both these types of scattering. The role of these combination processes in scanning near-field microscopy is studied in this paper for the case in which the magnetization (magnetic field) vector is perpendicular to the surface, i.e., the magnetooptical Kerr effect occurs.[10] Thus, the results obtained below are a generalization[7] to the case of media with a nonuniform distribution of the magnetization in the plane of the surface.

MODEL AND BASIC EQUATIONS

Following Ref. 7, we define the unperturbed model (Fig. 1) by the dielectric tensor $\varepsilon^0(z,\omega)\hat{I}$ with $\varepsilon^0(z)=\varepsilon_1\vartheta(-z)+\varepsilon_2\vartheta(z)$, where $\vartheta(z)=0$ for $z<0$ and $\vartheta(z)=1$ for $z>0$, while $\hat{I}$ is the unit tensor, whose components are the Kronecker delta symbols $\delta_{\alpha\beta}$ with the Cartesian indices α and β. The electric field $\mathbf{E}^0(\mathbf{r})$ of the light and the tensor Green function $\hat{D}^0(\mathbf{r},\mathbf{r}')$ in the absence of a perturbation and the total field $\mathbf{E}(\mathbf{r})$ in the presence of a perturbation in the dielectric polarization $\mathbf{P}(\mathbf{r})$ are described by the equations[7]

$$[\operatorname{curl}\operatorname{curl}-\varepsilon^0(z,\omega)k_0^2\hat{I}]\begin{Bmatrix}\mathbf{E}^0(\mathbf{r})\\ \hat{D}^0(\mathbf{r},\mathbf{r}')\\ \mathbf{E}(\mathbf{r})\end{Bmatrix}=\begin{Bmatrix}0\\ \delta(\mathbf{r}-\mathbf{r}')\hat{I}\\ 4\pi k_0^2\mathbf{P}(\mathbf{r})\end{Bmatrix}, \tag{1}$$

where $\mathbf{E}^0(\mathbf{r})$ and $\hat{D}^0(\mathbf{r},\mathbf{r}')$ satisfy Maxwellian boundary conditions with respect to the variable $\mathbf{r}$ and $k_0=\omega/c$. For components of the Green function, in the following we shall use the representation

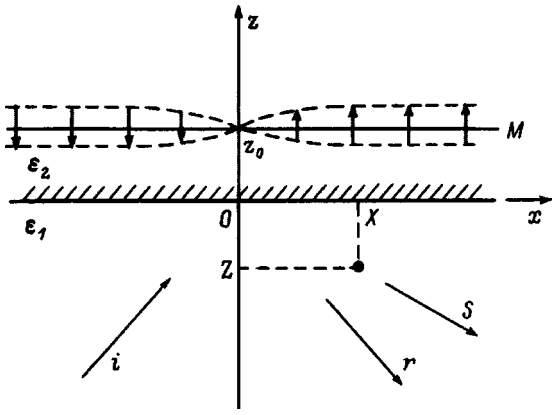


FIG. 1. Illustration of near-field magnetooptical microscopy.

$$D^0_{\alpha\beta}(\mathbf{r},\mathbf{r}',\omega)=\int \frac{d^2Q}{(2\pi)^2}\exp[i\mathbf{Q}(\mathbf{r}_\parallel-\mathbf{r}'_\parallel)]\times G^0_{\alpha\beta}(z,z';\mathbf{Q},\omega),$$

$$G^0_{\alpha\beta}(z,z';\mathbf{Q},\omega)=\sum_{\mu,\nu} S_{\alpha\mu}(\mathbf{Q}/Q)d^0_{\mu\nu}(z,z';Q,\omega)S_{\beta\nu}(\mathbf{Q}/Q), \quad (2)$$

where $\mathbf{r}_\parallel=(x,y)$, $\mathbf{Q}=(\mathbf{e}_xQ_x+\mathbf{e}_yQ_y)$, $S_{xx}=S_{yy}=Q_x/Q$, $-S_{xy}=S_{yx}=Q_y/Q$, $S_{zz}=1$, and $\mathbf{e}_\alpha$ is the unit vector for the α th Cartesian axis.

The components $d^0_{\alpha\beta}(z,z';Q,\omega)$ of the Green function in Eq. (2) are the solutions of inhomogeneous ordinary differential equations obtained from the second of Eqs. (1) with $\mathbf{Q}=\mathbf{e}_xQ$ when $(\partial/\partial\mathbf{r}=(iQ,0,d/dz)$.[11]

The perturbation $\mathbf{P}=\mathbf{P}^{\mathrm{I}}+\mathbf{P}^{\mathrm{II}}$ includes contributions from the nonmagnetic particle (the source of the near field) $\mathbf{P}^{\mathrm{I}}$ and from the magnetized medium $\mathbf{P}^{\mathrm{II}}$. In the following $\mathbf{P}^{\mathrm{I}}$ is taken into account self-consistently in all orders of perturbation theory, and $\mathbf{P}^{\mathrm{II}}$, is so taken to first order in the magnetization. In terms of the theory of the multiple scattering of light for small particles near a plane surface, we obtain[12]

$$P^{\mathrm{I}}_\alpha(\mathbf{r},\omega)=\chi^{(\alpha)}(\omega)\,\delta(\mathbf{r}-\mathbf{R})E^0_\alpha(\mathbf{R},\omega), \quad (3)$$

where $\mathbf{R}=(X,0,Z)$ and $Z<0$ (Fig. 1).

The components of the diagonal polarizability tensor for a particle near the surface are given by $\chi^{(\beta)}=[1/\alpha^{(\beta)}-\sigma^{(\beta)}]^{-1}$ in terms of the components of the diagonal polarizability tensor of the particle, $\alpha^{(\beta)}$, calculated as the response to an external field in a uniform medium.[13] The $\sigma^{(\beta)}$ are calculated in terms of $\hat{D}^0(\mathbf{r},\mathbf{r}')$ in Ref. 7. This transition from $\alpha^{(\beta)}$ to $\chi^{(\beta)}$ corresponds to including the near surface ''image'' charges of the particle (near field) in the absence of magnetization.[13]

For the magnetooptical contribution to the polarization (in the medium with $z>0$), in general we take

$$4\pi P^{\mathrm{II}}_\alpha(\mathbf{r})=\sum_\beta \Delta\varepsilon_{\alpha\beta}(\mathbf{r},\mathbf{H})E_\beta(\mathbf{r}) = i\varepsilon_B(\mathbf{H})h(\mathbf{r}_\parallel)f(z)\sum_\beta \Psi_{\alpha\beta}(\mathbf{H}/H)E_\beta(\mathbf{r}), \quad (4)$$

where it is assumed that the magnetization $\mathbf{M}$ and magnetic field $\mathbf{H}$ vectors are parallel to one another.

The presence of a spatial modulation in the dielectric (magnetization) tensor of Eq. (4), expressed by the function $h(\mathbf{r}_\parallel)$, is fundamental to an examination of scanning microscopy. In the following, we shall assume that the magnetic inhomogeneity is uniform in the surface plane, i.e., $h(\mathbf{r}_\parallel)=h(x)$ in Eq. (4).

ELEMENTARY PROCESSES

We shall consider linearly polarized *TE* (*s* polarization) and *TM* (*p* polarization) waves as the elementary incident and secondary waves. In them the electric field vector is, respectively, perpendicular and parallel to the plane formed by the wave vector and the normal to the interface at $z=0$. Here the field of *s*- and *p*-polarized waves with $\mathbf{Q}^i=\mathbf{e}_xQ^i=\mathbf{e}_x\sqrt{\varepsilon_1}k_0\sin\Theta^i$, incident at an angle Θ^i, is given by one of the following expressions ($z<0$):

$$\{\mathbf{E}^i_s(\mathbf{r}),\ \mathbf{E}^i_p(\mathbf{r})\}=\{\overline{E^i_s}\mathbf{e}_y,\ \overline{E^i_p}(\mathbf{e}_x\cos\Theta^i-\mathbf{e}_z\sin\Theta)\}\times\exp[i(Q^ix+\sqrt{\varepsilon_1k_0^2-(Q^i)^2}z)]. \quad (5)$$

The solution of the last of Eqs. (1) under conditions (3)–(5) is given by[6,7]

$$\mathbf{E}(\mathbf{r},\omega)=\sum_{n=0}^{5}\mathbf{E}^{(n)}(\mathbf{r},\omega). \quad (6)$$

Here $\mathbf{E}^{(0)}(\mathbf{r})=\mathbf{E}^0(\mathbf{r})=\exp(iQ^ix)\mathbf{E}^0(z;Q^i)$ corresponds to mirror reflection of the light in the absence of a particle and a magnetic inhomogeneity. The other contributions, illustrated by the diagrams of Fig. 2, refer to elastic (Rayleigh) scattering of light between the initial (i) and final (f) states of the radiation. Diagram *1* corresponds to scattering of light by a subsurface particle in the absence of magnetization, which has been studied before.[7]

The remaining diagrams ($n=2-5$) are a comprehensive representation of the set of magnetooptical scattering processes that are linear in the magnetization. Diagram *2* applies to diffraction of light on a magnetic inhomogeneity, and in the case of a uniform inhomogeneity with a distribution function $h(x)$, the scattered light field is given by

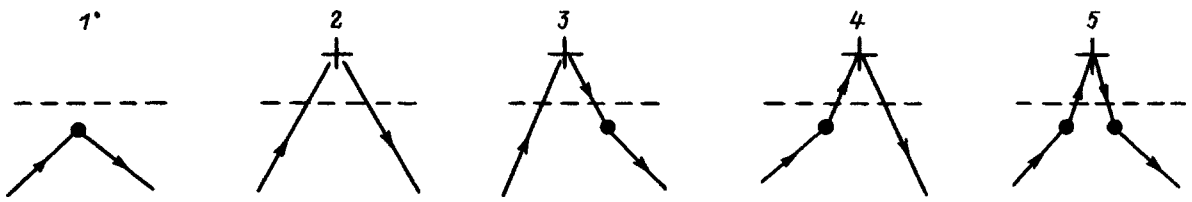


FIG. 2. A schematic representation of the zeroth and first order contributions, in terms of the magnetization, to the magnetooptical scattering of light by a small nonmagnetic particle and a magnetization inhomogeneity. The smooth lines are waves propagating in an unperturbed layered medium (the functions $\hat{D}^0$ and $\mathbf{E}^0$), the points represent the scattering of light by a particle (polarizability $\hat{\chi}$), the crosses represent the magnetooptical Kerr effect (susceptibility $\Delta\hat{\varepsilon}/4\pi$), and the dashed lines, an arbitrary interface between the media.

$$\mathbf{E}^{(2)}(\mathbf{r}) = k_0^2 i\varepsilon_0 \int_{-\infty}^{\infty} \frac{dQ_x}{2\pi} \times \exp[i(Q_x x - k_1(Q_x)z)]\hat{d}^0(0^-, 0^+, Q_x)\hat{\Psi} \times H(Q_x - Q_x^i)F[k_2(Q_x) + k_2(Q_x^i)]\mathbf{E}^0(0^+; Q_x^i). \quad (7)$$

Here $Q_x^i = Q^i$, while $k_m \equiv \sqrt{\varepsilon_m k_0^2 - Q^2}$ or, in general, $k_m^t \equiv k_m(Q^t) = \sqrt{\varepsilon_m k_0^2 - (Q^t)^2}$ for light in a medium with number m in the tth scattering step,

$$H(Q_x) = \int_{-\infty}^{\infty} dx \, \exp(-iQ_x x)h(x), \quad (8)$$

and

$$F[k_2(Q') + k_2(Q'')] = \int_0^{\infty} dz \, \exp\{i[k_2(Q') + k_2(Q'')]z\}f(z). \quad (9)$$

For $\sqrt{x^2+z^2} \gg 1/k_0$, the integral in Eq. (7) is calculated by the method of steepest descent[7] with a saddle point $Q_x^f = \sqrt{\varepsilon_1}k_0 \sin\Theta^f \operatorname{sgn}(x)$, where $\Theta^f = \arctan(|x/z|)$ and the scattering angle Θ^f is defined with respect to the negative direction of the z axis. Because of the uniform distribution $h(x)$, the light is scattered in the xz plane. Here in the function $H(Q_x^f - Q^i)$ of Eq. (7), $|Q_x^f - Q^i| \sim k_0$, i.e., this process is effective only in ordinary magnetooptical microscopy, where, according to the Rayleigh criterion, the resolution of the magnetic objects is limited to a scale length $\sim 1/k_0$ (Ref. 10).

NEAR-FIELD MAGNETOOPTICAL SCATTERING

The processes corresponding to diagrams $n = 3-5$ are of interest for near-field microscopy. A common feature of these processes is the possibility of double scattering (on a particle and on an inhomogeneity), in each event of which the component Q_x of the two-dimensional wave vector $\mathbf{Q}$ can change in magnitude by an amount much greater than k_0. If the corresponding components have a magnetic inhomogeneity in the spatial Fourier spectrum, then the elastic scattering of the light may include and X-dependent modulation in the response on a subwavelength scale.

Diagram *3* corresponds to the scattering of light on a magnetic inhomogeneity, after which each s- or p-polarized wave is scattered by a particle. The radiation field in these processes is given by

$$\mathbf{E}^{(3)}(\mathbf{r}) = 4\pi k_0^4 i\varepsilon_0 \hat{D}^0(\mathbf{r},\mathbf{R})\hat{\chi}\hat{U}(\mathbf{R};Q^i)\hat{\Psi}\mathbf{E}^0(0^+;Q^i). \quad (10)$$

Here

$$\hat{U}(\mathbf{R};Q^i) = \int_{-\infty}^{\infty} \frac{dQ_x}{2\pi} \exp[i(Q_x X - k_1(Q_x)Z)]\hat{d}^0 \times (0^-,0^+;Q_x)H(Q_x - Q^i)F(k_2 + k_2^i), \quad (11)$$

with $\hat{D}^0(\mathbf{r},\mathbf{r}')$ defined by Eq. (1). Diagram *4* describes the scattering of light by a particle followed by diffraction of the s- or p-polarized waves on a magnetic inhomogeneity. These processes make the following contribution to the radiation field:

$$\mathbf{E}^{(4)}(\mathbf{r}) = 4\pi k_0^4 i\varepsilon_0 \int \frac{d^2Q}{(2\pi)^2} e^{i(\mathbf{Q}\cdot\mathbf{r}_\parallel - k_1(\mathbf{Q})z)} \times \hat{G}^0(0^-,0^+;\mathbf{Q})\hat{\Psi}\hat{V}(\mathbf{Q};\mathbf{R})\hat{\chi}\mathbf{E}^0(\mathbf{R}), \quad (12)$$

with

$$\hat{V}(\mathbf{Q};\mathbf{R}) = \int_{-\infty}^{\infty} \frac{dQ_x'}{2\pi} \exp[-i(Q_x' X + k_1(\mathbf{Q}')Z)] \times \hat{d}^0(0^+,0^-;Q_x')H(Q_x - Q_x')F(k_2 + k_2'), \quad (13)$$

where $\mathbf{Q} = (Q_x, Q_y)$, $\mathbf{Q}' = (Q_x', Q_y)$.

Finally, diagram *5* refers to magnetooptical scattering on an inhomogeneity which takes place between two events of scattering by a particle. The corresponding contribution to the field (6) is given by

$$\mathbf{E}^{(5)}(\mathbf{r}) = 16\pi^2 k_0^6 i\varepsilon_0 \hat{D}^0(\mathbf{r},\mathbf{R})\hat{\chi}\hat{W}(\mathbf{R})\hat{\chi}\mathbf{E}^0(\mathbf{R}), \quad (14)$$

where

$$\hat{W}(\mathbf{R}) = \int \frac{d^2Q}{(2\pi)^2} \exp[i(Q_x X - k_1(\mathbf{Q})Z)] \times \hat{G}^0(0^-,0^+;\mathbf{Q})\hat{\Psi}\hat{V}(\mathbf{Q};\mathbf{R}). \quad (15)$$

For a given direction $\mathbf{r}/r$, with $|\mathbf{r}| \gg c/\omega$, each component of the field (6) can be represented locally in the form of a linear combination of two asymptotically plane waves $\sim \exp(i\sqrt{\varepsilon_1}k_0 r)$ with s and p polarizations. The amplitudes $\overline{E}_{\lambda'}^{(n)}$ of these waves are related to the amplitude $\overline{E}_{\lambda}^{i}$ of the λ-polarized incident wave (5) by the equations[6,7]

$$\overline{E}_{\lambda'}^{(2)} = (1/\sqrt{r})l_{\lambda'\lambda}^{(2)}\overline{E}_{\lambda}^{(i)}, \quad \overline{E}_{\lambda'}^{(n)} = (1/r)l_{\lambda'\lambda}^{(n)}\overline{E}_{\lambda}^{(i)}, \quad (16)$$

where the polarization indices of the incident λ and scattered λ' waves take on the designations s and p.

In Eq. (16) the first expression corresponds to a cylindrically diverging wave (7) and the second, to spherical waves with $n = 1, 3-5$. The transverse cross section for scattering of light into the element of solid angle $d\Omega^f = \sin\Theta^f d\Theta^f d\varphi^f$ with the polarization transformation $\lambda \to \lambda'$ is

$$\frac{d\sigma_{\lambda'\lambda}(\mathbf{M})}{d\Omega^f} = \left|\sum_{n=1}^{5} l_{\lambda'\lambda}^{(n)}\right|^2 \cong |l_{\lambda'\lambda}^{(1)}|^2 + 2\sum_{n=2}^{5} \operatorname{Re}[l_{\lambda'\lambda}^{(1)} l_{\lambda'\lambda}^{(n)*}] \equiv \sum_{n=1}^{5} \frac{d\sigma_{\lambda'\lambda}^{(n)}}{d\Omega^f}. \quad (17)$$

Here $d\sigma_{\lambda'\lambda}(0)/d\Omega^f \equiv d\sigma_{\lambda'\lambda}^{(1)}/d\Omega^f = |l_{\lambda'\lambda}^{(1)}|^2$ is the transverse cross section for elastic scattering of light by a surface particle in the absence of magnetization, and it is assumed that $|l_{\lambda'\lambda}^{(1)}| \gg |l_{\lambda'\lambda}^{(n)}|$ for all $n \geqslant 2$. Expressions for the angles of rotation of the axes of the polarization ellipse and the scattered light ellipses in terms of $l_{\lambda'\lambda}^{(n)}$ are given by Eqs. (35) and (36), respectively, of Ref. 7.

OPTICAL RESOLUTION AND IMAGE SHAPE

We now consider the previous results as applied to scanning near-field magnetooptical microscopy, i.e., we shall study the quantities (16) and (17) as functions of the coordinate X in the case $|\mathbf{R}| \ll 1/k_0$. To do this we examine an ultrathin ferromagnetic layer of atomic thickness l, which is magnetized perpendicular to its plane ($\mathbf{M} \| \mathbf{e}_z$) and is uniformly inhomogeneous in this plane, i.e, in Eq. (4) we have

$$\Psi_{\alpha\beta} = \delta_{\alpha x}\delta_{\beta y} - \delta_{\alpha y}\delta_{\beta x}, \quad f(z) = l\,\delta(z - z_0), \tag{18}$$

where $z_0 > 0$ and $l < z_0 \ll 1/k_0$.

We shall assume that in this layer there is a domain wall, in which the uniform distribution of the magnetization and its Fourier transform are given by

$$h(x) = (2/\pi)\arctan(x/\Delta),$$

$$H(Q_x) = 2\exp(-|Q_x|\Delta)/(iQ_x), \tag{19}$$

where Δ is the domain wall width ($0 < \Delta \ll 1/k_0$).

This model corresponds, in particular, to nanostructures formed by ultrathin layers of a ferromagnetic material with spacers made of a precious metal, where the axis of easy magnetization is perpendicular to the interfaces.[14]

Under conditions (18) and (19), the matrix elements $l^{(n)}_{\lambda'\lambda}$ are calculated by analogy with Ref. 7, where the polar magnetooptical Kerr effect was studied in the near field of a small particle in the case of a uniform magnetic medium (i.e., for $h(x) = 1$). In calculating the $l^{(n)}_{\lambda'\lambda}$ in the near field approximation ($|Z|, z_0, \Delta \ll 1/k_0$), we consider the domain of integration in Eqs. (11) and (13) to be determined by the wave vector $|Q_x| \sim 1/|Z| \gg k_0$. Assuming further that $|Q_x| \gg \sqrt{|\varepsilon_m|}k_0$, we shall neglect the contributions, in the near field, from the purely perpendicular s-polarized electric field, which are described by the functions d^0_{yy}: the latter are small by a factor $\sim(k_0Q)^2 \sim (k_0Z)^2 \ll 1$ compared to the functions $d^0_{\alpha\beta}$ corresponding to the longitudinal-transversely p-polarized field. (Here the subscripts α and β are equal to x or z.) Given Eq. (19), the integrals (11) and (13) are expressed in the quasistatic approximation in terms of the following functions:

$$\{K_X(\mathbf{R}), K_Z(\mathbf{R})\} = \{X, -L\}\,\frac{1}{X^2 + L^2}, \tag{20}$$

where $L = |Z| + z_0 + \Delta$ determines the characteristic distance between the particle and the probe object along the normal to the surface, with $L \ll 1/k_0$.

Under these conditions, the principal matrix elements are

$$\begin{Bmatrix} l^{(3)}_{ss} \\ l^{(3)}_{ps} \end{Bmatrix} = A\,\exp(-iQ^f X\cos\varphi^f) \times \begin{Bmatrix} g^+_s(Q^f)\chi^{(x)}K_X \sin\varphi^f \\ g^+_p(Q^f)\chi^{(x)}K_X \cos\Theta^f \cos\varphi^f + g^-_p(Q^f)\chi^{(z)}K_Z \sin\Theta^f \end{Bmatrix} \times t_s(Q^i), \tag{21}$$

and

$$\begin{Bmatrix} l^{(4)}_{sp} \\ l^{(4)}_{pp} \end{Bmatrix} = A \begin{Bmatrix} t_s(Q^f)\cos\varphi^f \\ -t_p(Q^f)\cos\Theta^f \sin\varphi^f \end{Bmatrix} [K_X\chi^{(x)} g^+_p(Q^i)\cos\Theta^i - K_Z\chi^{(z)} g^-_p(Q^i)\sin\Theta^i]\exp(iQ^iX). \tag{22}$$

Here

$$A = \frac{i\varepsilon_B k_0^2 \overline{t_p} l}{\pi\varepsilon_1}, \tag{23}$$

$$g^{\pm}_\lambda(Q) = \exp[ik_1(Q)Z] \pm r_\lambda(Q)\exp[-ik_1(Q)Z], \tag{24}$$

$\varphi^f = \arctan(y/x)$, and $\Theta^f = \arctan(\sqrt{x^2+y^2}/|z|)$, with the angle Θ^f taken from the negative z direction. The reflection $r_\lambda(Q)$ and transmission $t_\lambda(Q)$ coefficients for λ-polarized waves are given by

$$r_\lambda(Q) = \frac{\eta^\lambda_1(Q) - \eta^\lambda_2(Q)}{\eta^\lambda_1(Q) - \eta^\lambda_2(Q)}, \quad t_\lambda(Q) = 1 + r_\lambda(Q), \tag{25}$$

where $\eta^p_m(Q) = \varepsilon_m/k_m(Q)$, $\eta^s_m(Q) = k_m(Q)$, and $k_m(Q) = \sqrt{\varepsilon_m k_0^2 - Q^2}$ in medium number m.

In Eq. (23), $\overline{t_p} = 2\varepsilon_1/(\varepsilon_1 + \varepsilon_2)$ is the quasistatic ($k_0/Q \to 0$) limit of $t_p(Q)$, while for $|Z| \ll 1/k_0 \sim 1/Q^f$, Eq. (24) takes the form $\overline{g^\pm_\lambda} = 1 \pm \overline{r_\lambda}$, where $\overline{r_p} = (\varepsilon_1 - \varepsilon_2)/(\varepsilon_1 + \varepsilon_2)$ and $\overline{r_s} = 0$.

Using Eqs. (4) and (18), we write Eq. (15) in the form $\hat{W}(\mathbf{R}) = l\hat{J}(\mathbf{R})$, and in the quasistatic approximation ($k_0 L \ll 1$)

$$\hat{J}(\mathbf{R}) = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty} dQ_x \int_{-\infty}^{\infty} dQ'_x \int_{-\infty}^{\infty} dQ_y \times \exp[i(Q_x - Q'_x)X]\exp[-(|Q_x|) + |Q'_x|)(|Z| + z_0)] \times H(Q_x - Q'_x)\hat{G}^0(0^-, 0^+; \mathbf{Q})\hat{\Psi}\hat{G}^0(0^+, 0^-; \mathbf{Q}'), \tag{26}$$

where $\mathbf{Q} = (Q_x, Q_y)$ and $\mathbf{Q}' = (Q', Q_y)$, with the important values being those such that $|Q_x| \sim |Q'_x| \sim 1/|Z| \gg k_0$.

As above, from $\hat{G}^0$ in Eq. (26) we eliminate the functions d^0_{yy} and the terms linear in Q_y, which vanish after integration [Eq. (26)] owing to the translational symmetry of the model (19) with respect to the coordinate y. Since only the components J_{xy}, J_{yx}, J_{yz} and J_{zy} are nonzero, we obtain the following matrix elements:

$$\begin{pmatrix} l_{ss}^{(5)} \\ l_{ps}^{(5)} \end{pmatrix} = -B \begin{pmatrix} g_s^+(Q^f)\chi^{(x)} J_{xy} \sin\varphi^f \\ g_p^+(Q^f)\chi^{(x)} J_{xy} \cos\Theta^f \cos\varphi^f + g_p^-(Q^f)\chi^{(z)} J_{zy} \sin\Theta^f \end{pmatrix} g_s^+(Q^i), \tag{27}$$

and

$$\begin{pmatrix} l_{sp}^{(5)} \\ l_{sp}^{(5)} \end{pmatrix} = B \begin{pmatrix} g_s^+(Q^f)\cos\varphi^f \\ -g_p^+(Q^f)\cos\Theta^f \sin\varphi^f \end{pmatrix} \times [J_{yx}\chi^{(x)} g_p^+(Q^i)\cos\Theta^i - J_{yz}\chi^{(z)} g_p^-(Q^i)\sin\Theta^i], \tag{28}$$

in which

$$B = 4\pi i k_0^6 \varepsilon_0 l \chi^{(y)} \exp[i(Q^i - Q^f \cos\varphi^f)X]. \tag{29}$$

These matrix elements $l_{\lambda'\lambda}^{(n)}$, along with the elements found[7] in the absence of magnetization, completely determine the ellipsometric parameters of the scattered light, as well as the contributions to the transverse scattering cross section (17).

In order to estimate the near-field contributions to the field (6) and the observed quantities (17), we shall use the following estimate of the integrals (11) and (13): $U \sim V \sim (l/L)/k_0^2$. Since $D^0(\mathbf{r},\mathbf{R}) \sim 1/r$ for the scattered light, we find $E^{(3)} \sim E^{(4)} \sim (1/r)\varepsilon_0 l (k_0^3 \chi / k_0 L) E^0$ and $E^{(5)}/E^{(3)} \sim \chi/L^3$. In the general case of a metallic particle of characteristic size a, we have $\chi \sim a^3 \omega_p/\Gamma$ (Ref. 7), where the parameter $\omega_p/\Gamma \sim 10$ determines the ''quality factor'' of the possible plasma resonance in a particle at frequency ω_p. Thus, as it is small for sufficiently large $L \sim 1/k_0$, the field $E^{(5)}$ may be comparable in magnitude to $E^{(3)} \sim E^{(4)}$ when $L \geqslant a$.

In the near-field approximation with respect to the interaction of a particle with the surface, light-scattering by a particle in the presence of a magnetic inhomogeneity differs in a number of ways from light-scattering in the case of a uniformly magnetized medium. Thus, the field $\mathbf{E}^{(3)}$ is excited only by an s-polarized external wave, and the field $\mathbf{E}^{(4)}$, only by a p-polarized wave. A comparison of Eqs. (21)–(23) with the results of Ref. 7 shows that the component $\chi^{(y)}$ of the polarizability tensor for a surface particle is eliminated from these fields, since a transverse field is inefficient in the near-field processes $n=3$ and 4. At the same time, Eqs. (27) and (28), which describe the magnetooptical analog of the ''image force'' effect when $\mathbf{M} \| \mathbf{e}_z$, include $\chi^{(y)}$ through Eq. (29) for all the scattering channels with $n=5$. Light scattering due to the component of the polarizability $\chi^{(z)}$ normal to the surface is excluded at an angle of incidence $\Theta^i = 0$ for the field $\mathbf{E}^{(4)}$ and at a scattering angle $\Theta^f = 0$ for the field $\mathbf{E}^{(3)}$. The components $\mathbf{E}^{(3)} - \mathbf{E}^{(5)}$ of the field in Eq. (6) contain different combinations of the angles of incidence and scattering, which can also be used to isolate the different scattering channels. It is significant that, in the near-field approximation, the directional diagrams of the radiation, e.g., the dependences of the terms in Eq. (17) on the azimuthal angle φ^f, are simpler in form than their analogs in the general case discussed in Ref. 7 for a uniform magnetization $\mathbf{M} \| \mathbf{e}_z$.

From the standpoint of near-field microscopy, however, the major interest is in the dependence of observables, such as $d\sigma_{\lambda'\lambda}^{(n)}/d\Omega^f$, on the distance $|Z| + z_0$ between the particle and the magnetic inhomogeneity. In the model of Eq. (19), this dependence is determined by Eqs. (20) and (26), which enter Eq. (17) linearly. The quantities $L = |Z| + z_0 + \Delta$ or $|Z| + z_0$ in Eqs. (20) and (26) are a measure of the optical resolution: this can be justified in the usual scheme,[15] which considers functions of the form of the image (20) for two objects separated by a distance $\sim L$. Thus Eqs. (20) and (26) show that the optical resolution (minimum scale length for the observable spatial modulation in the magnetooptical image) is determined by the quantity $L = |Z| + z_0 + \Delta$, i.e., the resolution should improve as the separation $|Z| + z_0$ is reduced. At the same time, in view of the condition $|Z| \geqslant a$, even in the limit $z_0 \to 0$ and $\Delta \to 0$, the observable image size cannot be smaller than the characteristic particle size a, which is therefore the main factor limiting the optical resolution. The predicted enhancement in the resolution for the quantities $|d\sigma_{\lambda'\lambda}^{(n)}/d\Omega^f|$ with decreasing $|Z|$ correlates with experiment.[3,4] In these experiments it was noted that the image contrast (optical resolution) in a scanning magnetooptical microscope increases with decreasing distance between a particle and the surface of the magnetic medium, which consisted of a ferromagnetic superlattice. It is clear that when nonmagnetic surface microroughness or defects are present, this behavior should be observed in their optical resolution as well. This effect has been modeled numerically[8,9] and, in terms of our theory, it can be described analytically using instead of Eq. (18) a perturbation with the appropriate symmetry and spatial distribution of the dielectric constant.

CONCLUSION

An analytical theory of near-field scanning magnetooptical microscopy has been presented above for the case in which the polar magnetooptical Kerr effect occurs. It has been shown that in this optical arrangement, the scale length of the resolution of a magnetic inhomogeneity along the surface is limited by the size of the probe particle, while subwavelength resolution improves with decreasing relative distance between this particle and the magnetic inhomogeneity along the normal to the surface of the magnetic medium. The shape of the magnetooptical image in the surface plane also depends strongly on the distribution of the magnetization transverse to the surface. The specific features of magnetooptical scattering in the theory proposed above are related solely to the special (magnetooptical) symmetry of the perturbations in the dielectric tensor. Thus, when a perturbation of another type is used, this analytical theory can be generalized simply to describe near-field effects unrelated to the

magnetization. It might be expected that only the directional diagrams of the radiation would change in that case, while the above results regarding the optical resolution would remain unchanged.

This work was supported by the Russian Fund for Fundamental Research, Grant No. 96-02-17898.

The author thanks T. J. Silva for help discussions of a number of questions concerning the experiment.

[1] D. W. Pohl and D. Courjon (eds.), *Near Field Optics*, Kluwer Acad. Publ. (1993), 436 pp.
[2] E. Betzig, J. K. Trautman, R. Wolfe *et al.*, Appl. Phys. Lett. **61**, 142 (1992).
[3] T. J. Silva, S. Schultz, and D. Weller, Appl. Phys. Lett. **65**, 658 (1994); T. J. Silva, Doctoral Dissertation, Univ. of California, San Diego (1994), 263 pp.
[4] T. J. Silva and A. B. Kos, in *Near Field Optics*, edited by M. A. Paesler and P. J. Moyer, Proc. Soc. Photo-Opt. Instrum. Eng. (SPIE) **2**, 2535 (1995).
[5] V. I. Safarov, V. A. Kosobukin, C. Hermann *et al.*, Phys. Rev. Lett. **73**, 3584 (1994); Microsc. Microanal. Microstruct. **5**, 381 (1994); Ultramicroscopy **57**, 270 (1995).
[6] V. A. Kosobukin, in *Near Field Optics*, edited by M. A. Paesler and P. J. Moyer, Proc. Soc. Photo-Opt. Instrum. Eng. (SPIE) **9**, 2535 (1995).
[7] V. A. Kosobukin, Fiz. Tverd. Tela **39**, 560 (1997) [Phys. Solid State **39**, 488 (1997)].
[8] N. Garcia and M. Nieto-Vesperinas, Opt. Lett. **20**, 949 (1995); R. Carminati and J.-J. Griffet, Opt. Commun. **116**, 316 (1995).
[9] A. A. Maradudin, A. Mendoza-Suarez, E. R. Mendez, and M. Nieto-Vesperinas, in *Optics at the Nanometer Scale*, edited by M. Nieto-Vesperinas and N. Garcia, Kluwer Acad. Publ. (1996), p. 41.
[10] A. K. Zvezdin and V. A. Kotov, *Thin Film Magnetooptics* [in Russian] (Moscow, 1988), 190 pp.
[11] A. A. Maradudin and D. L. Mills, Phys. Rev. B **11**, 1392 (1975); Phys. Rev. B **12**, 2943 (1975).
[12] A. M. Brodskiĭ and I. M. Urbakh, *Electronics of the Metal/Electrolyte Interface* [in Russian] (Moscow, 1989), 296 pp; Fiz. Tverd. Tela (St. Petersburg) **35**, 884 (1993) [Phys. Solid State **35**, 457 (1993)]; Fiz. Tverd. Tela (St. Petersburg) **36**, 3015 (1994) [Phys. Solid State **36**, 1605 (1994)].
[13] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continous Media*, 2nd ed. [Pergamon Press, Oxford (1984); Nauka, Moscow (1982), 620 pp.].
[14] R. Allenspach, M. Stampanoni, and A. Bischof, Phys. Rev. Lett. **65**, 3344 (1990); R. Allenspach and A. Bischof, Phys. Rev. Lett. **69**, 3385 (1992).
[15] M. Born and E. Wolf, *Principles of Optics* [Pergamon, London (1967); Mir, Moscow (1970), 856 pp.].

Translated by D. H. McNeill

Extracting microwave energy from a cavity by mode conversion at a coupling window

S. N. Artemenko, V. A. Avgustinovich, and Yu. G. Yushkov

Scientific-Research Institute of Nuclear Physics, Tomsk Polytechnical University, 634050 Tomsk, Russia
(Submitted January 8, 1997)
Zh. Tekh. Fiz. **68**, 92–96 (July 1998)

A study is made of the production of high power nanosecond rf pulses by extracting microwave energy from an oversized cavity by means of conversion, at a coupling window, of the high-Q working mode to an auxiliary mode which is strongly coupled to an external load. It is shown that microwave rf pulse compressors with copper storage cavities and energy extraction by mode conversion at a coupling window can provide gains of 5–13 dB with output signal durations of 20–150 ns and peak powers of 5–10 MW in the 3-cm band and 50–100 MW in the 10-cm band. Rf pulses lasting 30 ns with peak powers of 0.5 MW have been obtained experimentally at a frequency of 9.4 GHz with a gain of 9 dB. © *1998 American Institute of Physics.* [S1063-7842(98)01607-9]

INTRODUCTION

The extraction of energy from a cylindrical oversized cavity through conversion of an H_{01n} mode into an H_{11p} mode which is strongly coupled to a external load has been reported.[1,2] Energy was extracted through a circular output waveguide which was smaller than the cutoff for the H_{01} mode but larger than cutoff for the H_{11} mode, and was attached coaxially to the cavity at one of its ends. Conversion was achieved using an electrical spark gap positioned in the hollow of the cavity at the maximum of the φ component of the electric field of the working mode parallel to the field lines. The feasibility of microwave rf pulse compressors was demonstrated with energy extraction by mode conversion, and gains of 13 dB with a copper storage cavity and 30 dB with a superconducting cavity were attained for output pulse durations of $\approx 20-50$ ns.

Meanwhile, the microwave compressors described in Refs. 1 and 2 have two major disadvantages, due to the placement of the spark gap directly in the hollow of the cavity. First, the Q of the cavity falls off during the field buildup because the working mode is converted on the structural elements of the spark gap device into other waves which radiate into the load through the output waveguide, and second, this arrangement reduces the electrical breakdown strength of the apparatus because these elements are located at the site where the electric component of the rf field is highest. The drop in Q reduces the gain of the compressor, while the lower electrical breakdown strength makes it harder to operate the device at a sufficiently high power level.

At the same time, energy extraction based on mode conversion is one of the few methods that have been realized in practice and can, under certain conditions, compete with the best-known and most often used method, which is based on extraction through an interference switch.[3,4] Compared to systems with an interference switch, one of the undoubted advantages of energy extraction by mode conversion is a less dense eigenmode spectrum of their storage elements and, when there are no spark gap elements in the cavity, the systems of the latter type can work at higher power levels. The first point is related to the presence of an output waveguide in cavities of this type, radiation through which causes dilation of the spectrum, and the second originates in the larger cross sectional area of this waveguide compared to the cross section of the waveguide in an interference switch.

In addition, with a mode-conversion compressor one can construct a ''hybrid'' device by attaching an interference switch to its output waveguide. In this case, energy can be extracted either by mode conversion with the switch closed, opening it only after energy has been transferred from the cavity working mode to the auxiliary mode, or by using a transition process in a system of two stationary coupled modes, as has been done[5] in a system of two coupled cavities. Here there is no longer a need for very strong coupling between the chosen modes and, therefore, less demanding specifications are imposed on the intermode conversion device (spark gap, coupling window, etc.). Furthermore, even in this case a compressor can operate at a higher power level than an ordinary system with an interference switch, since in such a device a high rf electric field will be across the switch for only a relatively short time, equal to the time to transfer energy from the cavity working mode to the auxiliary mode ($\sim 10-100$ ns).

For these reasons, there is definite interest in searching for effective methods and devices for rapid initiation of strong intermode coupling in the cavities of mode-conversion compressors, and for techniques and devices which will conserve the high electrical characteristics of the cavity during field buildup and ensure the required degree of coupling during extraction.

In this paper we study a method for rapidly initiation of strong mode coupling based on changing the intermode interaction coefficient at the coupling window between the cavity and a shorted waveguide stub as the length of the stub is varied. The efficiency of this method is evaluated. It is tested experimentally on a microwave compressor apparatus operating in the 3-cm band.

ESTIMATE OF THE INTERMODE COUPLING COEFFICIENT AT A WINDOW

The intermode coupling coefficient $\gamma_{1,2}$ for modes with the same resonance frequency f can be estimated using the well known formula[6]

$$\gamma_{12}=-\int (\mathbf{H}_1\mathbf{H}_2-\mathbf{E}_1\mathbf{E}_2)dV/V, \tag{1}$$

where $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{E}_1$, and $\mathbf{E}_2$ are the magnetic and electric vectors of the rf fields of the interacting modes and V is the cavity volume.

The integral in Eq. (1) is taken over the volume δV of the ''deformed'' part of the cavity within which the interaction takes place.

For an interaction at a coupling window whose characteristic dimensions are much smaller than the cavity dimensions and is located, for example, at the maximum of the rf field magnetic components, Eq. (1) can be replaced by the approximate expression

$$\gamma_{12}\approx \delta V H_{10}H_{20}/V, \tag{2}$$

where H_{10} and H_{20} are the magnetic field strengths of the fields at the center of the window.

On the other hand, besides inducing an intermode interaction, the ''deformed'' part of the cavity is known[6] to cause a relative drift in the frequency of the oscillations, $\sigma_{1,2}$, given by

$$\sigma_{1,2}=\delta f_{1,2}/f=2\int (\mathbf{H}_{1,2}^2-\mathbf{E}_{1,2}^2)dV/V, \tag{3}$$

in which, as in Eq. (1), the integral is taken over the volume δV. Then, we find for this coupling window that

$$\delta V\approx \delta f_{1,2}V/2\, fH_{10,20}^2. \tag{4}$$

Equation (2) can, therefore, be reduced to the form

$$\gamma_{12}\approx \delta f_{1,2}H_{10}H_{20}/2\, fH_{10,20}^2. \tag{5}$$

Then, since it is necessary to put the window where H_{10} and H_{20} are not only maximal or close to maximal, but also comparable, for efficient interaction of the modes, we can finally write Eq. (5) in the form

$$\gamma_{12}\approx \delta f/2\, f, \tag{6}$$

where $\delta f\approx \delta f_1\approx \delta f_2$.

We shall determine δf for a cavity coupled to a shorted waveguide stub using the scattering matrix method.[7] For this, we represent the device under study in the form of coupled waveguide segments, as shown in Fig. 1, where *1* is the input waveguide, *2* is the cavity, and *3* is the stub, while k and h are parameters characterizing the coupling of the cavity to the input waveguide and stub, respectively, $a_1\ldots a_4$ are the amplitudes of the incident waves, and $b_1\ldots b_4$ are the amplitudes of the reflected waves. Then, according to the method we have chosen, we can write the following equations for the wave amplitudes in the system:

$$\begin{vmatrix} a_1 \\ a_2 \end{vmatrix} = \begin{vmatrix} -\sqrt{1-k^2}jk \\ jk-\sqrt{1-k^2} \end{vmatrix} \begin{vmatrix} b_1 \\ b_2 \end{vmatrix}, \quad \begin{vmatrix} a_3 \\ a_4 \end{vmatrix} = \begin{vmatrix} -\sqrt{1-h^2}jh \\ jh-\sqrt{1-h^2} \end{vmatrix} \begin{vmatrix} b_3 \\ b_4 \end{vmatrix},$$

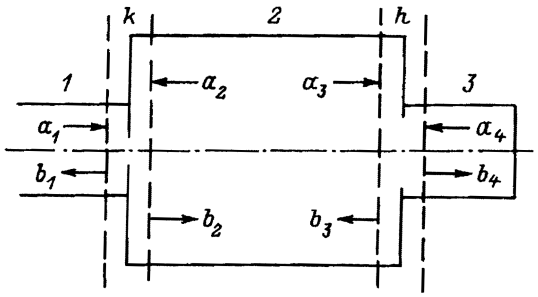


FIG. 1. Sketch of a cavity coupled to a short circuited stub.

$$a_2=b_3\exp(-\alpha-j\varphi_c),\quad a_3=b_2\exp(-\alpha-j\varphi_c),$$
$$a_4=-b_4\exp(-2\beta-2j\psi), \tag{7}$$

where α and β are the damping constants of the waves in the cavity and stub section when they have the same free path, and φ_c and ψ are the additional phase shifts of the waves in the cavity and stub owing to the difference between the working frequency of the system and the resonance frequencies of the cavity and stub.

Here ψ is set by the choice of the stub section length, while φ_c is given by

$$\varphi_c\approx \varphi_1+\varphi_2, \tag{8}$$

where $\varphi_1=\arctan(h/\sqrt{1-h^2})/2$ is the phase shift induced by the effect of the coupling window on the cavity frequency, while φ_2, as can easily be shown using Eq. (7), is given by the approximate formula $\varphi_2\approx\arctan[h^2\exp(-2\beta)\sin(2\psi)/(1-2\exp(-2\beta)\cos(2\psi)+\exp(-4\beta))]/2$ and is the phase shift owing to the effect of the shorted stub. (We neglect the effect of the input window.)

It can also be shown that

$$\varphi_c=\pi\delta f T, \tag{9}$$

where T is twice the transit time of the wave along the cavity.

Thus, from Eqs. (6) and (9), we find

$$\gamma_{12}\approx|\varphi_c|/2\pi fT=|\varphi_c|/\omega T. \tag{10}$$

If, further on, as a model of the interacting modes we consider a system of two coupled cavities with a coupling parameter m instead of a system of two coupled circuits, as in Ref. 6, then we find that m is expressed in terms of γ_{12} by

$$m=\omega T\gamma_{12}. \tag{11}$$

Therefore, for the parameter m characterizing the magnitude of the intermode coupling at the window, from Eqs. (10) and (11) we ultimately obtain

$$m\approx|\varphi_c(h,\beta,\psi)|. \tag{12}$$

This last expression allows us to use the results of Artemenko[8] for estimating the efficiency of intermode coupling. According to Ref. 8, energy transfer from the working mode to the external load becomes efficient when the intermode coupling becomes as efficient as the coupling between the auxiliary mode and the load. This means that, for efficient extraction of the energy by mode conversion, m must

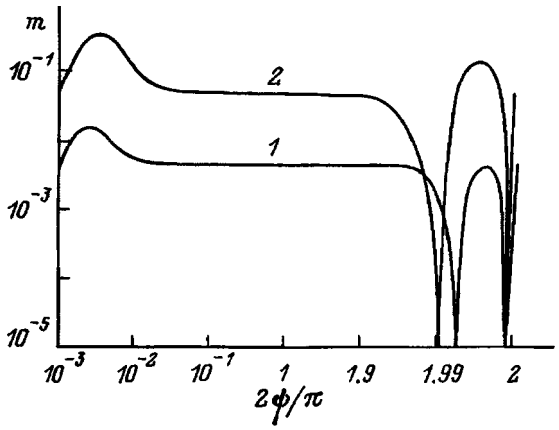


FIG. 2. The coupling parameter of the cavity working mode and the auxiliary mode as a function of the electrical length of the stub.

be at least comparable to the parameter $q \approx 2\sqrt{\alpha\beta_{\text{out}}}$, which characterizes the energy loss of the auxiliary mode to radiation through the output waveguide, where β_{out} is the coupling coefficient for this mode to the load.

Figure 2 shows plots of m as a function of ψ when $\beta = 5\times10^{-3}$ for $h = 1.0\times10^{-2}$ and 1.0×10^{-1} (curves *1* and *2*). These graphs show that m depends significantly both on the coupling h of the cavity to the stub and on the additional phase shift ψ (from the stub length). Regardless of the value of h, there are always two values of ψ at which the intermode coupling parameter m equals zero. At the same time, a stub length can be chosen such that m is ~ 0.1 or greater. Thus, it is evident that if the energy builds up over a stub length corresponding to $m = 0$ and it is then changed rapidly with a commutator (spark gap), then this method can be used to ensure rapid initiation of intermode coupling. (According to Ref. 8, the criterion for strong intermode coupling is $m \gg 4\sqrt{\alpha\alpha_1}$, where α_1 is the damping constant for the auxiliary wave mode in a single pass of the cavity).

ESTIMATE OF THE CHARACTERISTICS OF A MICROWAVE COMPRESSOR WITH MODE CONVERSION AT A COUPLING WINDOW

Using the above results, we now estimate the possible operating characteristics of microwave compressors for the 3- and 10-cm bands and investigate a method for energy extraction.

It is easy to show that for the typical intrinsic Q factors of an oversized copper cylindrical cavity with H_{01n} mode oscillations, $Q_{01} \approx 10^5$, and double transit times of the wave along the cavity, $T \approx 1-3$ ns, the damping constant α for the H_{01} wave is close to 10^{-4}. Thus, for an auxiliary mode with an intrinsic $Q_{02} \approx 5\times10^4$ and a load $Q_{l2} \approx 2\times10^3$, its coupling parameter q with the external load, will be of the order of 0.1–0.2.

It follows, further, from Eq. (9) that for a frequency drift $\delta f \approx 10-20$ MHz, corresponding to the actually attainable coupling of a cavity with a stub $h \approx 0.1-0.2$, and for $T \approx 1-3$ ns, the phase shift φ_c and, therefore, the parameter m can attain values of $\sim 0.05-0.1$ comparable to q. Because of this, the coupling window between the cavity and stub can serve as an effective energy transfer element from the cavity working mode to the auxiliary mode and from the auxiliary mode to the load. Here the gain of the compressor will be given by[8]

$$M^2 \approx (1 - \pi q^2/4m) M_0^2, \tag{13}$$

where $M_0^2 = q^2/4\alpha$ is the gain coefficient of a compressor with an interference switch.

Equation (13) implies that M^2 can be as high as $(0.8-0.9)M_0^2$.

For these values of m and q, energy extraction with mode conversion should,[8] take place with a sinusoidal modulation in the envelope of the output signal owing to the successive transfer of energy from the working cavity mode to the auxiliary mode and back. However, since the transfer period for m close to q becomes comparable to the damping constant τ of the signal ($T/m \approx T/8\alpha\beta_{\text{out}} = \tau$), oscillations in the envelope of the output pulse may be absent except for the first ''burst'' corresponding to the transfer of energy from the cavity mode to the auxiliary mode. The reverse process will not have time to occur in this case.

The gain coefficients of microwave compressors with mode conversion at the coupling window ($M^2 \leqslant q^2/4\alpha$) and their output signal durations ($t_p \geqslant T/m$) are estimated to be $\approx 5-13$ dB and $\approx 20-150$ ns in the 3- and 10-cm bands. The maximum output power level will be determined by the electrical breakdown strength of the window and the waveguide stub, and with gas insulation (e.g., sulfur hexafluoride) at excess pressure, giving a factor of three extra breakdown strength, it can reach $\approx 5-10$ MW in the 3-cm band and $\approx 50-100$ MW in the 10-cm band.

EXPERIMENTAL RESULTS

Experimental studies were conducted on a 3-cm band system with a cavity of diameter 90 mm and length 213.5 mm operating at a frequency of 9.4 GHz with $H_{01(02)}$ modes. The intrinsic Q of the cavity was 1.1×10^5. The system was excited through a coupling window in one of the end caps of the cavity at the site of the maximum in the H_r-th component of the rf field of the cavity working mode. A coupling window between the cavity and a shorted stub made of a standard rectangular waveguide with a transverse cross section of 28.5×12.6 mm was placed in the same end cap of the cavity, symmetrically opposite the input window relative to the center of the end cap.

An H-tee with a shorted half-wave side arm and a commutator (electrical spark discharge) mounted in it was mounted in the waveguide stub to study the effect of its length on the magnitude of the intermode coupling, as shown in Fig. 3, where *1* is the input waveguide, *2* is the tee, *3* is the commutator, *4* is the cavity, and *5* is the output waveguide. The half wavelength of the side arm provided a decoupling of about 45 dB between the inlet and output arms during buildup. This made it possible to avoid breakdown in the output arm of the tee as energy built up because of illumination of the waveguide tract by sparking on the plunger that shorts this arm. In addition, this design made it possible to follow the dynamics of the intermode interaction as the

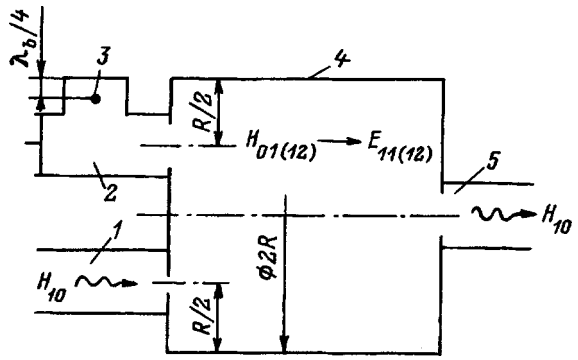


FIG. 3. Sketch of a resonant system for experimentally extracting energy by conversion of the cavity working mode at a coupling window.

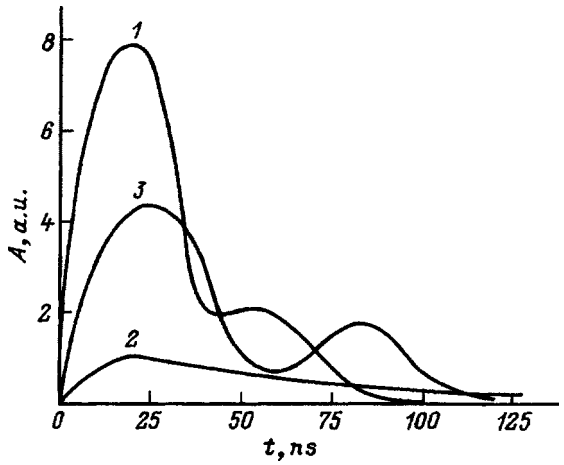


FIG. 4. Envelopes of rf output pulses for stubs with different electrical lengths.

length of the stub was changed, without having to reassemble it. The length of the input arm of the tee was chosen to be close to the half wavelength and to be such that at the working wavelength, the radiation of energy to the load through the output waveguide was minimal. In the system used in the experiments, the level of this radiation was no higher than the typical levels for an interference switch ($\sim$45 dB). As an output waveguide we used a standard rectangular waveguide with a cross section of 28.5×12.6 mm connected coaxially to the cavity on its other end cap, as shown in Fig. 3, and with its wide wall parallel to the wide walls of the input and stub waveguides. The dimensions of the oval coupling windows between the resonator and the stub and output waveguides were 12.6×16 and 12.6×18 mm, respectively.

The system was powered by a magnetron generator with a pulsed output power of $\approx$60 kW and pulse duration of $\sim1\ \mu$s. The operating regimes were switched using the plasma of a microwave discharge in air at atmospheric pressure, initiated by delivering a high-voltage control signal to the commutator. The spark gap response time was about 2 ns. The relative frequency detuning $\Delta f/f$ of the working cavity mode and of the $E_{11(12)}$ (auxiliary) mode degenerate with it, which was strongly coupled to the output waveguide, was less than $5\times10^{-4}-10^{-3}$ during switching, which means that the frequencies of the interacting modes are essentially the same[9] for a intermode coupling coefficient $\gamma_{12}\approx10^{-3}$ ($m\approx0.1$).

After operation of the switch, rf pulses with the envelopes shown in Fig. 4 were recorded at the system outlet. Curve *1* corresponds to a stub length for which the intermode coupling is close to maximal ($\psi\approx\pi/2-h$), curve *2*, to a length at which there is almost no coupling between the modes ($\psi\approx\pi-h$), and curve *3*, to an intermediate case. It can be seen that the experimental results are in good qualitative agreement with the theoretical estimates of this paper and Ref. 8. The maximum measured gain coefficient for this compressor was 9 dB with a peak output signal power $\approx$0.5 MW and pulse duration of 30 ns at a level of 0.5. The intermode coupling parameter m for these rf output pulse characteristics is estimated to be $\geqslant0.05$. The efficiency with which the energy of the compressed pulse was transferred to the load was $\sim0.2-0.25$ in the case of unoptimized buildup in the cavity.

CONCLUSION

In this paper we have provided a justification for and experimentally validated the feasibility of obtaining high power nanosecond rf pulses through the buildup of microwave energy in an oversized cavity followed by rapid extraction through conversion, at a coupling window, of the high-Q cavity working mode into an auxiliary mode that is strongly coupled to the external load. Our results give grounds for hope that rf microwave compressors with energy extraction by mode conversion may turn out to be sufficiently promising devices with characteristics close to those of compressors using interference switches for energy extraction.

[1] S. N. Artemenko, V. L. Kaminskiĭ, and Yu. G. Yushkov, Pis'ma Zh. Tekh. Fiz. **7**, 1529 (1981) [Sov. Tech. Phys. Lett. **7**, 656 (1981)].
[2] S. N. Artemenko, V. L. Kaminskiĭ, and Yu. G. Yushkov, Zh. Tekh. Fiz. **53**, 1885 (1983) [Sov. Phys. Tech. Phys. **28**, 1163 (1983)].
[3] R. Alvarez, D. Birks, D. Bern *et al.* [in Russian], Atomnaya Tekhnika za Rubezhom, Vol. 11 (1982) pp. 36–39.
[4] V. A. Avgustinovich, L. Ya. Avgustinovich, S. N. Artemenko, and Yu. G. Yushkov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **30**(2), 90 (1987).
[5] V. I. Ivannikov, Yu. D. Chernousov, and I. V. Shebolaev, Zh. Tekh. Fiz. **65**(5), 194 (1995) [Tech. Phys. **40**, 514 (1995)].
[6] V. B. Shteĭnshleĭger, *Wave Interaction Phenomena in Electromagnetic Cavities* [in Russian] (Oborongiz, Moscow, 1955), 112 pp.
[7] J. L. Altman, *Microwave Circuits* [Van Nostrand, Princeton, N.J. (1964), 462 pp; Mir, Moscow (1968), 488 pp.].
[8] S. N. Artemenko, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **30**, 1289 (1987).
[9] S. N. Artemenko and V. L. Kaminskiĭ, Zh. Tekh. Fiz. **59**(11), 161 (1989) [Sov. Phys. Tech. Phys. **34**, 1332 (1989)].

Translated by D. H. McNeill

Magnetostatic volume waves in exchange-coupled ferrite films

S. L. Vysotskiĭ, G. T. Kazakov, and Yu. A. Filimonov

Saratov Branch of the Institute of Radio Engineering and Electronics, Russian Academy of Sciences, 410019 Saratov, Russia

A. V. Maryakhin

Scientific-Research Institute of Materials for Electronics Technology, 410026 Kaluga, Russia
(Submitted March 31, 1997)
Zh. Tekh. Fiz. **68**, 97–110 (July 1998)

The influence of exchange coupling of layers on the propagation of magnetostatic dipole volume waves in normally and tangentially magnetized two-layer epitaxial ferrite structures is investigated. It is shown that the indicated influence is manifested in the form of dynamic spin pinning effects on the interlayer boundary and formation of a common dipole-exchange wave spectrum for the entire structure. In this case, at the synchronism frequencies of the dipole and exchange waves the losses of the dipole waves grow and anomalous segments appear in the dispersion. In films magnetized in the ''hard'' direction relative to the axis of normal uniaxial surface anisotropy the magnetostatic dipole volume waves can interact resonantly with the surface spin waves supported by the boundaries with pinned spins. © *1998 American Institute of Physics.* [S1063-7842(98)01707-3]

INTRODUCTION

The use of multilayer ferrite structures as waveguides for magnetostatic waves (MSWs) is of interest from the point of view of extending the possibilities of controlling their dispersion and damping, and of forming new types of spin-wave excitations.[1–3] These possibilities are realized most fully in those cases in which the frequency ranges of the existence of magnetostatic waves in individual layers are sufficiently similar and regions of degeneracy of the spectra of isolated films arise, where these isolated films are ferrite films comprising a multilayer structure and being in essence coupled waveguides for the magnetostatic waves. Neglecting the magnetoelastic interaction,[1)] one usually distinguishes two main mechanisms of coupling of film waveguides: via dipole fields[4–6] and via the exchange interaction at the film boundaries.[7,8] In the case when the relation $A_{12} \gtrsim D$ is fulfilled between the interlayer exchange constant A_{12} and the surface anisotropy constant D, both of the indicated mechanisms play a significant role in the formation of the spin-wave excitation spectrum.[9,10] In this case the influence of exchange coupling of layers on the propagation of magnetostatic waves is most distinctly manifested at the synchronism frequencies of the dipole and exchange waves of the structure and is associated, first, with the appearance of dynamic spin pinning on the interlayer boundary,[11] and second, with the formation of a common exchange wave spectrum of the structure accompanied by the appearance of ''repulsion'' [12] of the spin-wave resonance (SWR) modes of the layers in the degeneracy region. The aim of the present work is the study the influence of the interlayer exchange on the properties of the magnetostatic forward (MSFVW) and backward (MSBVW) volume waves in two-layer ferrite films.

Note that the influence of interlayer exchange on the propagation of magnetostatic surface waves (MSSWs) in a two-layer ferrite structure was considered in Refs. 13–15. It was shown there that dynamic spin pinning caused by exchange coupling, like surface anisotropy in the case of isolated films,[16–19] leads to resonant growth of losses and to the appearance of anomalous segments of the dispersion at the SWR frequencies. It may be expected that in the case of MSFVWs and MSBVWs propagating in exchange-coupled ferrite films, exchange coupling will lead to analogous effects. In this case, in the interpretation of the experiments it is important to distinguish the contributions to the changes in the dispersion and damping of the magnetostatic waves due to the parameters D and A_{12}. For this reason we have placed special emphasis on this aspect in the present work.

We also discuss the possibility of resonant interaction of magnetostatic dipole volume waves with surface spin waves, which can exist near the boundaries of films with pinned surface spins[20,21] or at an interlayer boundary in the case of interlayer exchange of antiferromagnetic type: $A_{12}<0$ (Ref. 7). Since the existence of such surface waves is possible only for a certain type of surface anisotropy, such a formulation of the problem can be of interest for diagnostics of surface spin states in a multilayer structure.

RESULTS OF CALCULATIONS

Let a two-layer ferrite structure be oriented such that the z axis is perpendicular to the surface of the structure, and the boundary between the films coincides with the xy plane. The films are characterized by their thickness d_l, saturation magnetization $4\pi M_{0l}$, exchange stiffness A_l, line width of the ferromagnetic resonance ΔH_l, where $l=1,2$ is the number of the layer, and the layer $l=2$ is located in the half space $z<0$. We characterize the exchange coupling at the boundary $z=0$ by the interlayer exchange parameter A_{12}. The

films are taken to be unbounded in the xy plane and homogeneous and isotropic and to possess a normal uniaxial surface anisotropy characterized by the surface anisotropy constant D_{li}, $i=1,2$, where the index $i=1,2$ corresponds to the lower and upper surfaces of the lth film. We consider waves propagating along the x axis.

MSFVW GEOMETRY

We direct the external magnetic field H_0 along the z axis. Following Ref. 9, we write the combined solution of the magnetostatic equation and the Landau–Lifshitz equation in layers in the form

$$h_{xl}=\left(\sum_{j=1}^{6} B_j \exp iK_jz\right)_l T,$$

$$h_{zl}=\left(\sum_{j=1}^{6} \frac{K_j}{q} B_j \exp iK_jz\right)_l T,$$

$$m_{yl}=\left(\sum_{j=1}^{6} \chi_j B_j \exp iK_jz\right)_l T,$$

$$m_{xl}=\left(\sum_{j=1}^{6} \alpha_j B_j \exp iK_jz\right)_l T, \tag{1}$$

where $T=\exp i(qx-\omega t)$, B_{jl} are unknown amplitudes, K_{jl} are the roots of the characteristic equation for the lth film

$$(p^3+ap^2+bp+c)_l=0, \tag{2}$$

where $p=K^2+q^2$, and the coefficients a,b,c have the form

$$a=2\omega_H/\omega_{ex}, \quad b=(\omega_H^2-\omega^2+q^2\omega_m\omega_{ex})/\omega_{ex}^2,$$

$$c=q^2\omega_m\omega_H/\omega_{ex}^2, \tag{3}$$

where $\omega_H=\gamma(H_0-4\pi M_0)$, $\omega_m=\gamma 4\pi M_0$, $\omega_{ex}=2\gamma A/M_0$, and γ is the gyromagnetic ratio.

We take the coefficients χ_j and α_j entering into the expressions for the components of the magnetization of the lth film to be in the form

$$\chi_j=\frac{i\omega\gamma M_0}{\tilde{\omega}_H^2-\omega^2}, \quad \alpha_j=\frac{\tilde{\omega}_H\gamma M_0}{\tilde{\omega}_H^2-\omega^2},$$

$$\tilde{\omega}_H=\omega_H+\omega_{ex}(K_j^2+q^2). \tag{4}$$

In expressions (3) and (4) we have dropped the subscript l, indicating that the corresponding coefficient belongs to the lth film.

The expressions for the fields in the regions $z>d_1$ and $z<-d_2$ have the form

$$h_x^3=B_3 \exp(-qz)T, \quad h_z^3=ih_x^3|_{z>d_1},$$

$$h_x^4=B_4 \exp(qz)T, \quad h_z^4=-ih_x^4|_{z<-d_2}. \tag{5}$$

We assume that the conditions of continuity of the tangential fields h_x and normal components of the magnetic induction $b_z=h_z+4\pi m_z$ are fulfilled at the boundaries of the structure at $z=d_1,0,-d_2$

$$h_x^3=h_x^1, \quad b_z^3=b_z^1|_{z=d_1},$$

$$h_x^1=h_x^2, \quad b_z^1=b_z^2|_{z=0},$$

$$h_x^2=h_x^4, \quad b_z^2=b_z^4|_{z=-d_2}. \tag{6}$$

We write the exchange boundary conditions in the form characteristic for normal uniaxial surface anisotropy,[21,9]

$$\frac{\partial m_1}{\partial z}+L_{11}m_1=0|_{z=d_1},$$

$$\frac{\partial m_2}{\partial z}-L_{22}m_2=0|_{z=-d_2},$$

$$\frac{\partial m_1}{\partial z}-L_{12}m_1-\frac{A_{12}}{A_1}\left(m_1-\frac{M_{01}}{M_{02}}m_2\right)=0\Big|_{z=0},$$

$$\frac{\partial m_2}{\partial z}+L_{21}m_2+\frac{A_{12}}{A_2}\left(m_2-\frac{M_{02}}{M_{01}}m_1\right)=0\Big|_{z=0}, \tag{7}$$

where the parameters $L_{li}=D_{li}/A_l$ characterize spin pinning on the upper $(i=1)$ and lower $(i=2)$ surface of the lth film, and $m_l=m_{xl},m_{yl}$.

The dispersion equation for the waves was derived from the condition of compatibility of system of equations (6), (7) upon substitution of the corresponding expressions for the fields and magnetizations with amplitudes B_{jl}, B_3, B_4 not equal to zero. In our case this reduces to finding the conditions of vanishing of a 14th-order determinant, which was done numerically. In the calculations of the dispersion curves, losses were not taken into account, the wave number q of the magnetostatic waves was prescribed, and the frequency ω corresponded to the roots of the dispersion equation. In the calculations of wave damping, losses were taken into account by the standard substitution $\omega_H\rightarrow\omega_H-i\gamma\Delta H$. The frequency of the magnetostatic waves in this case was taken to be assigned by a generator, and the values of the wave number $q=q'+iq''$ corresponded to the roots of the determinant, where the real part q' and the imaginary part q'' of the wave number determine respectively the dispersion law and the MSW losses. In the choice of the values of the wave number we chose those values that were closest to the values of q of the fundamental mode of the volume waves.

Note that in experiments it is possible to judge the nature of the losses of magnetostatic waves from the form of the frequency response characteristic (FRC) of a model of delay line type. If losses to conversion of the microwave signal to magnetostatic waves and the effect of ''direct'' induction between transformers are not taken into account in the calculations of the FRC, then the form of the FRC is given by

$$P=-8.68q''S, \tag{8}$$

where P is the level of the output signal in dB and S is the distance between the transformers in the model.

Since in experiments performed using standard radio measuring apparatus (such as an FK2-18 phase difference and attenuation meter) it is possible to record signal levels not lower than -60 dB, in the discussion of calculated results segments of the FRC with levels below -50 dB are

TABLE I. Parameters of the investigated structures.

Sample No.	Layer l	$4\pi M_0$, Gs	d, μm	A,10^{-7} erg/cm
1	1	400	13.8	1
	2	700	12.2	1.2
2	1	1650	4.92	3.8
	2	800	17	2.2
3	1	640	6	2
	2	1750	8	3.85

not considered. In normally magnetized two-layer films such segments of the FRC correspond to segments of the spectrum of spin-wave modes that coincide with the frequency region of existence of dipole MSFVWs in isolated layers[5–7]

$$\omega_{Hl} \leqslant \omega \leqslant \omega_{0l}, \tag{9}$$

where ω_{Hl} and $\omega_{0l}=\sqrt{\omega_{Hl}(\omega_{Hl}+\omega_{ml})}$ are respectively the long-wavelength $(q\to 0)$ and short-wavelength $(q\to\infty)$ boundaries of the MSFVW spectrum of the lth film.

In the calculations of the dispersion curves and the FRC we will return to the structures with the parameters indicated in Table I. We take the damping parameters and the distance between transformers to be equal to $\Delta H_1=\Delta H_2=0.2$ Oe and $S=4$ mm. We are interested in the behavior of the dispersion and losses of the waves as functions of the spin pinning parameters L_{li} and interlayer exchange parameter A_{12}. We assume that the surface anisotropy constants in the ferrite films are characterized by values $|D_{li}|\leqslant 0.1$ erg/cm^2 (Ref. 22). In this case the pinning parameters of the surface spins $L_{li}=D_{li}/A_l$ take values $|L_{li}|\leqslant 10^6$ cm^{-1}.

Figures 1, 2, 3, and 4 display, respectively, for the frequency intervals (9) the results of calculations of the FRC and dispersion laws of the spin-wave modes in structure 1 for a field $H_0=800$ Oe and different values of the parameters L_{li} and A_{12}. It can be seen that the FRC has two regions of signal transmission, denoted as *I* and *II*, which correspond to segments of the spectrum occupied by MSFVWs in isolated layers.

In the absence of interlayer exchange $A_{12}=0$ and for the pinning parameters equal to $L_{11}=L_{22}=0$, $L_{21}=L_{12}=-5\times 10^4$ cm^{-1} the FRC in regions *I* and *II* has the form typical of isolated films with one-sided spin pinning[24,25] (Fig. 1a). This is indicated, first, by the presence of deep ''dips,'' whose frequency positions are well described by the expression for the spin-wave resonance (SWR) frequencies in a normally magnetized film[21]

$$f_N=f_H+f_{ex}Q_N^2, \tag{10}$$

where $f=\omega/2\pi$, and $Q_N=(\pi N)/d$ is the wave number of the spin wave at the Nth SWR frequency. Second, the depth of the dips in each of regions *I* and *II* grows monotonically with growth of the MSFVW frequency.

It can be seen from a comparison of the form of regions *I* and *II* of the FRC and the corresponding segments of the spectrum in Fig. 4 that for $A_{12}=0$ in region *II* of the FRC the dips are noticeable only at the frequencies of intersection of

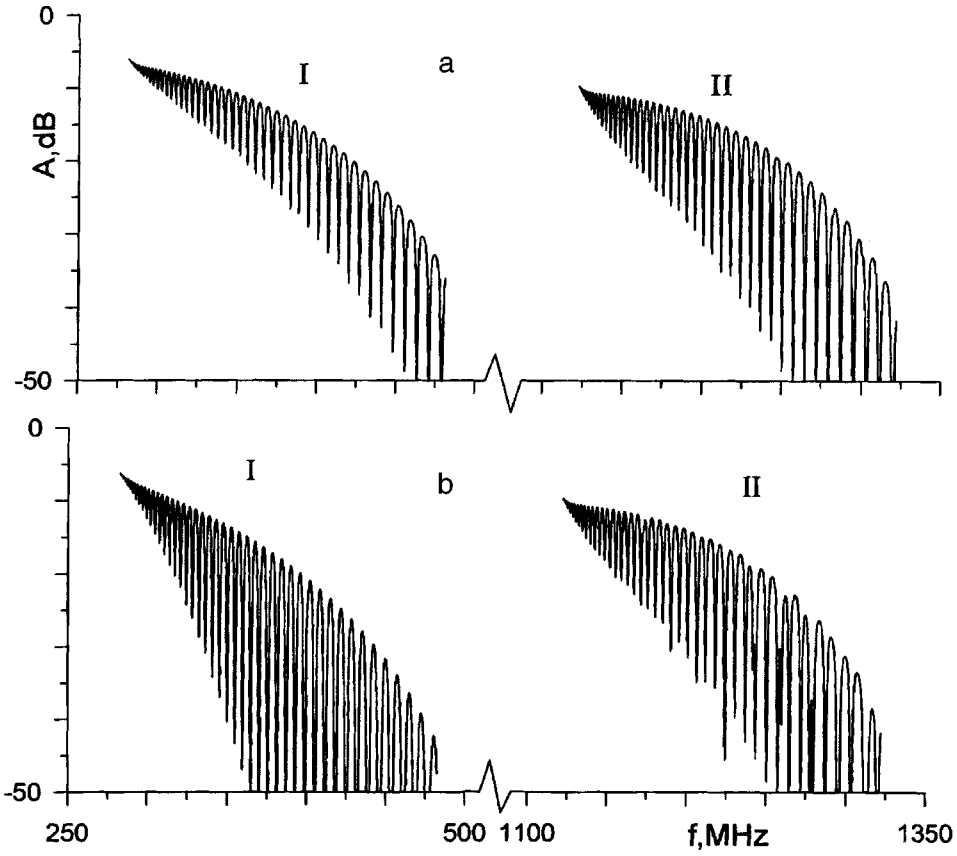


FIG. 1. FRC of a model of delay line type for the case of propagation of MSFVWs in the absence of exchange coupling in free spins on the outer surfaces of the structure. Magnitude of the spin pinning on the inner surfaces: a — $L_{21}=L_{12}=-5\times 10^4$ cm^{-1}, b — $L_{12}=-5\times 10^5$ cm^{-1}; $L_{21}=-5\times 10^4$ cm^{-1}.

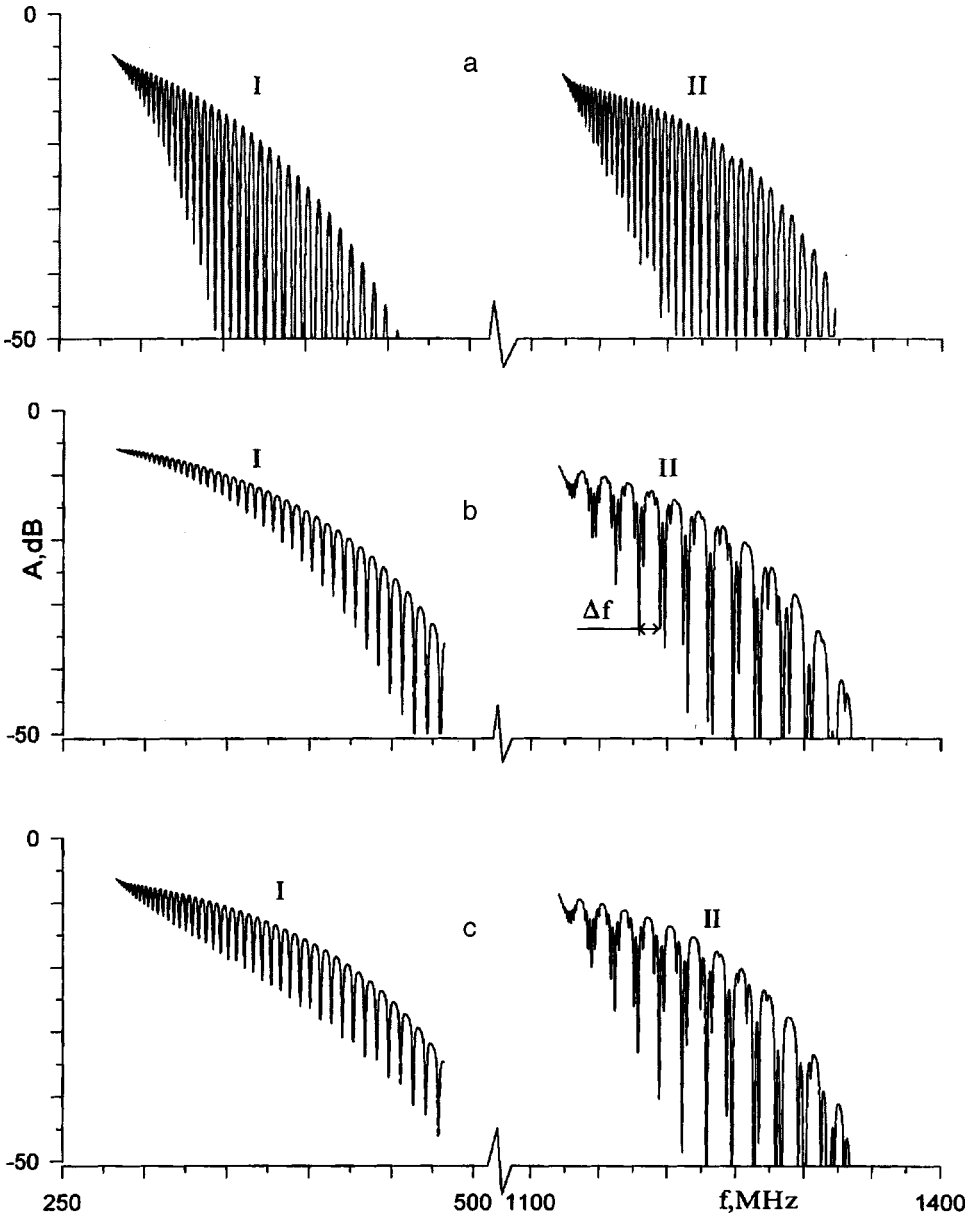


FIG. 2. FRC of a model of delay line type for the case of propagation of MSFVWs in the absence of exchange coupling in free spins on the outer surfaces of the structure. Magnitude of the spin pinning on the inner surfaces: a — $A_{12}=0.01$ erg/cm^2, $L_{21}=L_{12}=10^5$ cm^{-1}; b — $A_{12}=-0.01$ erg/cm^2, $L_{21}=L_{12}=10^5$ cm^{-1}; c — $A_{12}=0.01$ erg/cm^2, $L_{21}=L_{12}=-10^4$ cm^{-1}.

the MSFVW fundamental mode of film 2 and the spin-wave modes of this same layer (layer 2). However, in the given frequency interval spin-wave modes of the layer with larger magnetization (layer 1) are present, whose cutoff frequencies Ω are given by expression (10) upon substitution of the film parameter values of layer 1 and for the frequency range of region *II* in Fig. 1 they have mode numbers $N_1=101-119$ (in Fig. 4 the indicated frequencies are indicated by arrows). The indicated modes of layer 1, of course, interact with the modes of layer 2, which is manifested, in particular, in a ''repulsion'' of the dispersion curves. The efficiency of the given interaction is determined mainly by the values of the spin pinning parameters in layer 1 and is much smaller, as a rule, than the interaction of modes of layer 2 with each other. The above-said is illustrated by the inset to Fig. 4, which reveals the nature of the dispersion of the spin-wave modes of layer 2 with mode numbers $N_2\approx 14-16$ and the spin-wave mode of layer 1 with mode number $N_1\approx 103$, which is marked in the inset by an asterisk. For values of the surface spin pinning parameters $|L_{11}|,|L_{12}|>10^5$ cm^{-1} in region *2* of the FRC the dips at the SWR frequencies of layer 1 become noticeable (Fig. 1b). Note that for the indicated values of the spin pinning parameters in layer 1 in region *1* of the FRC both the depth of the dips and the attenuation of the signal at frequencies far from resonance of the MSFVW and spin-wave modes grow noticeably. On the whole, neglecting the exchange interaction the dependence of the spectrum and the FRC on the structure parameters is analogous to the case of normally magnetized isolated films.[23,24]

Figure 2 displays the results of calculation of the FRC simultaneously taking into account pinning of surface spins near the interlayer boundary in the exchange coupling of layers and the case when $D\sim A_{12}\approx 0.01$ erg/cm^2. When the signs of the surface anisotropy constant and the interlayer exchange constant are the same, the depth of the dips in the FRC increase as a group and an insignificant modulation appears in region *II* in the vicinity of the frequencies Ω

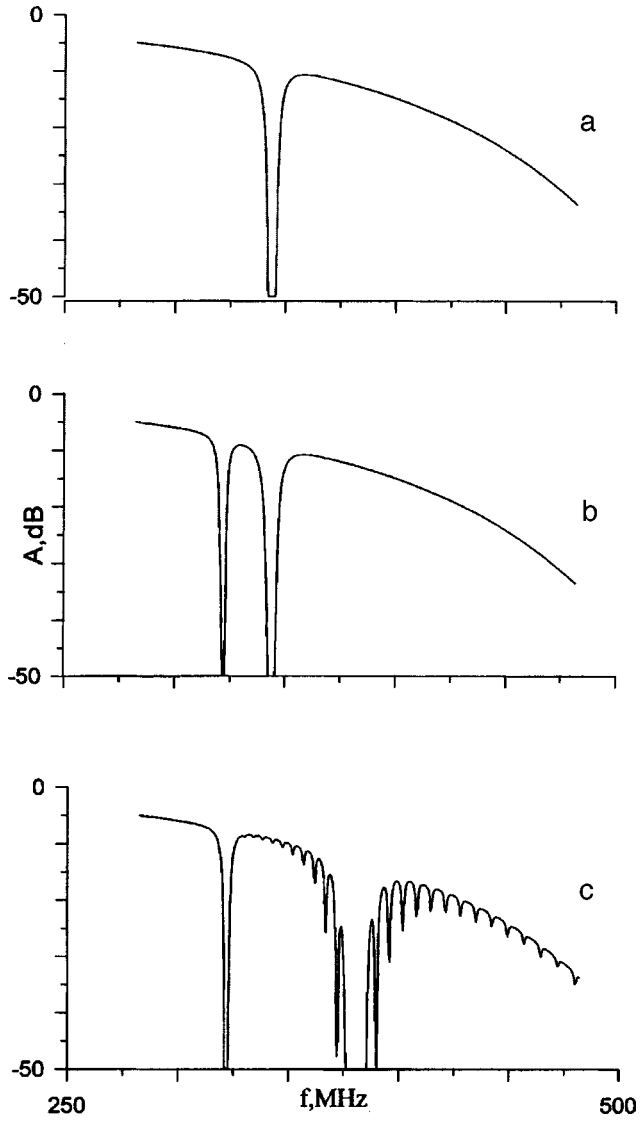


FIG. 3. Region *I* of the FRC of a model of delay line type for the case of propagation of MSFVWs. a — in the absence of exchange coupling for the case of free spins on the outer surfaces of the structure and $L_{12}=0$, $L_{21}=-2.1\times10^5$ cm^{-1}; b — in the absence of exchange coupling for the case of free spins on the surfaces of the first layer and $L_{21}=-2.1\times10^5$ cm^{-1}; $L_{22}=-2.13\times10^5$ cm^{-1}; c — for $A_{12}=0.004$ erg/cm^2 and magnitudes of the spin pinning on the same surfaces as in case b.

corresponding to coincidences of the SWR spectra of the isolated films. This is illustrated by Fig. 2a, where $L_{11}=L_{22}=0$, $L_{21}=L_{12}=10^5$ cm^{-1}, and $A_{12}=0.01$ erg/cm^2.

In the case when the signs of the surface anisotropy constant and the interlayer exchange constant are different, the modulation of the depth of the dips in region *II* of the FRC increases (Figs. 2b and 2c). The frequency positions of the deepest neighboring dips have values similar to those of the frequencies Ω, but do not coincide with them. Table II lists the values of the frequency intervals Δf between the deepest neighboring dips in the region of the FRC indicated in Fig. 2b by arrows and corresponding to MSFVWs with wave numbers $q\approx40$–400 cm^{-1}.

The indicated behavior of the FRC in region *II* is in line with the form of the conditions on spin pinning (7) at the boundary $z=0$ of the exchange-coupled films. Indeed, provided the condition $L_{12},L_{21}\sim A_{12}/A_1,A_{12}/A_2$ is fulfilled, surface anisotropy and exchange coupling give comparable contributions to the spin pinning. In this case, the magnitude of the contribution due to the exchange coupling of layers is determined by the ratio of the high-frequency components of the magnetic moments of the films and is of an oscillatory nature in the vicinity of the frequencies Ω, where $m_1\sim M_2$. The form of the FRC corresponds to the formation of a common spectrum of spin-wave modes of the structure, as is also confirmed by direct calculation of the SWR frequencies of the structure as functions of the exchange coupling parameter A_{12} (inset *2* to Fig. 4).

Note that for $A_{12}>0$ the ''repulsion'' of SWR frequencies at the frequencies Ω takes place ''above'' the frequency Ω (inset *2* to Fig.4). The fact that the frequency of one of the interacting modes remains practically identical with Ω means that the magnetizations in the layers on both sides of the boundary have similar amplitudes and are in phase. Therefore, ferromagnetic exchange coupling does not substantially alter the nature of the magnetization distribution for this mode and at the given frequency the spin pinning on the interlayer boundary falls. From this vantage point it is clear that growth of the frequency of the ''repelled'' mode with increase of A_{12} is due to the tendency of ferromagnetic exchange coupling to decrease the degree to which the oscillations of the magnetization on the two sides of the interlayer boundary are out of phase, which is equivalent to an increase in the spin-wave resonance number and can be identified with growth of spin pinning.[11,21] Taking the above into account, the existence in the vicinity of Ω of the deepest dip at frequencies $\omega>\Omega$ becomes understandable. Antiferromagnetic exchange coupling, on the contrary, most noticeably alters the nature of the magnetization distribution of those modes that oscillate in phase at the boundary $z=0$. In this case, the frequency of the ''repelled'' mode is lowered (inset *2* in Fig. 4) and the deepest dips in region *2* of the FRC are found at frequencies $\omega<\Omega$. The nonmonotonic variation with frequency of the depth of the dips in region *II* of the FRC of the type shown in Figs. 2b and 2c is also preserved in the case when exchange coupling of the films makes the main contribution to spin pinning on the interlayer boundary ($|A_{12}|\gg|D|$).

Let us now consider conditions under which resonance features arise in region *I* of the FRC, associated with the interaction of the spin-wave modes of the layers of the type noted for region *II* of the FRC in Figs. 1 and 2. From Eqs. (9) and (10) it is easy to see that for the structure under consideration consisting of films with different magnetizations and situated in a magnetic field normal to its surface, only surface modes of the FRC of layer 2 which have imaginary wave numbers Q_N can fall into the frequency band of existence of the MSFVWs. Thus, SWR modes exist in films with normal uniaxial surface anisotropy in the case when the films are magnetized in the ''hard'' direction relative to the surface anisotropy axis.[21] For normally magnetized films, SWR surface modes can exist for a normal uniaxial surface anisotropy of ''easy plane'' type, which under conditions (7) corresponds to $L_{ij}<0$. Figure 3a shows region *I* of the FRC for the pinning parameters $L_{11}=L_{12}=L_{22}=0$, $L_{21}=-2.1$

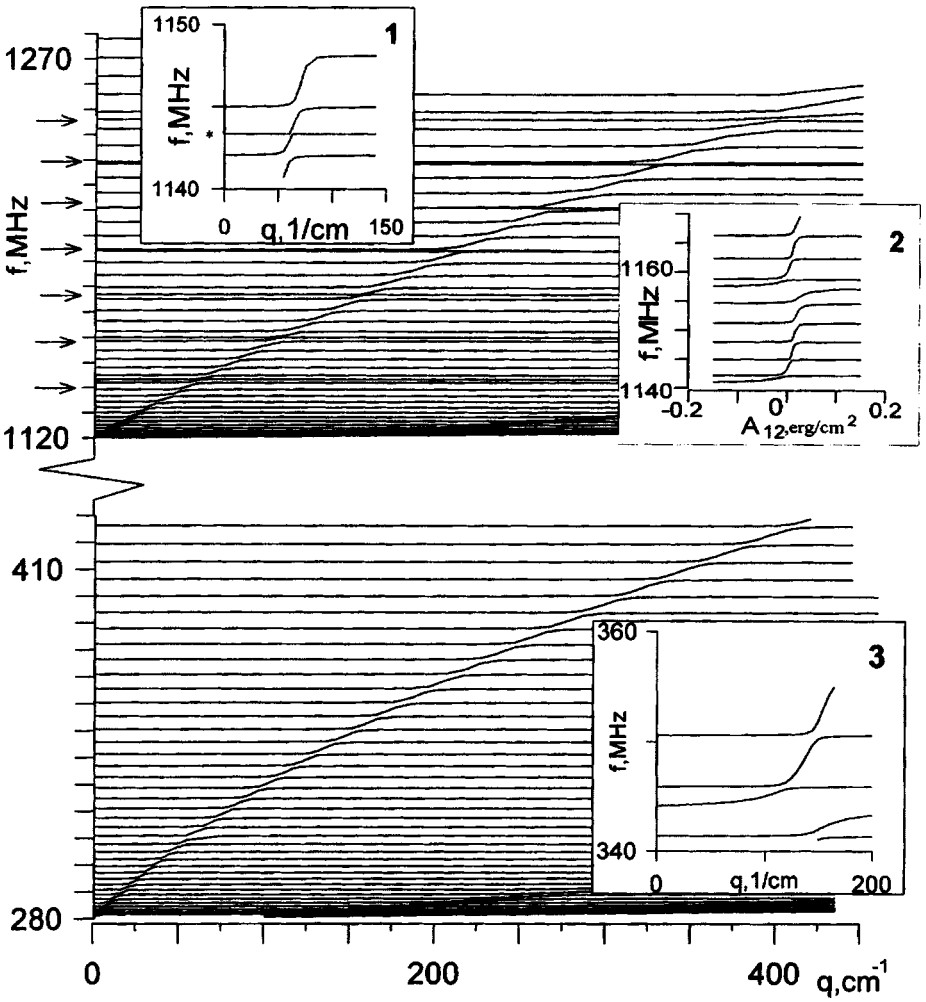


FIG. 4. Dispersion of MSFVWs for $H_0=800$ Oe, calculated in the absence of exchange coupling for the case of free spins on the outer surfaces of the structure and $L_{12}=L_{21}=-10^5$ cm^{-1}.

$\times 10^5$ cm^{-1} in the absence of exchange coupling between the layers. The solitary dip at frequencies $\omega_S \approx 340$–350 MHz is due to interaction of the MSFVWs of layer 1 with the surface exchange spin wave supported by the boundary with pinned spins at $z=0$. Inset *3* to Fig. 4 illustrates the nature of the spectrum of spin-wave modes in the vicinity of ω_S. The position of the frequency ω_S is determined mainly[2] by the value of the spin pinning parameter L in film 2 and falls in the frequency band of existence of MSFVWs of layer 1 for $-2.3\times 10^5 < L_{21} < -1.7 \times 10^5$ cm^{-1}.

Reasonably enough, in the case in which the spins on the boundary $z=-d_2$ in film 2 are characterized by a pinning parameter L_{22} whose value lies in the indicated range, a dip will be formed in the FRC which is due to the interaction of the MSFVWs with the surface spin wave traveling along its surface. Finally, for simultaneous spin pinning on both surfaces of film 2 in region *I* of the FRC two dips can arise whose positions are determined by the parameters L_{21} and L_{22}. This latter case is illustrated by the curve in Fig. 3b, which was calculated for $L_{21}=-2.1\times 10^5$ cm^{-1} and $L_{22}=-2.13\times 10^5$ cm^{-1}.

TABLE II. Widths of the frequency intervals between dips in the FRC of structure 1.

δf, MHz	Δf, MHz
12	17
12	17
12	13
24	18
14	19
13	15

Exchange coupling of the films affects mainly the position and shape of the dip formed due to resonance of the MSFVWs of layer 1 with the surface spin wave traveling in layer 2 near the interlayer boundary $z=0$ (Fig. 3c). The frequency ω_S at which the dip arises is increased if the exchange coupling is ferromagnetic ($A_{12}>0$) and decreased if it is antiferromagnetic ($A_{12}<0$). Together with a change in the position of the dip, exchange coupling leads to ''framing'' of the edges of the dip by a series of shallower dips which are arrayed at the SWR frequencies of layer 1 and whose depth falls with distance from the central frequency of the dip. The appearance of these features is tied up with spin pinning on the interlayer boundary, which is induced in the structure by exchange coupling. Indeed, the presence of natural excitations in the layers at the frequency ω_S means that on both sides of the boundary high-frequency magnetizations have nonzero amplitudes. The surface spin wave in

layer 2 decays exponentially into the film at distances determined by the roots K_j of the characteristic equation and, being primarily an exchange wave for $q=0$, it has nearly zero amplitude at distances $R\sim\sqrt{\omega_{ex}/(\omega_H-\omega_S)}\approx 2\times 10^{-5}$ cm. Such a rapid falloff of the magnetization amplitude in layer 2 can be understood in light of interlayer exchange of the magnetization of layer 1 as an indication of the presence of spin pinning on the boundary and leads to the appearance of dips in the FRC. Obviously, the efficiency of the indicated mechanism falls off outside the region of frequencies ω_S. It is also clear that in structures composed of films with thickness $d_l \gg R$, the influence of exchange coupling of the layers on the resonance of the MSFVWs with the surface spin waves pressed against the outer boundaries of the film (in the given case, up against the boundary $z=-d_2$) will be weak.

MSBVW GEOMETRY

To consider the propagation of magnetostatic backward volume waves (MSBVWs) in the structure, we direct the field along the x axis. We take the expressions for the fields h_{xl}, h_{zl} and the magnetization component m_{yl} to be in a form analogous to expressions (1), and we write the component m_{zl} as

$$m_{zl}=\left(\sum_{j=1}^{6}\xi_j B_j \exp iK_j z\right)_l T. \tag{11}$$

Here K_{jl} are the roots of the characteristic equation for the lth film in the form (2), where the coefficients a,b,c have the form

$$a=(\omega_m+2\omega_H)/\omega_{ex},$$
$$b=(\omega_m\omega_H+\omega_H^2-\omega^2-q^2\omega_m\omega_{ex}/\omega_{ex}^2),$$
$$c=-q^2\omega_m\omega_H/\omega_{ex}^2, \quad \omega_H=\gamma H_0. \tag{12}$$

We represent the coefficients ξ_j in the form

$$\xi_j=\frac{\tilde{\omega}_H^2\gamma M_0 K_j/q}{\tilde{\omega}_H^2-\omega^2}.$$

The expressions for the fields in the regions $z>d_1$ and $z<-d_2$ have the form (5).

We assume that the conditions of continuity of the tangential fields h_x and the normal components of the magnetic induction $b_z=h_z+4\pi m_z$ in the form (6) are satisfied at the boundaries of the structure $z=d_1,0,-d_2$. We write the exchange boundary conditions in the form characteristic for normal uniaxial surface anisotropy[22,9]

$$\frac{\partial m_{z1}}{\partial z}-L_{11}m_{z1}=0, \quad \left.\frac{\partial m_{y1}}{\partial z}=0\right|_{z=d_1},$$

$$\frac{\partial m_{z2}}{\partial z}+L_{22}m_{z2}=0, \quad \left.\frac{\partial m_{y2}}{\partial z}=0\right|_{z=-d_2},$$

$$\frac{\partial m_{z1}}{\partial z}+L_{12}m_{z1}-\frac{A_{12}}{A_1}\left(m_{z1}-\frac{M_{01}}{M_{02}}m_{z2}\right)=0,$$

$$\frac{\partial m_{x1}}{\partial z}-\frac{A_{12}}{A_1}\left(m_{x1}-\frac{M_{01}}{M_{02}}m_{x2}\right)=0\Bigg|_{z=0},$$

$$\frac{\partial m_{z2}}{\partial z}-L_{21}m_{z2}+\frac{A_{12}}{A_2}\left(m_{z2}-\frac{M_{02}}{M_{01}}m_{z1}\right)=0,$$

$$\frac{\partial m_{x2}}{\partial z}+\frac{A_{12}}{A_2}\left(m_{x2}-\frac{M_{02}}{M_{01}}m_{x1}\right)=0\Bigg|_{z=0}. \tag{13}$$

The dispersion equation was derived and studied in analogy with the above-considered case of MSFVWs. Figures 5 and 6 display the form of the spin-wave spectra and the FRC of a model of delay line type corresponding to wave propagation in structure 1 from Table I for $H_0=193$ Oe and the above-chosen values of the damping ΔH, distance S, and different values of A_{12} and L_{li}. The calculations were limited to the frequency region of the existence of dipole MSBVWs in the isolated layers $\omega_H<\omega<\omega_{0l}$, where ω_H is the short-wavelength ($q\to\infty$) boundary of the MSBVW spectrum (12), being common for both films, and $\omega_{0l}=\sqrt{\omega_H^2+\omega_H\omega_{ml}}$ are the long-wavelength boundaries ($q\to 0$) of the dipole MSBVW spectra. It is clear that for structure 1 the long-wavelength boundary of layer 1 with magnetization $4\pi M_0=700$ G is located at a higher frequency than the long-wavelength boundary of layer 2.

Before going on to a discussion of the results, note that the frequencies of the SWR exchange modes in tangentially magnetized films are given by[22]

$$f_N=\sqrt{[f_H+f_m+f_{ex}Q_N^2][f_H+f_{ex}Q_N^2]}. \tag{14}$$

Clearly, in isolated layers the resonant interaction of dipole MSBVWs with exchange spin-wave modes is impossible since $f_N>f_0=\omega_0/2\pi$. For two-layer structures like structure 1 in Table I in the frequency interval $\omega_{02}<\omega<\omega_{01}$ the dipole MSBVWs of the layer with greater magnetization can interact resonantly with the volume exchange spin-wave modes of the layer with smaller magnetization.[13] Note also that for $L>0$ in a tangentially magnetized film with normal uniaxial surface anisotropy the existence of surface exchange spin waves is possible at frequencies $\omega_S<\omega_0$ (Ref. 21). As in the case considered above of normally magnetized films with $L<0$, it may be expected that when such surface waves fall into the spectrum of dipole MSBVWs an interaction will arise between them. In two-layer structures containing films with different magnetizations, surface waves supported by all four surfaces of the structure can fall into the frequency region from ω_H to ω_{0m}, where $\omega_{0m}=\min(\omega_{01},\omega_{02})$, for certain values of the parameters $L_{li}>0$.

Figure 5 displays the form of the spin-wave spectrum calculated for structure 1 with pinning parameters $L_{11}=L_{22}=0$, $L_{12}=L_{21}=10^5$ cm^{-1} in the absence of exchange coupling. It can be seen that the nature of the spectrum as a whole corresponds to the above arguments—in the frequency interval 900–1200 MHz regions of ''repulsion'' of the dispersion curves of the layer spin-wave modes are visible, arising at the frequencies of degeneracy of the dipole MSBVW spectrum of layer 1 and the exchange spin-wave spectrum of layer 2. The magnitude of the ''repulsion'' of the dispersion

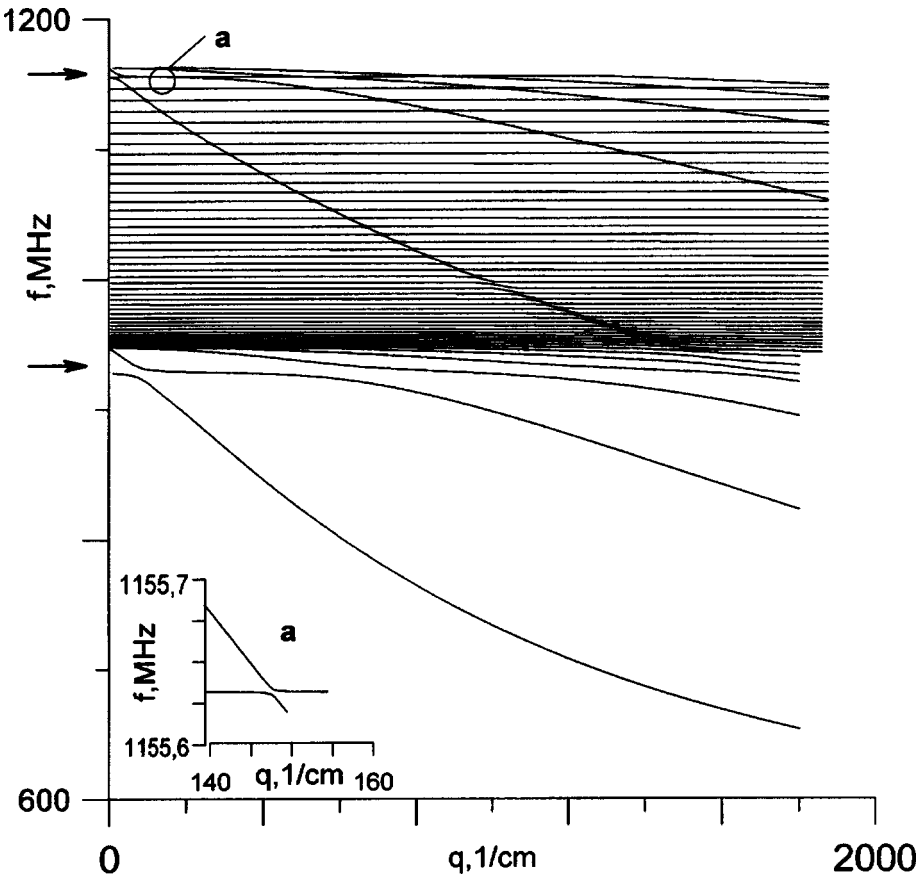


FIG. 5. Same as in Fig. 4, for $H_0=193$ Oe.

curves is governed mainly by the spin pinning parameters in layer 2 and the interlayer exchange parameter as shown in Refs. 9, 6, and 24. In addition, near the frequencies $\omega_{S1} \approx 970$ MHz and $\omega_{S2} \approx 1155$ MHz, whose positions in Fig. 5 are indicated by arrows, significant restructuring of the spectrum is seen to take place, due to interaction of the MSBVWs with the surface spin waves. Inset a to Fig. 5 reveals the nature of the ''repulsion'' of the spin-wave modes near the frequency ω_{S2}.

Figure 6 displays the form of the FRC for chosen values of the pinning parameters and the exchange coupling parameter $A_{12}=0$ (Fig. 6a), $A_{12}=0.005$ erg/cm^2 (Fig. 6b), and $A_{12}=-0.001$ erg/cm^2 (Fig. 6c). It can be seen that at the resonance frequencies of the MSBVWs with the surface waves and at the resonance frequencies of the MSBVWs of layer 1 with the volume exchange modes of layer 2, dips are formed in the FRC, and that in contrast to the case of MSFVWs the depth of the dips in the absence of exchange coupling between the layers at the frequencies of the MSBVW resonances and volume exchange modes turns out to be substantially less and grows as the MSBVW frequency is lowered. Note also that in the long-wavelength region of the FRC $q \leqslant 5\times 10^2$ cm^{-1}, which is the most accessible frequency range for experiment and is demarcated by arrows in Fig. 6a, variations in the resonance frequencies are essentially absent—the depth of the dips does not exceed 1–2 dB.

For the exchange coupling parameter $A_{12}\neq 0$ the depth of the dips in the long-wavelength region of the FRC grows by 5–10 dB. It can be seen that in contrast to the MSFVW cases shown in Figs. 3b and 3c, taking inhomogeneous exchange into account in the given case has hardly any effect on the nature of the frequency dependence of the depth or the mutual arrangement of the dips corresponding to the resonances of the MSBVWs of layer 1 with the volume spin-wave modes of layer 2, which is well described by expression (14).

Exchange coupling has a much stronger effect on the resonance of the MSBVWs with the surface spin waves propagating along the interlayer boundary $z=2$. For ferromagnetic exchange coupling the frequencies of the surface spin waves ω_S grow whereas for antiferromagnetic exchange coupling they fall. Clearly, for fixed values of the spin pinning parameters L at the boundary $z=0$ such values of the parameters $A_{12}>0$ will be found for which the surface spin waves cease to exist and the dips associated with them in the FRC will be absent (Fig. 6b). It is also clear that as the antiferromagnetic exchange coupling is increased, the frequency ω_S can fall so far that it drops below the lower boundary of the MSBVW spectrum: $\omega_S<\omega_H$. For the case shown in Fig. 6c, exchange coupling does not violate the conditions of existence of the resonance of the surface spin waves with the MSBVWs. In this case in region *II* of the FRC in the vicinity of ω_{S2} the dips corresponding to resonances of the MSBVWs with the volume spin-wave modes of layer 2 deepen as in the case of the MSFVWs (Fig. 3c).

A distinguishing feature of tangentially magnetized structures in comparison with the case of normal magnetization is the possibility of the appearance in the spectrum, for antiferromagnetic ($A_{12}<0$) exchange coupling, of waves of the two-layer structure at the frequencies $\omega_S<\omega_{0m}$ of the

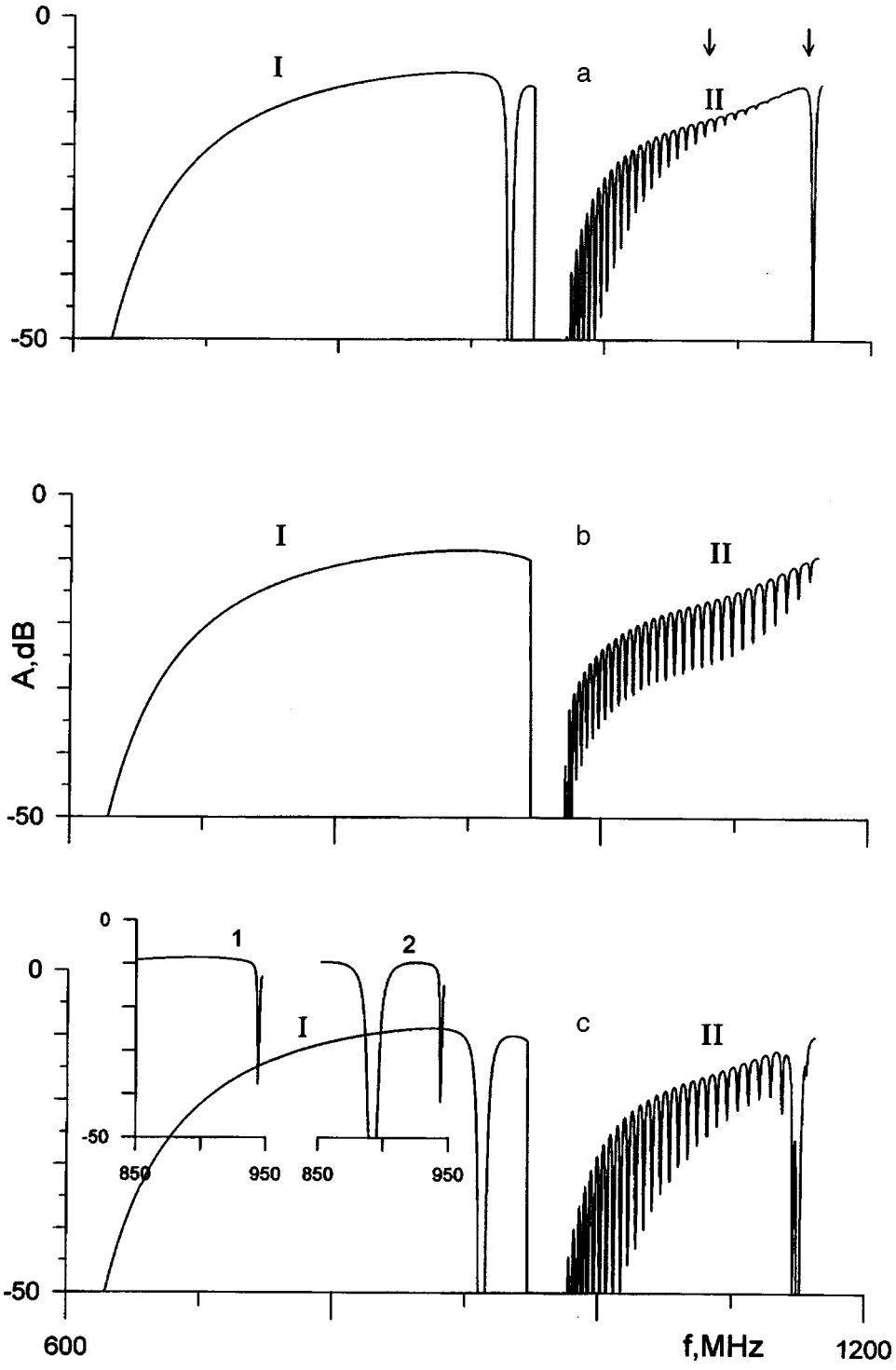


FIG. 6. FRC of a model of delay line type for the case of propagation of MSBVWs, calculated for the case of free spins on the outer surfaces for different values of the parameters of exchange coupling and spin pinning on the inner surfaces. a — for $A_{12}=0$, $L_{12}=L_{21}=-10^5$ cm^{-1}; b — $A_{12}=0.005$ erg/cm^2, $L_{12}=L_{21}=-10^5$ cm^{-1}; c — $A_{12}=-0.001$ erg/cm^2, $L_{12}=L_{21}=-10^5$ cm^{-1}.

surface spin waves supported by the interlayer boundary.[7] Such waves, which like surface spin waves are due to surface anisotropy, can also interact resonantly with the MSBVWs and lead to the appearance of dips in region *I* of the FRC. For this case the numeral 1 in Fig. 6c indicates region I of the FRC, calculated for the case of free surface spins $L=0$ and $A_{12}=-0.001$ erg/cm^2. Note that in the case of antiferromagnetic exchange coupling between the layers and spin pinning on the outer boundary of the film with the lower magnetization two dips can appear in the FRC for $10^4<L_{22}<1.7\times10^5$ cm^{-1} and $0>A_{12}>-0.01$ erg/cm^2, corresponding to resonances of the MSBVWs and surface spin waves localized near the boundaries $z=0$ and $z=-d_2$ (curve *2* of the FRC in Fig. 6c). If the spins are pinned on interlayer boundary, then taking the interlayer exchange interaction into account does not lead to the appearance of an additional dip but only alters the shape and frequency of the dip caused by the surface anisotropy. Also note that for values of the spin pinning parameter $1.9\times10^5<L<2.2\times10^5$ in layer 1 the surface spin waves supported by these boundaries will also interact with the MSBVWs of layer 2.

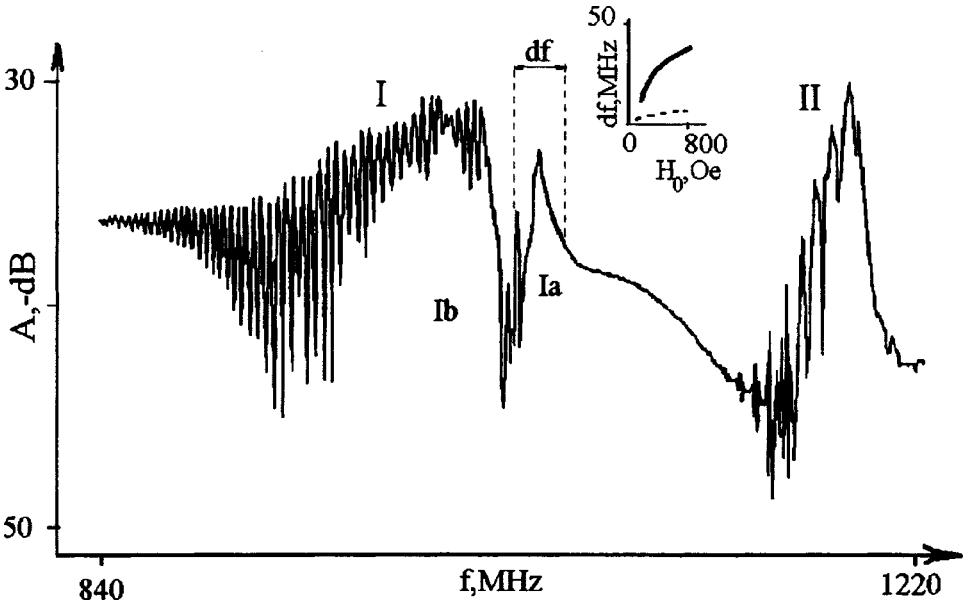


FIG. 7. FRC of a model of delay line type for the case of propagation of MSBVWs for $H_0=193$ Oe.

EXPERIMENTAL RESULTS

In the experiments we used two-layer ferrite structures prepared by liquid-phase epitaxy on substrates of gadolinium–gallium garnet with (111) orientation by successive growth of layers of yttrium–iron garnet with the composition $Y_3Fe_{2-y}Sc_yFe_{3-x}Ga_xO_{12}$, $y\leqslant 0.3$, $x\leqslant 1.0$. The layer parameters are indicated in Table I. The gyromagnetic ratios in both layers had values typical for YIG films $\gamma=2.8$ MHz/Oe.

The samples were placed in an input and an output microstrip transformer, both of which had a width of 15μm and a length of 5 mm. The distance between the two transformers could be varied, which made it possible to use the ''mobile probe'' method to measure the dispersion and damping of the magnetostatic waves.[25] The external magnetic field H_0 was oriented either normal to the plane of the structure (''MSFVW geometry'') or in the plane of the structure perpendicular to the microstrips (''MSBVW geometry'').

We investigated the amplitude and phase–frequency characteristics of an MSW delay-line model in the frequency range 0.1−6 GHz at room temperatures. Special attention was given to the appearance of segments of resonant growth of losses, typical for resonances of dipole MSWs with the spin-wave modes of the structure, in the FRC of models like those shown in Figs. 1, 2, 3 and 6.

Figure 7 shows the form of the FRC of the model with structure 1 for the MSBVW geometry for the external field $H_0=193$ Oe and the distance between the transformers equal to 4 mm. It is possible to distinguish two regions of signal transmission which according to Figs. 5 and 6 should be ascribed to propagation of MSBVWs in layers 1 and 2. The frequencies of the long-wavelength boundaries of the regions of existence of dipole MSBVWs in the films f_{0l} are shifted by 30–50 MHz relative to the values shown in Figs. 5 and 6, which may be due to the influence of volume anisotropy in the layers. Inside region II of the FRC, segments of fade-out of interference are visible, as are dips, whose positions are accurately given by formula (14) upon substitution of the parameters corresponding to layer 2, and the resonance numbers $N_2\approx 29$–36. At the same time, segments of anomalous dispersion are observed in the dispersion curves, and the standing wave ratio of the input transformer grows as described in Ref. 13.

The indicated features in the damping, dispersion, and excitation are characteristic signs of resonant interaction of the dipole and exchange waves and, according to the results of calculations shown in Figs. 5 and 6, should be linked with resonant interaction of the fundamental mode of the dipole MSBVWs of layer 1 with the volume spin-wave modes of layer 2. Note that in the frequency interval corresponding to region *2* of the FRC, MSBVWs in the structure have wave numbers $q\leqslant 500$ cm^{-1}, and the depth of the dips in this case is 10–20 dB. As was noted in the discussion of the results of calculations of the FRC of the MSBVWs, in the long-wavelength region the appearance of dips of such significant depth can be due only to the presence of exchange coupling between the layers. Comparison with experiment shows that the exchange coupling parameter in this case is $A_{12}\sim 0.01$ erg/cm^2.

Note also that at frequencies 970–980 MHz of region *I* of the FRC a dip is observed that divides it into two parts: *Ia* and *Ib* (Fig. 7). From its position and shape, the indicated dip is analogous to the dip in the FRC in Fig. 6 due to resonant interaction of the MSBVWs with the surface spin wave. However, it seems that its appearance cannot be linked with resonant interaction of a dipole MSBVWs with a surface spin wave. Indeed, one of the properties of pinning of surface spins in tangentially magnetized films with normal uniaxial surface anisotropy is the frequency dependence of the effective pinning parameter[21]

$$L^{\text{eff}}=\frac{L^*}{2}\left(1-\frac{1}{\sqrt{1+(2\omega/\omega_m)^2}}\right). \tag{15}$$

It can be easily seen from Eq. (15) that in the investigated frequency range 0.1−6 GHz the parameter L^{eff} varies by

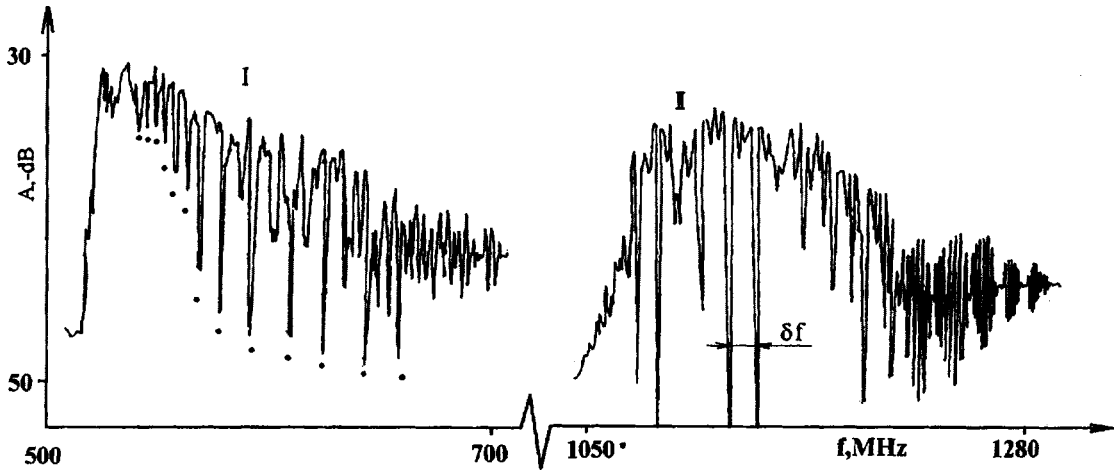


FIG. 8. Same as in Fig. 7, for $H_0=800$ *Oe*.

almost an order of magnitude. In line with this the position of the resonance dip in the FRC should change. This position can be characterized by the magnitude of the detuning dF of its central frequency from the long-wavelength frequency boundary f_0 of the MSBVWs in the film (Fig. 6). The inset to Fig. 7 plots the calculated (solid curve) and experimental (dotted) dependence of the parameter dF on the magnetic field H_0. It can be seen that the indicated curves are quite different. Etching of the outer surface of the film to a depth of $\sim 1\,\mu$m did not result in disappearance of this dip.

In addition, upon turning the model relative to the direction of the tangential field H_0 in such a way that the field is finally aligned with the transformers, region Ia took on the form of the FRC corresponding to a Damon–Eshbach magnetostatic surface wave. Consequently, the appearance of region *Ia* in the FRC shown in Fig. 7 can, in analogy with Ref. 26, be linked with resonant excitation of the structure by the high-frequency magnetic field of the strip transformer and with the setting up in the film at the frequencies of existence of the forward magnetostatic waves of magnetization oscillations.

Figure 8 reveals the shape of the FRC of the model with structure 1, normally magnetized in a field $H_0=800$ Oe for the distance between the transformers equal to 4 mm. It can be seen that both regions of signal transmission I and II have an indented appearance. Differences in the positions of the calculated and measured boundaries can be ascribed to the influence of anisotropy and other factors outside the scope of the approximations used in the calculations, e.g., nonuniformity of the layers.[27] The form of the FRC as a whole corresponds to the calculated results shown in Fig. 2b, where we took $A_{12}=0.01$ erg/cm^2 and $L=-10^5$ cm^{-1}. The deep dips in region I of the FRC, marked by dots, are located at frequencies given by formula (10) upon substitution of the parameters corresponding to layer 1 and the mode numbers $N_1\approx 5-23$ and on the whole have the form characteristic for isolated films with asymmetrically pinned surface spins.[6,24]

In region II of the FRC the frequencies at which the deepest dips are located are separated from one another by $\delta f\approx 8-20$ MHz. Here the values of δf behave nonmonotonically with growth of the frequency in the region of the FRC corresponding to MSFVWs with wave numbers $q\approx 40-400$ cm^{-1}, and in a way completely analogous to the pattern shown in Figs. 2b and 2c (see also Table II).

The adduced experimental results clearly indicate the presence in the structure of an interlayer exchange interaction. Comparison of the FRCs shown in Figs. 7 and 8 with the calculated results shows first that the relation $A_{12}\leqslant D$ between the interlayer exchange constant and the surface anisotropy constant is fulfilled at the interlayer boundary and, second, that the interlayer exchange parameter in the structure under consideration varies in the range $0.005<|A_{12}|<0.05$ erg/cm^2. However, it is difficult to uniquely determine the nature of the exchange coupling on the basis of the experimental data shown in Figs. 7 and 8 since the magnitudes and signs of the spin coupling parameters at the boundaries of the structure are not known with sufficient accuracy. At the same time, resonance of the MSBVWs propagating in the layer with lower magnetization with the surface spin wave was not observed experimentally whereas such a resonance can observed for $-0.001>A_{12}>-0.01$ erg/cm^2. This gives us reason to believe that in the investigated structure interlayer exchange is ferromagnetic: $A_{12}>0$. Note that this conjecture, like the above determined interval of absolute values of A_{12}, is in good agreement with known results for multilayer epitaxial garnet structures.[28,29] Within the framework of the two-layer structure model with normal uniaxial surface anisotropy, not taking account of either thickness nonuniformity of either of the layers or the presence of a transitional layer between them, agreement of the experimental and calculated results for the chosen values of the structure parameters d_l, $4\pi M_{0l}$, A_l, and ΔH_l can be obtained for $A_{12}\sim 0.01$ erg/cm^2 and $L\sim -10^5$ cm^{-1}.

Note that propagation of MSFVWs and MSBVWs in the remaining structures of Table I is accompanied by similar peculiarities, and the structures themselves are characterized by interlayer exchange and surface spin pinning parameters of the same order as for structure 1.

CONCLUSION

We have shown that the influence of interlayer exchange interaction on the propagation of magnetostatic dipole vol-

ume waves in two-layer epitaxial ferrite structures is manifested mainly at the frequencies of resonant interaction of the magnetostatic dipole volume waves primarily with the volume spin-wave modes of the structure and has a form completely analogous to the case of solitary films: at the indicated frequencies the losses of the magnetostatic waves grow resonantly, and the spectra of the interacting waves repel. We have shown that anomalies arise in the frequency dependence of the indicated spectral transformations and of the wave damping in structures with exchange coupling and that these anomalies arise at the degeneracy frequencies of the spin-wave modes of the structure. The indicated anomalies are caused by repulsion of the spectra of the degenerate spin-wave modes accompanied by the effect of dynamic pinning of the ''interlayer'' spins at the frequency of the ''repelled'' spin-wave mode, and are configured by the magnitude of the exchange coupling parameter A_{12}, and also by the relation between A_{12} and the spin pinning parameters at the film boundaries L.

We have shown that in structures with uniaxial normal surface anisotropy, magnetized in the hard direction relative to the anisotropy axis, magnetostatic dipole volume waves can be found under conditions of phase synchronism and can interact efficiently with the surface spin waves propagating along the film boundaries with pinned spins. In these structures, exchange coupling of the layers leads to appearance of dynamic spin pinning for those spin modes whose frequencies are close enough to the resonance frequency of the magnetostatic dipole volume waves and the surface exchange wave.

In structures comprised of films with different saturation magnetization, we have experimentally investigated the influence of interlayer exchange on the propagation of magnetostatic dipole forward and reverse volume waves. In the case of backward volume waves we have shown that the losses to propagation of the dipole wave propagating in the layer with greater magnetization grow resonantly at the hybridization frequencies of the dipole waves of this layer with the SWR modes of the layer with lower magnetization, and that for isolated layers the indicated changes in the properties of the MSBVWs are fundamentally impossible. In the case of MSFVWs the influence of interlayer exchange is most strikingly manifested at the degeneracy frequencies of the SWR spectra of the isolated films, i.e., in the frequency region corresponding to propagation of MSFVWs of the layer with lower magnetization. A study of the indicated anomalies in the propagation of magnetostatic waves would make it possible to estimate the interlayer exchange parameter and spin coupling parameters in the structure.

This work was supported by the Russian Fund for Fundamental Research (Project No. 97-02-18614).

[1)]The phonon mechanism of formation of the spin-wave excitation spectrum in a multilayer structure was considered in Ref. 4.

[2)]Since the frequencies of the surface spin waves ω_S in normally magnetized films have values[21] $\omega_S < \omega_H$, where ω_H is given by formula (3), the surface spin waves of layer 1 cannot interact resonantly with the MSFVWs of the structure and are not considered here.

[1] G. M. Vapné, Obzory po Élektronnoĭ Tekhnike, Ser. 1, No. 8(1060) (1984).
[2] Yu. I. Bespyatykh and V. I. Zubkov, Zh. Tekh. Fiz. **45**, 2386 (1975) [Sov. Phys. Tech. Phys. **20**, 1485 (1975)].
[3] R. E. Camley, T. S. Rahman, and D. L. Mils, Phys. Rev. B **27**, 261 (1993).
[4] S. V. Tarasenko, Fiz. Tverd. Tela (St. Petersburg) **36**, 2554 (1994) [Phys. Solid State **36**, 1390 (1994)].
[5] K. Vayhinger and H. Kronmuller, J. Magn. Magn. Mater. **62**, 159 (1986).
[6] B. A. Kalinikos and P. A. Kolodin, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **32**, 1290 (1989).
[7] H. Puszkarski, Surf. Sci. Rep. **20**, No. 2 (1994).
[8] K. Vayhinger and H. Kronmuller, J. Magn. Magn. Mater. **72**, 307 (1986).
[9] B. Hillebrandt, Phys. Rev. B **37**, 9885 (1988).
[10] S. L. Vysotskiĭ, G. T. Kazakov, M. L. Kats, and Yu. A. Filimonov, Fiz. Tverd. Tela (St. Petersburg) **35**, 1191 (1993) [Phys. Solid State **35**, 606 (1993)].
[11] F. Hoffmann, Phys. Solid State **41**, 807 (1970).
[12] C. Vittoria, Phys. Rev. B **37**, 2387 (1988).
[13] S. L. Vysotskiĭ, G. T. Kazakov, B. P. Nam, A. V. Maryakhin *et al.*, Pis'ma Zh. Tekh. Fiz. **19**(11), 65 (1993) [Tech. Phys. Lett. **19**(6), 349 (1993)].
[14] Yu. A. Filimonov, G. T. Kazakov, S. L. Visotsky *et al.*, J. Magn. Magn. Mater. **131**, 235 (1994).
[15] S. L. Vysotskiĭ, G. T. Kazakov, A. S. Khe, A. V. Maryakhin, and Yu. A. Filimonov, JETP Lett. **61**, 693 (1993)].
[16] B. A. Kalinikos, Izv. Vyssh. Uchebn. Zaved. Fiz. **24**, No. 8, 42 (1981).
[17] Yu. V. Gulyaev, P. E. Zil'berman, and A. V. Lugovskoĭ, Fiz. Tverd. Tela (Leningrad) **23**, 1136 (1981) [Sov. Phys. Solid State **23**, 660 (1981)].
[18] B. A. Kalinikos, N. G. Kovshikov, and A. N. Slavin, Zh. Éksp. Teor. Fiz. **94**, 159 (1986) [Sov. Phys. JETP **67**, 89 (1986)].
[19] A. S. Andreev, Yu. V. Gulyaev, P. E. Zil'berman *et al.*, Zh. Éksp. Teor. Fiz. **86**, 1005 (1984) [Sov. Phys. JETP **59**, 586 (1984)].
[20] V. M. Sokolov and B. A. Tavger, Fiz. Tverd. Tela (Leningrad) **10**, 1793 (1968) [Sov. Phys. Solid State **10**, 1412 (1968)].
[21] N. M. Salanskiĭ and M. Sh. Erukhimov, *Physical Properties and Uses of Magnetic Films* [in Russian] (Novosibirsk, 1975), 222 pp.
[22] A. V. Lugovskoĭ and V. I. Shcheglov, Radiotekh. Elektron. **27**, 518 (1982).
[23] B. A. Kalinikos, N. V. Kozhus, M. P. Kostylev, and A. N. Slavin, J. Phys.: Condens. Matter **2**, 9861 (1990).
[24] P. E. Zil'berman and A. V. Lugovskoĭ, Zh. Tekh. Fiz. **57**(1), 3 (1987) [Sov. Phys. Tech. Phys. **32**, 1 (1987)].
[25] P. E. Zil'berman, G. T. Kazakov, and V. V. Tikhonov, Radiotekh. Elektron. **30**, 1164 (1985).
[26] S. L. Vysotskiĭ, G. T. Kazakov, A. V. Maryakhin *et al.*, Radiotekh. Elektron. **37**, 1086 (1992).
[27] P. E. Zil'berman, A. G. Temiryazev, and M. P. Tikhomirova, Zh. Tekh. Fiz. **108**, 281 (1995) [JETP **81**, 151 (1995)].
[28] H. U. Uchishiba, H. Tominaga, and T. Namikata, Jpn. J. Appl. Phys. **12**, 928 (1973).
[29] A. H. Bobek, S. L. Blank, and H. J. Levinstein, Bell Syst. Tech. J. **51**, 1427 (1972).

Translated by Paul F. Schippnick

Mechanisms and kinetics of the initial stages of growth of films grown by chemical vapor deposition

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

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

(Submitted February 13, 1997)
Zh. Tekh. Fiz. **68**, 111–117 (July 1998)

The initial stages of growth of films and coatings by chemical vapor deposition are investigated. A system of equations is derived which describes the evolution of an island film at the stage of Ostwald ripening under conditions characteristic of vapor deposition. Solving this system of equations yields the dependence of all of the main characteristics of island films (the size distribution function of the islands, the dependence of the mean radius and density of the islands) as functions of time and the spatial coordinate. Suggestions are given for the preparation of films with prescribed properties. © *1998 American Institute of Physics.*
[S1063-7842(98)01807-8]

INTRODUCTION

The method of chemical vapor deposition (CVD) is widely used to prepare films and coatings for different purposes, including semiconductor films, HTSC films, and many others.[1–5] There are a number of works, both experimental[2,5] and theoretical,[3,4] that examine processes of film growth using this method. The theoretical studies, as a rule, examine only the hydrodynamics of the flow around the substrate, while the influence of the surface and the processes taking place on it reduces for the most part to just a calculation of the fluxes of the components to be deposited.[3,4] In particular, no attention has been given thus far to the influence of the fluxes of the deposited components on the kinetics of film growth as well as such characteristics of island films as the size distribution of the islands, the mean radius and critical radius of the islands, etc.

On the other hand, a number of works devoted to processes taking place on the surface[6–8] consider growth of the films only in the case when the material is delivered to the substrate uniformly over its area. The present paper represents an effort to consider the theory of growth of island films under conditions characteristic of CVD processes.

STATEMENT OF PROBLEM AND PHYSICAL ESSENCE OF THE PROCESSES

The essence of the method of chemical vapor deposition consists in blasting the substrate with a carrier gas containing one or more impurity components which serve as the source of material for film growth. The film can grow both directly from the impurity components or products of their reactions with each other (e.g., in the growth of zinc selenide films[3]) and from their decay products (when using metalorganic compounds as the impurities[2]). In this regard, for convenience of description we can divide the process of vapor deposition into two stages. In the first stage the impurity components are brought to the surface of the substrate and the chemical reaction takes place liberating the materials from which the film will grow. Examples of such a reaction is the reaction between gaseous zinc and selenium with the formation of zinc selenide,[3] the decomposition of metalorganic compounds with the liberation of metals,[2] etc. This process is described by the equation of convective diffusion with the corresponding boundary conditions, which we will consider in greater detail below.

In the second stage, at first there takes place an accumulation on the substrate of an excess quantity of material (in comparison with equilibrium), from which the film will grow (e.g., zinc selenide[3]). The accumulation of material for growth can take place both directly from the components brought to the surface, and as the result of a chemical reaction between them. Next, a first-order phase transition takes place with the formation of solid islands of a new phase. A first-order phase transition on a solid surface, as a rule, can be broken down into several stages.[1,6] First takes place fluctuational formation of islands of the new phase on the substrate surface, followed by independent growth of islands without change in their number, followed by coalescence, or Ostwald ripening. This last process is characterized by the formation of generalized thermal and diffusion fields in which islands with size less than critical dissolve and those with size greater than critical grow. During this stage the total number of islands decreases and their size and phase distribution is established (in the case of a multicomponent system). This stage is the most prolonged, and as a rule final formation of the structure of the film takes place during this stage, as has been shown[5–8] in a number of experimental and theoretical works.[1)] By virtue of the hydrodynamics of the flow of the carrier gas around the substrate, the quantity of the reagents delivered to its surface and reacting on it is different over its area. This leads to the result that the concentration of the material from which the film grows will be nonuniform over the area of the substrate; consequently the structure of the film will also be nonuniform over its area.

For definiteness, we will consider a variant of the method of vapor deposition in which the substrate is aligned

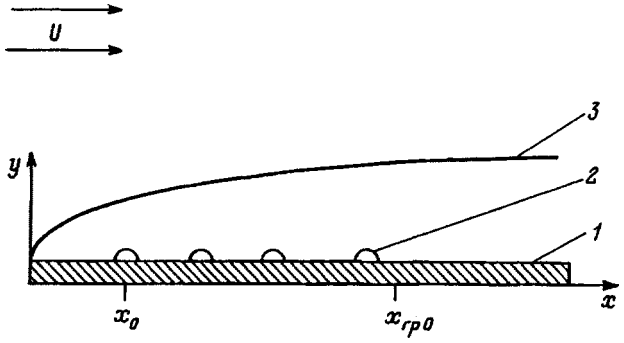


FIG. 1. Diagram of the process. Arrows indicate direction of motion of the carrier gas: *1* — substrate, *2* — islands, *3* — boundary layer.

with the carrier-gas flow (Fig. 1). Below we show how it is possible to extend the proposed theory to other orientations of the substrate. To elucidate the physical essence of the processes involved, we will examine the case in which islands of only one composition, i.e., of one phase, are formed on the surface. These islands are formed of at least two components (the substrate surface may serve formally as the second component). We let the process take place under isothermal conditions, i.e., the substrate temperature is taken to be constant. A diagram of the process is shown in Fig. 1. It was shown in Refs. 1 and 6 that a necessary condition of occurrence of the process of Ostwald ripening is that the ensemble of islands be found within the boundary layer of the gas. In the case of detachment of the boundary layer the process of Ostwald ripening breaks down, the islands begin to grow independently of one another, the generalized diffusion field breaks down, and the film will not have the predicted structural and phase composition. This phenomenon, obviously, can be used to explain the fact that it is possible to deposit high-quality films and coatings by vapor deposition, as a rule, on items with good hydrodynamic flow around them.

SAMPLE CALCULATION OF THE FLUX OF REACTING MATERIAL FROM THE CARRIER GAS TO THE SUBSTRATE

We take the length and thickness of the wafer that we have chosen as to serve as the substrate (Fig. 1) to be much greater than its thickness. In this case it may be treated as infinitesimally thin. The substrate is blasted by the carrier gas at below atmospheric pressure and with a constant temperature. The gas velocity far from the substrate is equal to U. The reacting components are mixed into the carrier gas with concentrations C_i. If the concentration of the impurity components is significantly less than the concentration of the carrier gas and they do not interact, then their behavior can be treated separately. We assume that the process takes place under stationary conditions. We write the two-dimensional (Fig. 1) stationary equation of convective diffusion (1) for each impurity component in the incompressible carrier gas moving rectilinearly under conditions of laminar flow. If the parameters of the process vary substantially in time, then it is necessary to use the nonstationary equation of convective diffusion, which is usually solved numerically[3,4]

$$V_x \frac{\partial C_i}{\partial x} + V_y \frac{\partial C_i}{\partial y} = D_{C_i} \frac{\partial^2 C_i}{\partial y^2}. \quad (1)$$

Here C_i is the concentration of the ith impurity component in the carrier gas; V_x and V_y are the components of the gas velocity; D_{C_i} is the diffusion coefficient of the ith component in the gas.

The boundary condition far from the wafer for the concentration C_i is $C_i|_{y\to\infty} = C_{0i}$, where C_{0i} is the concentration of the impurity component in the gas column. This condition reflects the fact that the falloff of the concentration of the reacting component takes place in the thin boundary layer. The boundary condition on the reaction surface (i.e., at $y=0$) is the equation of mixed kinetics

$$D_{C_i}\left(\frac{\partial C_i}{\partial y}\right)\Bigg|_{y=0} = k^* C_i^j,$$

where k^* is the rate constant of the chemical reaction of formation of the material from which the film will grow (e.g., the reaction of decomposition of a metalorganic compound[2]).

Solution of Eq. (1) in the given case does not present difficulties; therefore we can at once write down the expression, important in what follows, for the source strength of the ith component incident on the substrate at the point $y=0$ as a function of the coordinate x

$$g_{gi}(x) = D_{C_i}\left(\frac{\partial C_i}{\partial y}\right)\Bigg|_{y=0} \approx \frac{3^{2/3}}{2} \frac{C_{0i} D_{C_{0i}} \sqrt{0.665} U^{3/4}}{\eta^{1/4} \Gamma(1/3) \sqrt{x+x_0}}, \quad (2)$$

where η is the dynamic viscosity of the gas, U is the velocity of the carrier gas far from the substrate, x is the distance from the edge of the substrate, x_0 is the left boundary of the region in which the process of Ostwald ripening takes place (Fig. 1), $\Gamma(z)$ is the gamma function, whose value can be found in mathematical tables, and z is its argument.

Note that for illustration we have considered here the simplest stationary form of the equation of convective diffusion. To obtain a more accurate solution taking account of details of the technology (e.g., nonstationarity of the process, configuration of the reactor, rotation of the substrate, etc.) it would be necessary to use different numerical methods.[3,4] The technique of using values of the material fluxes obtained by other methods to examine the evolution of an island film is considered below.

Now let us go on to a description of the processes taking place on the substrate surface when growing films by the vapor deposition method.

EVOLUTION OF ISLAND FILMS DURING VAPOR DEPOSITION

Let an ensemble of islands of the same composition, i.e., a single phase,[2)] be found on the surface of the substrate (Fig. 1), having equilibrium shape, e.g., spherical with radius R. Islands of cylindrical shape can be described analogously. A chemical reaction takes place on the substrate surface between the materials being delivered from the carrier gas (see above), described by the equation

$$\nu\mu = k, \tag{3}$$

where μ and ν are the concentrations of the components being delivered to the substrate, and k is the reaction rate constant.

We take the stoichiometric coefficient of the reaction to be unity.

As a result of nonuniformity with length of the delivery of materials from the gas, diffusion fluxes arise over the substrate. The distribution of material over the substrate in this case should be described by the diffusion equation for each of the components. These equations should contain the sink strength of material into the islands of new phase $(d/dt)I(x,t)$, where $I(x,t)=(1/2)\chi\int_0^\infty f(R,x,t)R^3(x,t)dR$ is the volume of material in the islands; $f(R,x,t)$ is the size distribution function of the islands

$$\chi = \frac{1/3\pi(2-3\cos\Theta+\cos^3\Theta}{V_m N_n Q_0};$$

Θ is the angle of contact, V_m is the volume per atom (molecule) in the new phase, N_n is the number of adsorption sites per unit surface area, Q_0 is the amount of material of the new phase at the beginning of the process of Ostwald ripening, and $f(R,x,t)$ is the size distribution function of the islands. The factor 1/2 is included in the expression for $I(x,t)$ so as not to double count the flux sinking to the islands. In the case when the stoichiometric coefficients in Eq. (3) are not equal to unity, this factor must be replaced by the ratio of the corresponding stoichiometric coefficients. The diffusion equation should also include the phase source $d_g^s(x)$, which is determined by the minimum of the sources of the components (2) also with the factor 1/2. Thus, the diffusion equations of the components on the substrate surface with the corresponding boundary conditions has the form

$$\frac{\partial\nu}{\partial t} = D_\nu\frac{\partial^2\nu}{\partial x^2} + \frac{1}{2}d_g^s(x) - \frac{d}{dt}I(x,t), \tag{4}$$

$$\frac{\partial\mu}{\partial t} = D_\mu\frac{\partial^2\mu}{\partial x^2} + \frac{1}{2}d_g^s(x) - \frac{d}{dt}I(x,t), \tag{5}$$

$$x=0;\quad \mu=\mu_0;\quad \nu=\nu_0;$$

$$x=\infty;\quad \mu=\mu_\infty;\quad \nu=\nu_\infty. \tag{6}$$

According to Ref. 7, the system of equations describing the process of Ostwald ripening of an ensemble of islands should also include the equation of continuity for the island distribution function in size space

$$\frac{\partial f(R,t,x)}{\partial t} + \frac{\partial}{\partial R}(f(R,t,x)V_R(R)) = 0,$$

$$f(R,t,x)|_{t=0} = f_0, \tag{7}$$

where f_0 is the initial distribution function, and $V_R(R)$ is the rate of growth of the islands. It was shown in Ref. 7 that $V_R(R)$ is found by solving the diffusion equation for an individual island in the generalized field defined by the remaining islands and depends on the mean concentration of the components in the system. In our case, if this concentration varies weakly over the length of the substrate, it can be set equal to its mean value, employing data from Ref. 7.[3)] In particular, in the case when the rate of growth of the islands is bounded by the rate of formation of chemical bonds on the surface of a particle, i.e., by the rate of incorporation of the components into the crystalline lattice of an island

$$V_R = \frac{2\sigma\beta V_m^2\psi_1(\Theta)\alpha(\Theta)}{K_B TR}\left(\frac{R}{R_{\rm cr}}-1\right), \tag{8}$$

where β is the specific border flux onto an island, σ is the surface tension, K_B is Boltzmann's constant, $R_{\rm cr}$ is the critical radius, and T is the process temperature.

System of equations (4)–(7) is a complete, closed, nonlinear system of equations describing the process of Ostwald ripening in an island film under conditions characteristic of vapor deposition. First we simultaneously solve Eqs. (3)–(5). We subtract Eq. (4) from Eq. (5) and denote the concentration difference $\mu-\nu$ as M. For simplicity, we take the diffusion coefficients of the components to be equal $D_\mu = D_\nu = D$. We introduce the new independent variable

$$m = \frac{x+x_0}{D(t+t_0)}.$$

Equations (4) and (5) then transform to

$$\frac{d^2M}{dm^2} + 2m\frac{dM}{dm} = 0, \tag{9}$$

and boundary conditions (6) transform respectively to

$$m=0;\quad \mu=\mu_0;\quad \nu=\nu_0 \text{ and } M=\mu_0-\nu_0=M_0,$$

$$m=\infty;\quad \mu=\mu_\infty;\quad \nu=\nu_\infty \text{ and } M=\mu_\infty-\nu_\infty=M_\infty.$$

Solving Eqs. (9) and (3) simultaneously, we obtain expressions for μ and ν

$$\mu = \frac{1}{2}(M+\sqrt{M^2+4k}),\quad \nu = \frac{1}{2}(M-\sqrt{M^2+4k}), \tag{10}$$

where

$$M = M_0 - (M_0-M_\infty)\Phi\left(\frac{x+x_0}{\sqrt{D(t+t_0)}}\right),$$

and $\Phi(z)$ is the error function.

Now, knowing the explicit form of the functions $\mu(x,t)$ and $\nu(x,t)$ from Eqs. (4) and (5), we obtain an expression for the total flux to the surface

$$g\Sigma(x,t) = \frac{1}{2}g_g(x) - \frac{d}{dt}I(x,t),$$

at late times t

$$g\Sigma(x,t) = \frac{2k(M_0-M_\infty)^2\exp\left(-\frac{1}{2}\frac{(x+x_0)^2}{D(t+t_0)}\right)}{\pi D(t+t_0)(M_0+4k)^{3/2}}. \tag{11}$$

We write the equation of mass balance on the substrate[8]

$$\int_0^t g\Sigma(x,t) = \chi\int_0^\infty f(R,x,t)R^3(x,t)dR, \tag{12}$$

we substitute the value of the flux (11) in it and, integrating the left-hand side of Eq. (12), we obtain

$$\frac{2k(M_0-M_\infty)^2}{\pi(M_0^2+4k)^{3/2}}\left\{\mathrm{Ei}\left(-\frac{1}{2}\frac{(x+x_0)^2}{D(t+t_0)}\right)-\mathrm{Ei}\left(-\frac{(x+x_0)^2}{D(t_0)}\right)\right\}$$
$$+\frac{3^{2/3}}{2}\frac{D_{C_i}\sqrt{0.665}U^{3/4}}{\eta^{1/4}\Gamma(1/3)\sqrt{x+x_0}}t=\chi\int_0^\infty f(R,x,t)R^3(x,t)dR, \tag{13}$$

where $\mathrm{Ei}(z)$ is the exponential integral function.

We now consider two physically important limiting cases.

1. Let the following relation hold between the terms on the left-hand side of Eq. (14) in the considered interval of space and time:

$$\frac{2k(M_0-M_\infty)^2}{\pi(M_0^2+4k)^{3/2}}\left\{\mathrm{Ei}\left(-\frac{1}{2}\frac{(x+x_0)^2}{D(t+t_0)}\right)-\mathrm{Ei}\left(-\frac{(x+x_0)^2}{D(t_0)}\right)\right\}$$
$$\ll\frac{3^{2/3}}{2}\frac{D_{C_i}\sqrt{0.665}U^{3/4}}{\eta^{1/4}\Gamma(1/3)\sqrt{x+x_0}}t. \tag{14}$$

Physically, this means that the flux of material from the carrier gas to the surface is large and the first term in Eq. (14) can be neglected. Then Eq. (14) takes the form

$$\frac{3^{2/3}}{2}\frac{D_{C_i}\sqrt{0.665}U^{3/4}}{\eta^{1/4}\Gamma(1/3)\sqrt{x+x_0}}t=\chi\int_0^\infty f(R,x,t)R^3(x,t)dR. \tag{15}$$

We introduce the new variables $u=R/R_{\mathrm{cr}}$ and $X=\Delta_0/\Delta(x,t)$, where

$$\Delta(x,t)=\frac{\nu-\nu_\infty}{\nu_\infty}+\frac{u-u_\infty}{u_\infty}$$

is the relative supersaturation of the phase, where Δ_0 is its initial value. As the ''time'' it is necessary to choose the quantity $\tau=\ln(X^2)$, which for large t has a unique relation with the time. Rewriting relation (22) in terms of the new variables, we obtain

$$1=\chi^* e^{\frac{3}{p}\tau}\int_0^p\chi(\tau+\psi)\frac{u^3}{d^s(u)}du, \tag{16}$$

where

$$\chi^*=\frac{3^{2/3}}{2}\frac{\chi\eta^{1/4}\Gamma(1/3)\sqrt{x+x_0}}{D_{C_i}\sqrt{0.665}U^{3/4}},$$

$$\psi=\int_0^u\frac{du}{g(u)}=\ln(2-u)^2-\frac{4}{2-u}-2-\ln 4,$$

$\chi(\tau+\psi)$ is an arbitrary function that awaits to be determined from the condition of conservation of mass

$$g^s(u)=\frac{1}{2u}(u-2)^2,$$

p takes the values $p=2,3,4$ depending on the mechanism of mass transport in the system.[7]

We have reduced Eq. (15) to the form of Eq. (16), which corresponds to the equation solved in Ref. 7, with the sole difference that χ^* in the given case depends on the spatial coordinate x. Note that having transformed expression (2) to the form $g_g t^{n-1}$, we see that this corresponds formally to $n=1$. As was shown in Ref. 8, the material sources are decaying for $n<3/p$, while for $n\geqslant 3/p$ the sources are called nondecaying. Thus, under the stationary conditions considered here (i.e., $n=1$) the source can be both decaying for the mass transport mechanism corresponding to $p=2$, and constant for $p=3$ or growing for $p=4$. Consequently, when depositing different materials the same conditions of the process can lead to the formation of fundamentally different structures, from monodisperse (at a given point in space) for $p\leqslant 3$ to polydisperse with size distribution which will be obtained below. We consider here only the case $p=2$; the remaining cases can be obtained by analogy in accordance with the results of Ref. 8 and the present study.

Without going through the intermediate calculations, we write out the main functional dependences characterizing an island film at the stage of Ostwald ripening when growing the film by vapor deposition under the condition that the distribution of the component be governed mainly by hydrodynamic delivery of material from the gas phase. Thus, in this case the critical radius of the islands is equal to $R_{\mathrm{cr}}^p=R_{\mathrm{cr}0}^p+A_p t$, where $R_{\mathrm{cr}0}$ is the critical radius of the islands at the beginning of the process of Ostwald ripening, and the values of the coefficient A_p can be found in Ref. 7. For example, for $p=2$ the coefficient A_2 is given by

$$A_2=\frac{\beta\sigma V_m^2\psi_1(\Theta)\alpha(\Theta)}{K_B T}.$$

The mean radius of the islands varies in time as

$$R(x,t)=C_{pn}(A_p t)^{1/p}, \tag{17}$$

where

$$C_{pn}=\frac{\int_0^{u_0}P_p^s(u)u\,du}{\int_0^{u_0}P_p^s(u)du},$$

u_0 is a reference point; for $p=2$ this quantity $u_0=2$ (see Ref. 7).

The number of islands varies according to the law

$$N_g(x,t)=\frac{G_{gp}}{(3/p-1)(A_p t)^{3/p-1}\sqrt{x+x_0}},$$

$$G_{gp}=\frac{3^{2/3}}{2}\frac{\gamma_0 D_{C_i}\sqrt{0.665}U^{3/4}V_m R_{k0}^{p-2}N_n}{\eta^{1/4}\Gamma(1/3)\pi A_p}$$
$$\times\int_0^{u_0}\frac{e^{-(3/p-n)\tau(u)}u^3du}{(du^{p-1}/d\tau)^s}, \tag{18}$$

where $\tau(u)=\int_0^{u_0}du/(du^{p-1}/d\tau)^s$ (Ref. 7); N_n is the number of adsorption sites per unit surface area; γ_0 is a coefficient determined in accordance with Ref. 7.

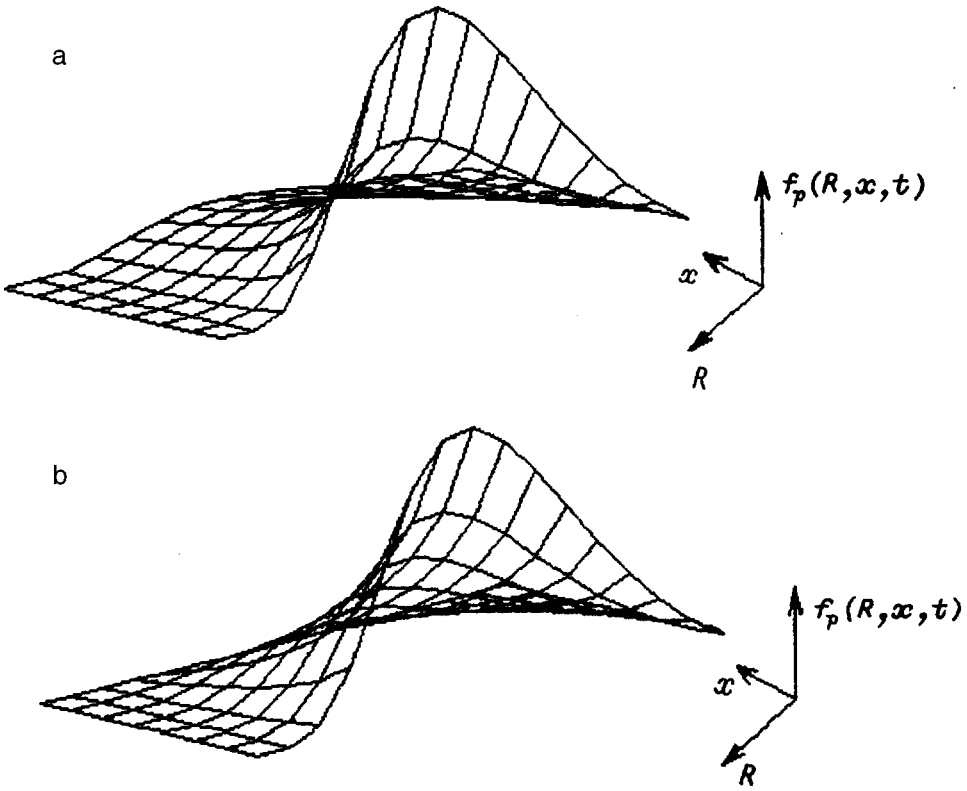


FIG. 2. Dependence of the island distribution function $f_p(R,x)$ on radius R and spatial coordinate x. a — the evolution of the ensemble of islands is governed by the hydrodynamic material source, b — the evolution of the ensemble of islands is governed by diffusion over the substrate.

In general, the size distribution function of the islands has the form

$$f(R,x,t)=\frac{N(x,t)}{R_{\rm cr}(t)}P_p(u). \tag{19}$$

The form of the function $P_p(u)$ was introduced in Refs. 1,6, and 7. For example, for $p=2$

$$P_2(u)=\begin{cases}\dfrac{2e^{3-2n}u\exp\left(-\dfrac{3-2n}{1-u/2}\right)}{(2-u)^{2+2(3/2-n)}} & u<2,\\ 0 & u\geqslant 2,\end{cases}$$

where n is equal to 1 for the stationary case under consideration.

For a nonstationary process n will differ from unity. Figure 2a displays the form of the function $f(R,x,t)$ at the some given moment of time t.

The degree of filling of the substrate in the case of hemispherical islands has the form

$$\xi(x,t)=\frac{4\pi G_{pg}Cpn^2(A_pt)^{1-1/p}}{(3/p-1)\sqrt{x+x_0}}. \tag{20}$$

2. Let us now consider the case when the material flux from the gas phase to the substrate is small in comparison with diffusion of the components over the surface of the substrate (15). This case can be realized, for example, when the concentration of the impurity components in the gas column is small [see Eq. (2)]. In this case Eq. (14) takes the form

$$\frac{2k(M_0-M_\infty)^2}{\pi(M_0^2+4k)^{3/2}}\left\{\mathrm{Ei}\left(-\frac{1}{2}\frac{(x+x_0)^2}{D(t+t_0)}\right)-\mathrm{Ei}\left(-\frac{(x+x_0)^2}{D(t_0)}\right)\right\}$$
$$=\chi\int_0^\infty f(R,x,t)R^3(x,t)dR. \tag{21}$$

Let us consider the expression on the left-hand side of Eq. (21) at late times t. We replace the exponential integral function by its approximate value at late times t

$$\frac{2k(M_0-M_\infty)^2}{\pi(M_0^2+4k)^{3/2}}\left\{\ln\left(\frac{t+t_0}{\gamma}\right)-\ln\left(\frac{1}{2}\frac{(x+x_0)^2}{D}\right)\right\}$$
$$=\chi\int_0^\infty f(R,x,t)R^3(x,t)dR, \tag{22}$$

where $\gamma=1.781\ldots$ is the Euler constant.

We equate the left-hand side of the equation to zero and, solving the resulting equation for x, we obtain the right boundary of the region of space at which the process of Ostwald ripening takes place (Fig. 1)

$$x_b=\sqrt{\frac{2D(t+t_0)}{\gamma}}-x_0.$$

Thus the process of Ostwald ripening of an ensemble of islands takes place in the region $x_0<x<x_b$, which expands with the passage of time. Transforming in Eq. (22) to the variables u and τ, we have

$$\frac{2k(M_0-M_\infty)^2}{\pi(M_0^2+4k)^{3/2}}\ln\left(\frac{2D}{\gamma(x+x_0)^2}\right)$$

$$=\chi e^{\frac{3}{p}\tau}\int_0^p\chi(\tau+\psi)\frac{u^3}{g^s(u)}du-\tau, \tag{23}$$

and in the limit $r\to\infty$ we obtain

$$\frac{2k(M_0-M_\infty)^2}{\pi(M_0^2+4k)^{3/2}}\ln\left(\frac{2D}{\gamma(x+x_0)^2}\right)$$

$$=\chi e^{\frac{3}{p}\tau}\int_0^p\chi(\tau+\psi)\frac{u^3}{g^s(u)}du. \tag{24}$$

It can be seen from Eq. (22) that it is formally similar to the equation of mass balance in Ref. 7; therefore, without going through the intermediate steps, we can immediately write out all the main parameters of an island film as functions of space and time. The mean radius of the islands in this approximation is given by expression (17), and the number of islands per unit surface area is equal to

$$N_d(x,t)=\frac{G_{pd}\ln\left(\frac{2D}{\gamma(x+x_0)^2}\right)}{3/p(A_pt)^{3/p}}, \tag{25}$$

where

$$G_{pd}=\frac{2\gamma_0(M_0-M_\infty)^2V_mR_{k0}^{p-2}N_n}{(M_0^2+4k)^{3/2}\pi}\int_0^{u_0}\frac{e^{-(3/p-n)\tau(u)}u^3du}{(du^{p-1}/d\tau)^s}.$$

The distribution function in this case has the form (19), in which $n=0$ in the expression for $P_p(u)$. The form of this function is shown in Fig. 2b.

The degree of filling of the substrate by the islands varies with time according to the law

$$\xi(x,t)=\frac{4\pi G_{pd}}{3/p(A_pt)^{1/p}}\ln\left(\frac{2D}{\gamma(x+x_0)^2}\right). \tag{26}$$

DISCUSSION OF RESULTS

On the basis of the above analysis we can draw some conclusions which are important for an understanding of the kinetics of film formation by vapor deposition and of methods for controlling its growth. Let us first consider the case of growth of islands of a new phase, controlled by the delivery of material from the carrier gas (15). Note that we have considered the simplest stationary case while most works on the hydrodynamics of vapor deposition consider nonstationary processes in three-dimensional space. The solution in this case, as a rule, is obtained numerically. To consider the kinetics of film growth as set forth in the previous section, we can make use of the following method.

In Refs. 6–8 it was shown that at long times all the external sources can be majorized by polynomials of the type g_gt^{n-1}, where g_g is the sink strength and n is some number $n>0$. Therefore, if we are to consider a nonstationary process it is necessary to majorize the obtained numerical dependence of the material flux to the substrate[3,4] by a function of the form $g_g(x,t)=g_g(x)t^{n-1}$ and consider the evolution of the island film on the surface of the substrate in accordance with the method proposed above. For the stationary flux considered above the following peculiarities of the evolution of an island film may be noted. If in the system the mass transport mechanism corresponding to $p=2$ is realized,[8] then at each point on the substrate the particles will have a wide size spectrum (Fig. 2). The degree of filling of the substrate in this case grows, and the formation time of a continuous film as a function of the spatial coordinate can be estimated by the formula

$$t_{\mathrm{mid}\,g}=\frac{1}{A_p}\left(\frac{(3/p-1)\sqrt{x+x_0}}{4\pi G_{pg}C_{pn}^2}\right)^{\frac{p}{p-1}}. \tag{27}$$

Employing formula (27), we can estimate the process time up to formation of a continuous thin film at a given point of the sample. In the nonstationary case for $n\leqslant 1/p$, according to Ref. 8, a porous noncontinuous film is formed. Note that if we consider a system with islands of different composition, then not only the size but also the composition of the islands will vary in space since there is a connection between these two characteristics.[7,8] If the mechanism of material transport corresponding to $p=3,4$ is realized in the system, then this will correspond to nondecaying material sources[4)] and the film will have a more finely dispersed structure with a practically monodisperse distribution of the islands over size and composition.[8] In this case, if the flux from the gas phase depends weakly on x, then it is possible to obtain a film that is uniform in its structural and phase composition over a large part of the substrate. In the case when the material flux to the substrate from the gas phase is small and the material is redistributed as a result of diffusion over the substrate, then formally there is no material source, the degree of filling of the substrate falls, and as a result a noncontinguous film will grow and the islands will be distributed in size according to the general law given by relation (19). If the growing film consists of islands of different composition, then, in addition the islands in this case will be distributed over composition and as a result will have a nonuniform spatial structure.

We point out that control of the evolution of an ensemble of islands requires a preliminary determination of the growth mechanism of the chosen material on the given substrate. A method of obtaining these data is considered in detail in Ref. 8.

This work was carried out with the partial support of the Russian Fund for Fundamental Research (Project No. 96-03-32396).

[1)] In the present work we consider only the process of Ostwald ripening; we will consider the process of nucleus formation separately.

[2)] We do not consider the evolution of multiphase films here in order not to overshadow the physical essence of the processes taking place.

[3)] It follows from Ref. 7 and the present study that in this approximation the critical and mean radii of the islands do not depend on the spatial coordinate. For a substrate of moderate size, this is in accord with experiment.[1] We will take into account the influence of nonuniformity of the concentration over the length of the substrate on the rate of growth of the islands and their mean and critical radii in a separate communication.

[4)]For the variant of a stationary flux we have considered. To study a nonstationary process, it is necessary to carry out the solution and analysis in accordance with Refs. 7 and 8 and the present work.

[1] S. A. Kukushkin and V. V. Slezov, *Dispersive Systems on Solid Surfaces (Evolutionary Approach): Mechanisms of Thin Film Formation* [in Russian] (Nauka, St. Petersburg, 1996), 312 pp.
[2] G. A. Razuvaev, B. G. Gribov *et al.*, *Metalorganic Compounds in Micro Electronics* [in Russian] (Nauka, Moscow, 1972), 479 pp.
[3] A. N. Vorob'ev, E. A. Garibin *et al.*, Vysokochistye Veshchestva, No. 3, 41 (1996).
[4] E. A. Garibin, I. A. Mironov, S. E. Khorushnikov, and A. N. Vorob'ev, Mater. Sci. Eng., B **39**, 8 (1996).
[5] A. B. Zhidkov, S. A. Kukushkin, E. P. Smirnov, and A. T. Denisova, Poverkhnost', No. 4, 71 (1990).
[6] S. A. Kukushkin and A. V. Osipov, Prog. Surf. Sci. **51**(1), 1 (1996).
[7] S. A. Kukushkin, Fiz. Tverd. Tela (St. Petersburg) **35**, 1582 (1993) [Phys. Solid State **35**, 797 (1993)].
[8] S. A. Kukushkin, Fiz. Tverd. Tela (St. Petersburg) **35**, 1597 (1993) [Phys. Solid State **35**, 804 (1993)].

Translated by Paul F. Schippnick

BRIEF COMMUNICATIONS

Optical properties of liquid gallium–indium alloy

L. A. Akashev and V. I. Kononenko

Institute of Solid State Chemistry, Ural Branch of the Russian Academy of Sciences, 620219 Ekaterinburg, Russia

(Submitted September 2, 1997)
Zh. Tekh. Fiz. **68**, 118–119 (July 1998)

Beatte ellipsometry is used to study the refractive index and absorption coefficient of liquid gallium and of a gallium–30 at. % indium alloy for wavelengths of 0.4–2.0 μm at a temperature at 373 K. The dispersion rel curves of the photoconductivity and reflectivity are calculated from the experimental data. The measured data in the infrared are used to calculate the concentration of conduction electrons and the relaxation rate. It is found that when 30 at. % of indium is introduced into gallium, the concentration of charge carriers does not change, while the relaxation rate of the conduction electrons changes substantially, and this leads to a decrease in the electrical conductivity of the liquid alloy. © *1998 American Institute of Physics.* [S1063-7842(98)01907-2]

Alloys and compounds based on gallium and indium have some extremely valuable properties. They are used in semiconductor devices based on III–V and III–VI compounds and superconductors, as heat transfer agents in nuclear reactors, as lubricating materials, in high temperature thermocouples, etc. There is also a relatively new domain of application of gallium–indium alloys, the fabrication of low temperature solders which are of great importance in semiconductor technology. Their major advantage is that they are used over a wide range of temperatures and can be used to solder ceramic as well as metallic materials. Information on the optical properties of liquid metals and alloys is of great importance for the development of the theory of liquid metal solutions and for solving applied problems concerning the synthesis of materials with specified properties. Data on the reflection and absorption capacities of liquid metal solutions are part of the reference material required for calculations of the crystallization process with allowance for radiative heat transfer, etc.

There is no information about studies of the optical properties of liquid gallium–indium alloys in the literature. In this paper we present data from a study of liquid gallium–30 at. % indium alloy over wavelengths of 0.4–2.0 μm at a temperature of 373 K (the liquidus temperature of this alloy is $T_l = 325$ K), as well as of liquid gallium ($T_{\text{melt}} = 303$ K). The optical constants were measured by Beatte ellipsometry on an LÉFZM ellipsometer.[1] The apparatus on which the measurements were made has been described in detail elsewhere.[2] This device has been upgraded: the vacuum system has been improved (10^{-4}–10^{-5} Pa) and the temperature inside the chamber has been raised (to 1600 K). The optical constants were measured for an angle of incidence of the light on the sample of 82°. The principles of ellipsometry are discussed in Refs. 3 and 4.

In these studies metals of grade Ga−000 and In−000 were used (principal metal content 99.999%). The alloys were synthesized in a vacuum at a pressure of 10^{-4}–10^{-5}Pa and a temperature of 473 K with a hold of 30 min. To obtain a flat, horizontal melt surface, molybdenum crucibles in the shape of a saucer were used (25−30 mm in diameter and 5 mm high). The angle between the crucible wall and its bottom was about 140°, so that the melt surface remained essentially horizontal, despite the comparatively small size of the crucible.

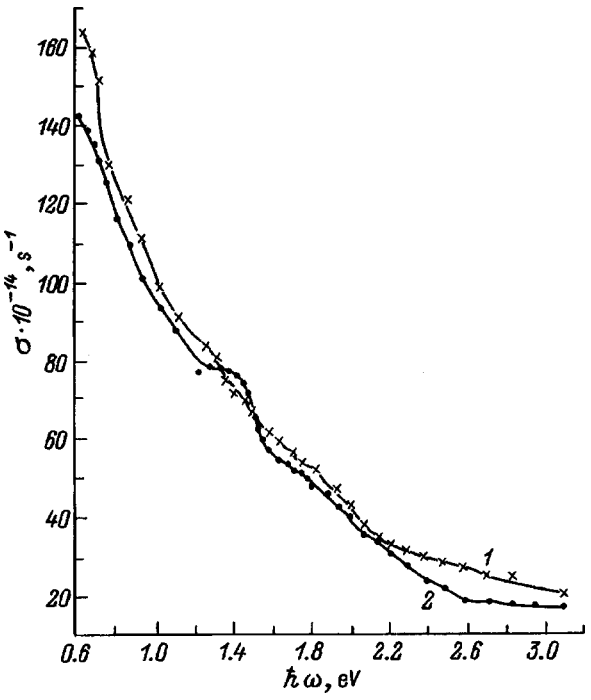


FIG. 1. The photoconductivities of liquid gallium (*1*) and gallium–30 at. % indium alloy (*2*) as functions of photon energy.

The method for measuring the optical constants of liquid metals is described in Ref. 2. An error analysis of the measurements of the optical constants of these melts showed that the error was less than 5%. The high frequency photoconductivity $\sigma(\omega)=nkc/\lambda$ and reflectivity R were calculated in this spectral range (0.40−2.0 μm) from the measured n and k for these liquid systems at 373 K.

There are, as yet, few experimental and theoretical studies of liquid metals, but it has already been reliably established that, within a certain spectral range, the frequency dispersion of the optical coefficients of a number of liquid metals can be explained in terms of a model of almost free electrons, using the Drude–Zener formulas. Compared to solid metals, this range is extended toward the visible.[5] Data on the optical properties of liquid indium at four wavelengths between 0.37 and 2.5 μm have been published by Hodgson[6] and of liquid gallium in the infrared (to 8 μm) by Comins.[7]

Figure 1 shows plots of the photoconductivity of liquid gallium and Ga–30 at. % In alloy as functions of photon energy. Curve *1* (Fig. 1), for liquid gallium, has no anomalies associated with interband electronic transitions. The optical properties of liquid gallium in this range of photon energies are determined by intraband transitions of the conduction electrons.

For the liquid Ga–30 at. % In alloy, a significant absorption band was observed at photon energies between 1.25 and 1.52 eV. This occurs because, when impurity atoms (in this case indium atoms) are introduced into the matrix (gallium atoms), additional electron energy levels develop owing to an interaction of the matrix atoms with the impurity atoms and these new levels form an additional band near the Fermi level (we observed it in this energy range).

Figure 2 shows the dispersion curves of the reflectivity of these melts. The reflectivity R varies from 83 to 90% for liquid gallium (Fig. 2, curve *1*) and from 83 to 87% for the alloy (Fig. 2, curve *2*), and in the near-IR, the reflectivity R of pure gallium is somewhat higher than that of the liquid Ga–30 at. % In alloy.

Using the refractive index and absorption coefficient of these melts measured in the IR, we have calculated N (the concentration of conduction electrons) and γ (the relaxation rate) in the nearly-free-electron model. The concentration of conduction electrons in liquid gallium was $N=1.28\times10^{23}$ cm^{-3} and $\gamma=1.26\times10^{15}$ s^{-1}. Here the maximum photoconductivity, given by $\sigma(0)=Ne^2/m\gamma$, where e and m are the electron charge and mass, is equal to 2.90×10^6 s/m, which is less than the static conductivity[8] $\sigma_{st}=3.76\times10^6$ s/m. This is equivalent to a larger value of the optical relaxation rate, $\gamma_{opt}>\gamma_{st}$.

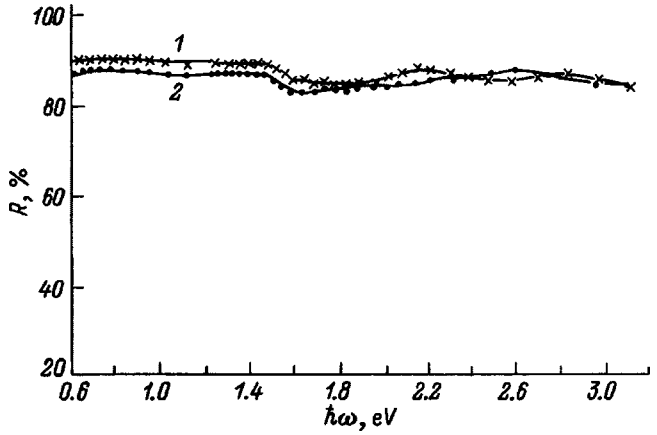


FIG. 2. Dispersion curves of the reflectivity of liquid gallium (*1*) and of gallium–30 at. % indium alloy (*2*).

For the liquid Ga–30 at. % In alloy, the concentration of conduction electrons was $N=1.27\times10^{23}$ cm^{-3}, i.e., hardly changed on introducing 30 at. % In into the gallium; however, the relaxation rate was substantially higher: $\gamma=1.78\times10^{15}$ s^{-1}, which led to a reduction in the maximum photoconductivity of the alloy, $\sigma(0)=2.04\times10^6$ s/m, which is also less than the static conductivity of the alloy, $\sigma_{st}=2.88\times10^6$ s/m.[8]

Therefore, introducing 30 at. % In into gallium causes an additional absorption band to appear in the photoconductivity spectrum of the alloy; here the charge carrier concentration remains constant, but there is a substantial increase in the relaxation rate of the conduction electrons, so that the electrical conductivity of the liquid alloy decreases.

This work was supported by the Russian Fund for Fundamental Research, Project No. 95-03-08005a.

[1] J. R. Beatte, Philos. Mag. **46**, 235 (1955).
[2] L. A. Akashev, V. I. Kononenko, and V. A. Kochedykov, Rasplavy **2**(4), 53-57 (1988).
[3] R. Azzam and N. Bashara, *Ellipsometry and Polarized Light* [North-Holland, Amsterdam (1977); Mir, Moscow (1981), 583 pp.].
[4] A. V. Rzhanov (ed.), *Principles of Ellipsometry* [in Russian] (Nauka, Novosibirsk, 1979), 422 pp.
[5] M. M. Noskov, *Optical and Magnetooptical Properties of Metals* [in Russian] (UNTs AN SSSR, Sverdlovsk, 1983), 225 pp.
[6] J. R. Hodgson, Philos. Mag. **7**, 229 (1962).
[7] N. R. Comins, Philos. Mag. **25**, 817 (1972).
[8] V. I. Kononenko, Doctoral Dissertation [in Russian], Institute of Chemistry, Ural Science Center, Academy of Sciences of the USSR, Sverdlovsk (1972), 332 pp.

Translated by D. H. McNeill

Analysis of the focusing efficiency of pulsed pressure waves as a function of the initial amplitude distribution and temporal profile

É. V. Ivanov

St. Petersburg State Institute of Precision Mechanics and Optics (Technical University), 197101 St. Petersburg, Russia

Yu. V. Sud'enkov

St. Petersburg State University, 198904 St. Petersburg, Russia

(Submitted September 22, 1997)
Zh. Tekh. Fiz. **68**, 120–123 (July 1998)

Solutions of the parabolic diffraction equation for the focusing of pressure pulses with a monopolar or bipolar initial profile and a bell-shaped or annular amplitude distribution are presented. These results are analyzed and conditions determined under which more efficient focusing occurs during the transition from a bell-shaped to an annular distribution of the initial amplitude, as well as when a monopolar pulse is replaced by a bipolar pulse.
 [S1063-7842(98)02007-8]

Studies of the focusing of pulsed elastic waves in liquids are, as a rule, restricted to the case of pulses with a monopolar initial profile and a uniform or bell-shaped radial amplitude distribution. This happens because of limitations which arise on the theoretical, as well as experimental, level on going to more complex problems. In our experiments,[1,2] we have demonstrated the possibility of greatly enhancing the focusing efficiency by using pulsed beams with annular distributions of the initial amplitude, and Komissarova *et al.*[3] have observed analogous phenomena during studies of the focusing of pulses with a bipolar initial profile.

The theoretical analysis of the focusing of pulsed pressure waves with different spatial and temporal characteristics is rather complicated, especially when dissipative and nonlinear effects are taken into account, and can only be carried out numerically. Nevertheless, for some special cases, it is possible to obtain exact analytical results in the linear parabolic approximation which indicate a way to raise the focusing efficiency by optimizing the initial temporal profile of the pulse and the radial distribution of its amplitude.

STATEMENT OF THE PROBLEM

We limit ourselves to axially symmetric pulsed beams, neglect nonlinear and dissipative effects, and assume that

$$k_r/k_z \ll 1, \tag{1}$$

where k_z and k_r are, respectively, the axial and radial components of the wave vector.

In this case, the parabolic diffraction equation[4]

$$(\partial^2 p/\partial r^2 + r^{-1}\partial p/\partial r)c_0/2 = \partial^2 p/\partial z\,\partial\tau \tag{2}$$

with the boundary condition

$$p(z=0,r,\tau) = p_0 f(r)\varphi(\tau + r^2/2Rc_0), \tag{3}$$

can be used to describe the focusing process, where p is the pressure increment, z and r, respectively, are the axial and radial coordinates, $\tau = t - z/c_0$ is the retarded time, c_0 is the speed of sound, $f(r)$ and $\varphi(\tau)$ are the radial amplitude distribution function and the temporal profile of the pulse, normalized to their maxima, R is the radius of curvature of the initial spherical front, and p_0 is the maximum initial amplitude of the pulse.

SOLUTIONS

1. Let us consider the focusing of a pressure pulse with a monopolar initial profile and a bell-shaped amplitude distribution,

$$p_0 = p_0^{(1)}, \quad f(r) = \exp(-r^2/r_0^2),$$
$$\varphi(\tau) = \exp(-\tau^2/\tau_0^2). \tag{4}$$

Using the Fourier transform technique,[4] we find the solution of Eq. (2),

$$p(z,r,\tau) = p_0(2\pi)^{-1}\int_{-\infty}^{+\infty} s(\omega)D(z,r,\omega)\exp(-i\omega\tau)d\omega, \tag{5}$$

where $s(\omega) = \int_{-\infty}^{+\infty}\varphi(\tau)\exp(i\omega\tau)d\tau$, $D = (1 - z/R + iz/d)^{-1}\exp[-(1 - z/R + iz/d)^{-1}(1 + id/R)r^2/r_0^2]$, and $d = \omega r_0^2/2c_0$ is the diffraction (Rayleigh) length, i.e., the pressure field in the pulsed beam is represented by the sum of an infinite number of converging monochromatic Gaussian beams with continuously variable frequency.

The focus of a monochromatic Gaussian beam is always shifted relative to the $z = R$ plane in the direction toward the source.[4] Accordingly, the focus of the pulsed beam described by Eq. (5) must also be located in a plane $z = z_f < R$. For a Gaussian beam with frequency ω the shift in the focus will be more negligible, and the focusing efficiency greater, the better the following condition is satisfied:[4]

$$\omega r_0^2/2Rc_0 \gg 1. \tag{6}$$

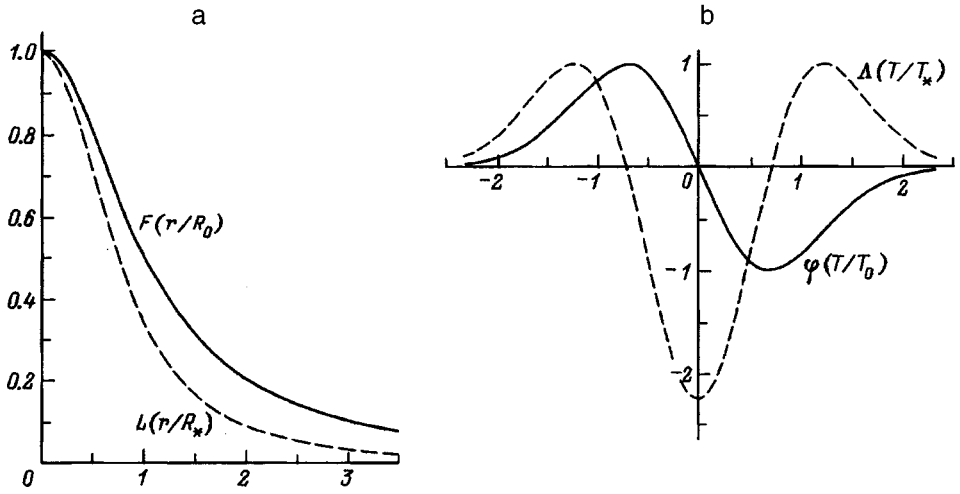


FIG. 1. The radial distribution of the pressure pulse amplitude (a) and its temporal profile (b) in the $z=R$ plane.

Thus, the larger the fraction of the pulse energy attributable to harmonics whose frequencies satisfy Eq. (6), the less the parameters of the pressure field in the planes $z=z_f$ and $z=R$ will differ and the more sharply the pulsed beam will be focused. Since the case of greatest interest is precisely the one with the most efficient focusing, we shall assume that $(R-z_f)/R \ll 1$, i.e., the real focus of the pulsed beam is essentially coincident with its geometric focus. Let us calculate the pressure field at in the plane $r=R$,

$$p(z=R,r,\tau)=p_f^{(1)}F(r/R_0)\phi(T/T_0), \tag{7}$$

where $F(r/R_0)=(1+r^2/R_0^2)^{-1}$ and $\phi(T/T_0)=-(2e)^{1/2}(T/T_0)\exp(-T^2/T_0^2)$ are the radial distribution of the amplitude and the temporal profile of the pulse, normalized to the maximum (Fig. 1), $T=\tau-r^2/2Rc_0$, $T_0=\tau_0(1+r^2/R_0^2)^{1/2}$ is the characteristic pulse duration, $R_0=Rc_0\tau_0/r_0$ is the radius of the beam waist at half maximum, $p_f^{(1)}=p_0^{(1)}G^{(1)}$ is the maximum value of the amplitude of the compression phase of the pressure pulse, and $G^{(1)}=t_0^2\sqrt{2e}Rc_0\tau_0$ is the gain in the compression phase.

2. We shall specify a pulse with an annular distribution of the initial amplitude and a monopolar initial profile as follows:

$$p_0=p_0^{(2)},$$
$$f(r)=K[\exp(-r^2/r_2^2)-\exp(-r^2/r_1^2)],$$
$$\varphi(\tau)=\exp(-\tau^2/\tau_0^2), \tag{8}$$

where $K=(1-x^{-2})^{-1}\exp[2(x^2-1)^{-1}\ln x]$ is a normalizing coefficient with $x=r_2/r_1>1$.

Using the linearity of the problem, we find the corresponding solution in the $z=R$ plane:

$$p(z=R,r,\tau)=p_f^{(2)}[F(r/R_2)\phi(T/T_2) - x^{-2}F(r/R_1)\phi(T/T_1)]/(1-x^{-2}), \tag{9}$$

where $R_{1,2}=Rc_0\tau_0/r_{1,2}$, $T_{1,2}=\tau_0(1+r^2/R_{1,2}^2)^{1/2}$, $p_f^{(2)}=p_0^{(2)}G^{(2)}$, $G^{(2)}=r_2^2\exp[2(x^2-1)^{-1}\ln x]/(2e)^{1/2}Rc_0\tau_0$.

3. We now consider the case of a pulse with a bipolar initial profile and a bell-shaped amplitude distribution:

$$p_0=p_0^{(3)}, \quad f(r)=\exp(-r^2/r_0^2),$$
$$\varphi(\tau)=-(2e)^{1/2}(\tau/\tau_*)\exp(-\tau^2/\tau_*^2). \tag{10}$$

By analogy with Eq. (7), we write

$$p(z=R,r,\tau)=p_f^{(3)}L(r/R_*)\Lambda(T/T_*), \tag{11}$$

where $L(r/R_*)=(1+r^2/R_*^2)^{-3/2}$ and $\Lambda(T/T_*)=(e^{3/2}/2)\times(2T^2/T_*^2-1)\exp(-T^2/T_*^2)$ are the radial distribution of the amplitude and the temporal profile of the pulse, normalized to their maxima (Fig. 1), $R_*=Rc_0\tau_*/r_0$ and $T_*=\tau_*(1+r^2/R_*^2)^{1/2}$ are the characteristic radius of the focal waist and pulse duration, and $p_f^{(3)}=p_0^{(3)}G^{(3)}$ and $G^{(3)}=r_0^2 2^{1/2}/eRc_0\tau_*$.

Unlike the case of the function $\phi(T/T_0)$, the maximum and minimum of the function $\Lambda(T/T_*)$ are not the same in absolute value (Fig. 1b): $|\Lambda(0)|=e^{3/2}/2\cong 2.24$. We shall therefore also determine the maximum amplitude of the rarefaction phase in the focal plane and the corresponding gain,

$$P_f^{(3)}=p_f^{(3)}|\Lambda(0)|=p_0^{(3)}r_0^2(e/2)^{1/2}/Rc_0\tau_*,$$
$$g^{(3)}=P_f^{(3)}/p_0^{(3)}. \tag{12}$$

Thus we have obtained exact analytical solutions of the linear parabolic wave equation which can be used to test various numerical algorithms and can also be used to determine methods for raising the focusing efficiency of pressure pulses by optimizing their spatial and temporal characteristics.

DISCUSSION OF RESULTS

1. We shall compare the results on the focusing of pulses with bell-shaped and annular initial amplitude distributions. The ratio of the corresponding gains gives

$$\ln(G^{(2)}/G^{(1)})=2(\rho_2^2\ln\rho_2-\rho_1^2\ln\rho_1)/(\rho_2^2-\rho_1^2), \tag{13}$$

where $\rho_{1,2}=r_{1,2}/r_0$, with $\rho_1<\rho_2$.

An analysis of Eq. (13) shows that three different situations are possible: for $\rho_2\geqslant 1$, $G^{(2)}/G^{(1)}>1$; for $\rho_2\leqslant e^{-1/2}$, $G^{(2)}/G^{(1)}<1$; and, for $e^{-1/2}<\rho_2<1$, the ratio $G^{(2)}/G^{(1)}$ can be either greater than or less than unity, depending on ρ_1. For example, for the case $\rho_1=1$ and $\rho_2=2$, Eq. (13) predicts a substantial increase in the gain on replacing a bell-shaped initial amplitude distribution by an annular one: $G^{(2)}/G^{(1)}$

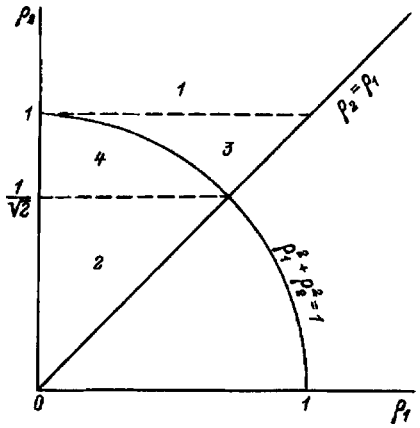


FIG. 2. A diagram of the focal pressures for a pulse with an annular distribution of the initial amplitude.

$=2^{8/3}\cong 6.35$. We shall write down expressions for $G^{(1)}$ and $G^{(2)}$ using the concept of the effective convergence angle β.[5] For axially symmetric, weakly converging beams, the parameter β is defined by

$$\beta^2 \cong 2\int_0^{\alpha} f(\theta)\,\theta\, d\theta, \tag{14}$$

where $\theta = r/R$ is the angle reckoned from the acoustic axis and α is the geometric convergence angle.

It is assumed that the angle α is small enough ($\alpha \leqslant 16°$) to ensure the validity of the parabolic approximation[6,7] but at the same time is large compared to the characteristic apex angles of the wave front, θ_0 and θ_2, where $\theta_{0,2} = r_{0,2}/R$.

Using Eq. (14), we can write the gain coefficients $G^{(1)}$ and $G^{(2)}$ in the form

$$G^{(1,2)} = R(\beta^{(1,2)})^2/(2e)^{1/2} c_0 \tau_0. \tag{15}$$

Therefore, the possible changes in the gain on going from a bell-shaped initial amplitude distribution to an annular one are determined exclusively by the effective convergence angle.

For comparing the focal pressures, we begin with the condition that the acoustic energy on the surface of the initial spherical front is constant,[4,5] i.e.,

$$p_0^2 \int_0^{\alpha} f^2(\theta)\,\theta\, d\theta = \text{const}, \tag{16}$$

where we have taken the axial symmetry of the problem and the smallness of θ into account.

From this we find that the focal pressures for bell-shaped and annular initial amplitude distributions are related by the simple formula

$$p_f^{(2)}/p_f^{(1)} = (\rho_2^2 + \rho_1^2)^{1/2}, \tag{17}$$

which describes an arc of a circle of radius $p_f^{(2)}/p_f^{(1)}$ centered at the coordinate origin and in the (ρ_1, ρ_2) plane for $\rho_1 < \rho_2$. Evidently, as in the case of the ratio $G^{(2)}/G^{(1)}$, three different situations are possible (Fig. 2): for $\rho_2 \geqslant 1$, $p_f^{(2)}/p_f^{(1)} > 1$ (zone *1*); for $\rho_2 \leqslant 2^{-1/2}$, $p_f^{(2)}/p_f^{(1)} < 1$ (zone *2*); and, for $2^{-1/2} < \rho_2 < 1$, the ratio $p_f^{(2)}/p_f^{(1)}$ can be either greater than or less than unity, depending on ρ_1 (zones *3* and *4*, respectively). In particular, substituting $\rho_1 = 1$ and $\rho_2 = 2$ in Eq. (17) gives $p_f^{(2)}/p_f^{(1)} = 5^{1/2} \cong 2.24$. Note that the increase in the focal pressure observed on going from a bell-shaped to an annular initial amplitude distribution is considerably less than the corresponding increase in the gain ($G^{(2)}/G^{(1)} \cong 6.35$). This is explained by a large reduction in the initial pressure p_0 in accordance with the condition (16) that the acoustic energy be constant. Thus these results can be used to estimate the focusing efficiency for pressure pulses, depending on the geometric parameters of the radial distributions of their initial amplitudes.

2. Let us compare the results for the focusing of pulses with monopolar and bipolar initial profiles. In this case, for the gain in the compression and rarefaction phases we obtain the following equations:

$$G^{(3)}/G^{(1)} = 2e^{-1/2}\vartheta \cong 1.21\vartheta,$$

$$g^{(3)}/G^{(1)} = e\vartheta \cong 2.72\vartheta, \tag{18}$$

where $\vartheta = \tau_0/\tau_*$.

Again using the condition that the acoustic energy be constant,[4,5]

$$p_0^2 \int_{-\infty}^{+\infty} \varphi^2(t)\, dt = \text{const}, \tag{19}$$

we also find the ratio of the pressures at the focus,

$$p_f^{(3)}/p_f^{(1)} = 8^{1/2}\vartheta^{3/2}/e \cong 1.04\vartheta^{3/2},$$

$$p_f^{(3)}/p_f^{(1)} = (2e)^{1/2}\vartheta^{3/2} \cong 2.33\vartheta^{3/2}. \tag{20}$$

Thus, replacing a monopolar pressure pulse by a bipolar one can lead to a substantial increase in the gains and focal pressures, especially in the rarefaction phase. It is clear from Fig. 1a that here the diameter of the beam waist will also be contracted. As an analysis of the spectral density functions $s(\omega)$ for the cases of monopolar and bipolar pulses shows, the observed effect involves the transfer of acoustic energy from low frequencies ($\omega \ll 1/\tau_0$) to high frequencies ($\omega \sim 1/\tau_*$).

We have, therefore, found exact analytical solutions to the parabolic diffraction equation for the focusing of pulses with monopolar or bipolar initial profiles with a bell-shaped or annular amplitude distribution.

These results show that the initial temporal profile of a pressure pulse and the radial distribution of its amplitude have a significant effect on the focusing process. In particular, by going from a bell-shaped to an annular distribution of the initial amplitude or by going from a monopolar pulse by a bipolar pulse, it is possible to increase the focusing efficiency substantially. Evidently, the greatest positive effect can be attained by simultaneous optimization of both the temporal and the spatial characteristics of the pulse.

The present analysis in terms of the linear parabolic approximation does not include nonlinear and dissipative effects and, also, is only valid for weakly converging beams.

Despite these limitations, the results obtained here are in qualitative agreement with appropriate experiments for submicrosecond pulses over a wide range of initial pressures and convergence angles[1–3] and can therefore be used for estimating focusing efficiencies.

[1] Yu. V. Sud'enkov and E. V. Ivanov, Pis'ma Zh. Tekh. Fiz. **22**(22), 27 (1996) [Tech. Phys. Lett. **22**(11), 916 (1996)].
[2] Yu. V. Sud'enkov and I. V. Ivanov, *Biomedical Systems & Techologies*, Proc. Soc. Photo-Opt. Instrum. Eng. (SPIE) **2928**, 262 (1996).
[3] I. I. Komissarova, G. V. Ostrovskaya *et al.*, Zh. Tekh. Fiz. **64**(7), 115 (1994) [Tech. Phys. **39**, 694 (1994)].
[4] M. B. Vinogradova, O. V. Rudenko, and A. P. Sukhorukov, *Theory of Waves* [in Russian] (Nauka, Moscow, 1990), 432 pp.
[5] I. N. Kanevskiĭ, *Focusing of Sound and Ultrasonic Waves* [in Russian] (Nauka, Moscow, 1977), 336 pp.
[6] J. N. Tjotta, S. Tjotta, and E. H. Vefring, J. Acoust. Soc. Am. **89**, 1017 (1991).
[7] M. F. Hamilton, J. Acoust. Soc. Am. **92**, 527 (1992).

Translated by D. H. McNeill

Analysis of the conditions for ultradeep penetration of powder particles into a metallic matrix

S. E. Aleksentseva and A. L. Krivchenko

Scientific-Research Institute for Problems of Conversion and High Technologies, Samara State University, 443100 Samara, Russia

(Submitted September 12, 1997)
Zh. Tekh. Fiz. **68**, 124–125 (July 1998)

The critical conditions for ultradeep penetration of particles when a flux of high velocity particles interacts with a matrix material are examined from the standpoint of nonequilibrium thermodynamics. The problem of the change in entropy consistent with their deformation during loading is solved qualitatively for the example of aluminum and copper samples. It is shown that ultradeep penetration of particles is a system process of mass and energy transfer owing to a developed instability in the material caused by the shock-wave action of the particle flux. The degree of disequilibrium of the process is described by the dependence of the change in entropy on the deformation of the matrix material. It turns out that the ultradeep particle penetration process occurs only in a region lying beyond a bifurcation point. © *1998 American Institute of Physics.* [S1063-7842(98)02107-2]

The working of metals by a flux of high velocity particles formed in a shock wave leads to the phenomenon of ultradeep particle penetration. Particles thrown at velocities of 1.5–2 km/s can penetrate into a metallic matrix to a depth of up to 1000 times their diameters, and the pressure during collision is $\sim$15 GPa or higher.[1]

The main task of this paper is to examine the critical conditions under which ultradeep dynamic microdoping takes place. Despite the complexity of the phenomena during shock wave interactions with metals, it is possible to estimate the thermodynamic conditions for the process by examining a number of sequential states through which the matrix passes when a flux of high velocity particles acts on it.

As a dynamic system, a metal is in a stable state before working. After a certain pressure above the dynamic yield point is applied, the system undergoes irreversible changes. According to the second law of thermodynamics and the principle of conservation of energy, a deformed metal in a plastic state is a highly nonequilibrium thermodynamic system.[2] Here the entropy is an index of the reversibility of the process and a measure of the degree of disorder. After the dynamic yield point is reached, the processes are irreversible owing to changes in the entropy.

The metallic matrix is a subsystem which interacts with a flux of high velocity particles formed in a shock wave. The material in the matrix exchanges both energy and matter with this flux, as evidenced by the penetration of the particle material deep into the interior of the matrix. Thus, the interaction of the directed particle flux with the metallic matrix can be regarded as a single thermodynamic system.

The flux of high velocity particles forms a shock wave in the matrix material. Plastic flow of the material sets in at some characteristic pressure p_1, and this corresponds to a sharp bend in the shock adiabat. Table I lists this pressure, along with the dynamic yield point σ_{dyn} and the static yield point $\sigma_{0.2}$, for Al and Cu.[3,4]

It is clear from Table I that when a flux of high velocity particles is applied, the stresses in copper and aluminum matrices exceed the yield points beyond which irreversible plastic deformation sets in. The degree of deformation of these metals as a function of the matrix volume V during pulsed interactions is given approximately by[5] $\varepsilon = 4/3 \ln V/V_0$. When ultradeep particle penetration takes place, the deformation can be 16–21%.

The flow stress for ultradeep penetration of the particles exceeds the Chernov–Lüders deformation. It can be described by the formula[6] $\sigma_T = \sigma_0 \varepsilon^n$, where ε is the plastic deformation, n is a constant which usually $\approx 0.3-0.4$,[7] and σ_0 is a constant for different metals, which, for pulsed loading, has a physical significance corresponding to the elastic limit on the Hugoniot adiabat, in good agreement with the data of Ref. 7.

Plots of $\sigma(\varepsilon)$ for Al and Cu are shown in Fig. 1. These curves clearly are qualitatively correlated and reflect the initial state (point 0), attainment of the dynamic yield point (point c), and a state in which the metallic matrix is deformed with ultradeep penetration of particles (point d).

The states of the deformed metal and the change in the entropy can be characterized using a method proposed by Grigor'ev *et al.*,[8] based on a rheological model for an elastoviscous–plastic body with linear hardening. This model describes a deformed metal with a continuous spectrum of relaxation times and yield stresses τ.

The change in the statistical entropy from the initial state

TABLE I.

Metal	p_1, GPa	σ_{dyn}, GPa	$\sigma_{0.2}$, GPa	$\sigma_{\text{dyn}}/\sigma_{0.2}$
Al	0.22	0.119	0.07	1.7
Cu	0.60	0.233	0.08	2.8

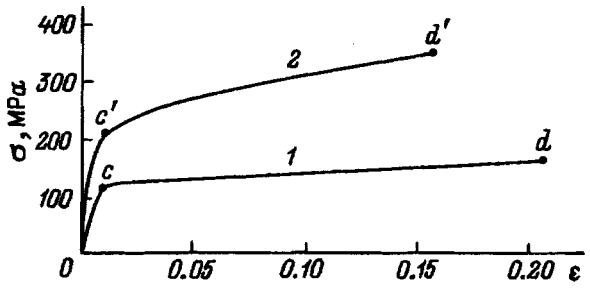


FIG. 1. Deformation state plots of the metal matrix during working with a flux of high velocity particles: *1*—aluminum, *2*—copper.

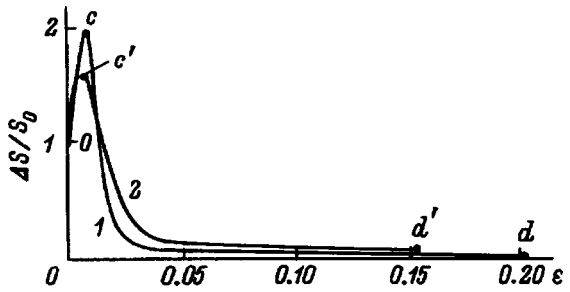


FIG. 2. Relative change in entropy as a function of the degree of deformation in aluminum *1* and copper *2*.

is based on a definition of entropy given by Landau and Lifshitz,[9]

$$\Delta S = -R \int_0^{\tau} f(\tau) \ln f(\tau) d\tau,$$

where R is the universal gas constant and $f(\tau)$ is the probability density characterizing the thermodynamic state of the system as a whole and corresponding to a transition to an irreversible state. The value τ of the dimensionless yield stress reflects the energy state in a local region.

Figure 2 shows plots of the change in entropy with the degree of deformation of the metal during pulsed loading by a flux of high-velocity particles. The transition to a new state is characterized by a change in entropy and a maximum on the curve. After the dynamic yield stress is reached, new structural levels begin to form. This region determines the zone within which ultradeep dynamic microdoping takes place. The sharp drop in the entropy change curve corresponds to and confirms the large amount of experimental data on the avalanche formation of dislocations during shock loading and the conversion of elastic stress energy into the surface energy of a new intergrain boundary. By the time deformations of 5–7% and pressures of ~ 1 GPa have been attained, multiple slip and a cellular structure appear in copper.[4] Pulsed loading causes a change in the interatomic interaction energy. The jump frequency of atoms in the kernel is higher than in an equilibrium lattice and this corresponds to a rise in the diffusion coefficient.

The entropy change shown in Fig. 2 may explain why deep penetration does not occur when a matrix is bombarded by single particles, since the nature of the interaction during single particle bombardment corresponds to the region $0-c$, where the metal is still in an equilibrium state and possesses an energy of the elastic stress fields.

In summary, by examining the critical conditions for occurrence of the ultradeep particle penetration process, we have established that a metallic matrix experiences a complex of states and a change in its thermodynamic potential in accordance with the amount of deformation. Ultradeep particle penetration is a systemic process of energy and mass transfer resulting from a developed instability in the material owing to the shock interaction of the high-velocity particles. Below the bifurcation point, the process probably cannot occur.

[1] S. K. Andilevko, V. A. Shilkin, S. M. Usherenko *et al.*, Int. J. Heat Mass Transf. **36**, 1113 (1993).
[2] G. Nicolis and I. Prigogine, *Self Organization in Nonequilibrium Systems* [Wiley, New York (1977); Mir, Moscow (1979), 512 pp.].
[3] F. A. Baum, M. P. Orlenko, and K. P. Stanyukovich, *The Physics of Explosions* [in Russian] (Nauka, Moscow, 1975), 704 pp.
[4] G. N. Épshteĭin (ed.), *The Structure of Metals Deformed by Explosions* [in Russian] (Metallurgiya, Moscow, 1988), 280 pp.
[5] R. Prümmer, *Explosivverdichtung Pulvriger Substanzen* [Springer-Verlag, Berlin (1987); Mir, Moscow (1990), 128 pp.].
[6] R. W. Cahn and P. Haasen (eds.), *Physical Metallurgy* [North-Holland, Amsterdam (1983); Vol. 3, Metallurgiya, Moscow (1987), 663 pp.].
[7] George E. Dieter Jr. (ed.), *Mechanical Metallurgy* (McGraw-Hill, New York, 1961), 615 pp.
[8] A. K. Grigor'ev, N. G. Kolbasnikov, and S. G. Fomin, *Structure Formation during Plastic Deformation of Metals* [in Russian] (Izd. Univ., St. Petersburg, 1992), 244 pp.
[9] L. D. Landau and E. M. Lifshitz, *Statistical Physics* [Pergamon Press, London (1980); Nauka, Moscow (1976), 680 pp.].

Translated by D. H. McNeill

Numerical analysis and synthesis of electron-optical systems of complex structure. II

S. K. Dyomin, S. I. Safronov, and R. P. Tarasov

Scientific-Research Institute of Pulse Technique, 115304 Moscow, Russia
(Submitted June 25, 1996; resubmitted December 25, 1997)
Zh. Tekh. Fiz. **68**, 126–129 (July 1998)

The numerical analysis and synthesis of intrinsically three dimensional electron-optical systems with complex structures are examined.
[S1063-7842(98)02207-7]

INTRODUCTION

In an earlier examination[1] of the numerical synthesis of complex electron-optical systems, primary attention was devoted to the modeling of systems whose electrodes form a surface S, or have a finite symmetry group, or admit supplementation to a surface $S_1 \supset S$ with a finite symmetry group. In this case, a numerical solution of the corresponding boundary-value integral equation on the surface S can be obtained with sufficient accuracy using the methods of Sec. 3 of Ref. 1. In the meantime, a multiply connected boundary surface S either may not have a finite symmetry group at all or may have such a group, but of low order; however, its components corresponding to the individual cascades of an electron-optical system can be described by higher-order symmetry groups. This last point is especially important in modeling specific electron beam devices, since, because of their design or in setting the technological tolerances during assembly, the symmetry of the system as a whole (if it existed) may be destroyed.

The domain of applicability of the methods of Ref. 1, which explicitly take into account the symmetry of the boundary surface, can be extended significantly by using them together with iterative algorithms. In fact, let the boundary surface S of the original boundary-value problem be asymmetric, but permit a representation of the form

$$S=\bigcup_{i=1}^{M} S_i; \quad S_i \cap S_j = 0; \quad i \neq j, \quad i,j = 1, 2, ..., M,$$

where some of the S_i either have a finite symmetry group or admit supplementation to symmetric elements S_i'.

If each element of the boundary, S_i, $i=1, 2, \ldots, M$, in this case is treated independently of the rest, as a separate element located in the external field of the remaining elements, then it is possible to construct an iterative process, each step of which reduces to solving the appropriate integral equation on an element S_i of the boundary S. In the course of formalizing this process, it turns out that it is an operator analog of the matrix method of Seidel, in which, in each step, an operator matrix, whose elements are operators corresponding to one section S_i of the boundary S, is treated. Since the economical algorithms described in Ref. 1 can be used to handle these operators, this iterative process makes it possible to reduce greatly the volume of calculations. In addition, the order of the matrices being handled in a numerical implementation of this algorithm is considerably reduced, and this makes the numerical process of finding an approximate solution to the original boundary value problem more stable. Naturally, not only Seidel's matrix method but also the entire class of triangular partition methods, at least, is suitable for realizing this analysis.

ITERATIVE METHODS

Let the set of boundary points of a Dirichlet, Newmann, or Cauchy boundary-value problem for the Laplace equation (problem (9) of Ref. 1) form a surface S of the form

$$S=\bigcup_{i=1}^{3} S_i, \quad S_i \cap S_j = 0, \quad i \neq j, \tag{1}$$

where S_1 has a finite symmetry group $\{\tau_k\}$, $k=1,2,\ldots,N_1$ of order N_1, and the components S_2 and S_3 of the surface S are asymmetric, but a surface $\overline{S_2}: S_2 \subset \overline{S_2}$ with a finite symmetry group $\{\overline{\tau_k}\}$, $k=1, 2, \ldots, N_2$ of order N_2, exists such that the area of S_2 is greater than the area of $\overline{S_2} \backslash S_2$. Then the boundary integral equation (Eq. (10) of Ref. 1) corresponding to the original boundary value problem with boundary surface S of the form (1) can be represented in the form

$$\sum_{j=1}^{3} A_{ij}\rho_j(x) = \varphi_i(x), \quad x \in S_i, \quad i=1, 2, 3, \tag{2}$$

where

$$[A_{ij}u](x) = \chi_{S_i}(x) \int A(x,x_1)u(x_1)d\mu(x_1),$$

$$x \in S_j, \quad x \in S_i.$$

It is easy to show that the operator $l(S_1)$, which is a contraction of the boundary operator l of the original boundary-value problem to the surface $S_1: l(S) \to l(S_1)$, is invariant under transformations from the group $\{\tau_k\}$,

$$T_k^{-1} l(S_1) T_k = l(S_1), \quad T_k u(x) = u(\tau_k^{-1} x),$$

$$\forall \tau_k \in \{\tau_k\}, \quad k=1, 2, ..., N_1. \tag{3}$$

In addition, the operator $l(S_2)$, which is a contraction of the boundary operator l to the surface $S_2: l(S) \to l(S_2)$, can be extended to the surface $\overline{S_2}: l(S) \to l(\overline{S_2})$, so that

$$\overline{T_k}^{-1} l(\overline{S_2}) \overline{T_k} = l(\overline{S_2}), \quad \overline{T_k} u(x) = u(\overline{\tau_k}^{-1} x),$$

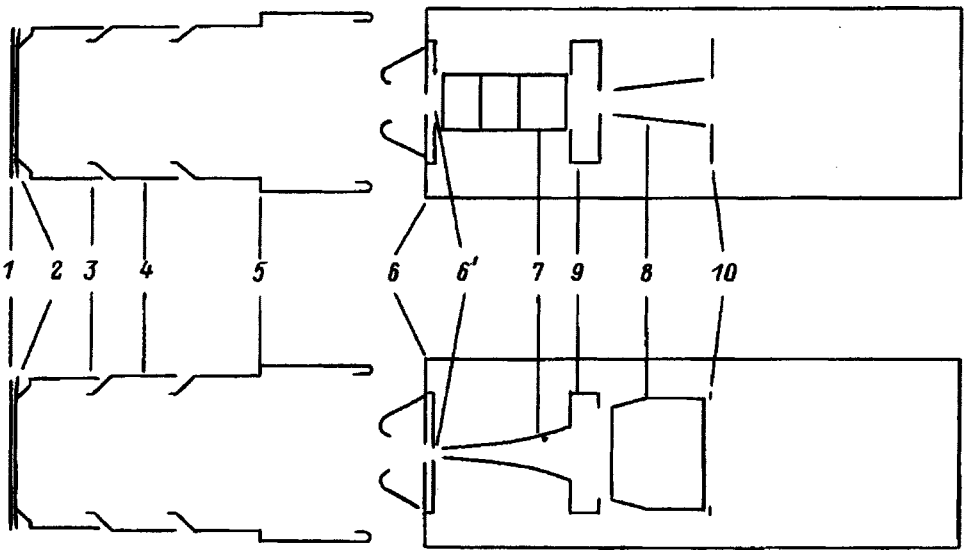


FIG. 1.

$$\forall \overline{\tau_k} \in \{\overline{\tau_k}\}, \quad k=1, 2, ..., N. \tag{4}$$

For numerical solution of Eq. (2) we shall use the iterative process employing one of the matrix triangular iteration methods,

$$(D+\omega\hat{L})\,\frac{u_{k+1}-u_k}{\omega}+Au_k=f, \quad k=1, 2, ...,$$

where D and $\hat{L}$ are operators from the expansion of $A=\|A_{ij}\|$ to a sum of diagonal, lower, and upper triangular matrices, $A=D+\hat{L}+U$, i.e., for Eq. (2) we have

$$A_{11}u_1^{(n+1)}=(1-\omega)A_{11}u_1^{(n)}+\omega[\varphi_1-A_{12}u_2^{(n)}-A_{13}u_3^{(n)}],$$

$$A_{22}u_2^{(n+1)}=(1-\omega)A_{22}u_2^{(n)}+\omega[\varphi_2-A_{21}u_1^{(n+1)}-A_{23}u_3^{(n)}],$$

$$A_{33}u_3^{(n+1)}=(1-\omega)A_{33}u_3^{(n)}+\omega[\varphi_3-A_{31}u_1^{(n+1)}-A_{32}u_3^{(n)}]. \tag{5}$$

Equation (3) implies that $T_k^{-1}A_{11}T_k=A_{11}$, $k=1, 2, \ldots, N_1$, i.e., the algorithms of Sec. 3 of Ref. 1 can be used for numerically solving the first of Eqs. (5) in each stage of the iteration. In turn, Eqs. (4) mean that for an efficient numerical solution of the second of Eqs. (5), in each step of the iteration one can use the reduction method, i.e., the inverse operator to $A_{22}(S_2)$ is constructed on the basis of the inverse operator to $\overline{A}_{22}(\overline{S_2})$, which is the extension of the operator A_{22} to the surface $\overline{S_2}$.

Therefore, using matrix triangular iteration methods makes it possible to take into account the symmetry of the

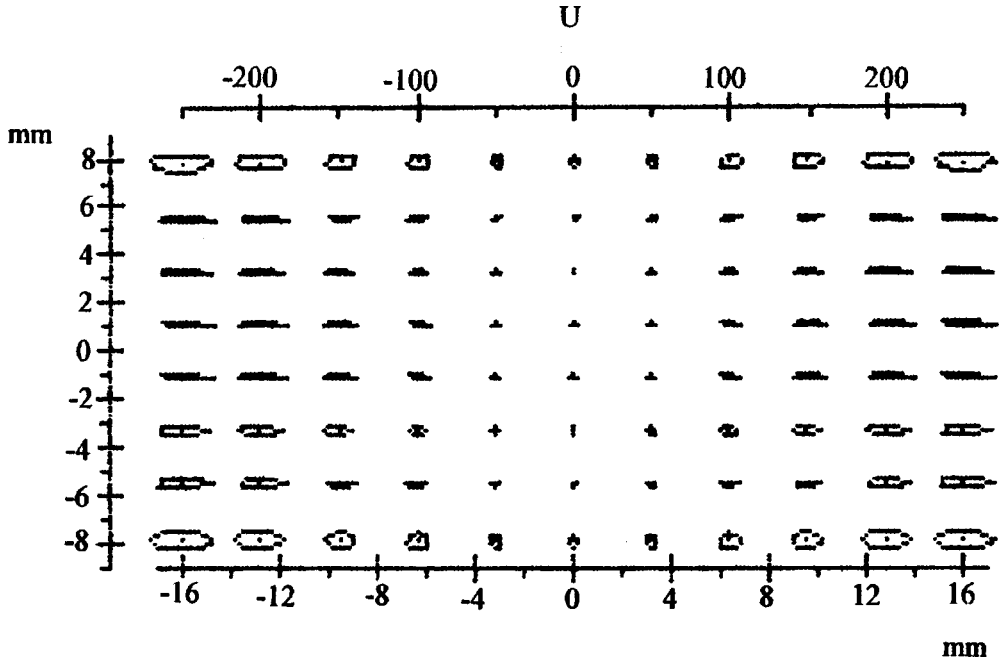


FIG. 2.

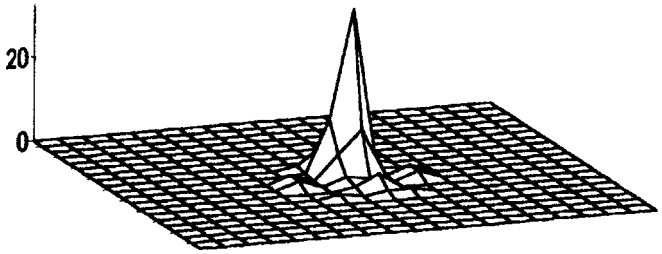


FIG. 3.

components of a surface S which is asymmetric on the whole, while representing S in the form of symmetric boundary elements or of boundary elements which admit supplementation to symmetric elements serves, in turn, as an algorithm for constructing the operator $B=D+\omega\hat{L}$ of the desired iteration process. In addition, it is known that iteration methods of this class have a rather high rate of convergence, and their major disadvantage is related to the need to deal with a diagonal matrix D whose dimensions may be large for a grid problem. Furthermore, if D has some symmetry, then in order to construct D^{-1} one can use the economical methods of Ref. 1. This makes it possible to reduce greatly the machine time employed, both for organizing the iterative process and for obtaining an approximate solution, as a whole. Reducing the size of the matrices dealt with, in turn, makes the process of finding an approximate solution more stable computationally, and this is especially important in solving boundary-value problems with a complicated boundary. And last: since in each step of the iteration process we are dealing with an integral operator associated with an individual element S_i of the surface S that has a certain symmetry group, including modules corresponding to symmetries of other types in a program for the given symmetry presents no special difficulty.

MODELING SPECIFIC ELECTRON-OPTICAL SYSTEMS

The mathematical model described in Ref. 1 and this paper was implemented in a software package of modeling programs permitting the numerical analysis and synthesis of complex, three dimensional electron-optical systems. Thus, Fig. 1 shows projections of the actual design of an electron-optical image converter on two orthogonal planes, YZ and XZ, passing through the device axis Z. This device consists of a photocathode (disk *1*), a grid modeled as a transparent electrode (disk *2*), accelerating *3*, intermediate *4*, and focusing *5* electrodes in the form of surfaces of revolution whose generators are piecewise-continuous curves consisting of line segments and circular arc segments. The anode chamber *6* has a rather complicated configuration that includes a rectangular slit *6′* as well as two circular apertures. Inside the anode chamber are two pairs of plates: deflection plates *7* and calibration plates *8*. The deflection plate consists of three rectangular segments set at fixed angles relative to one another. The calibration plate is broken into two parts: rectangular and trapezoidal (they can, anyway, be positioned at specified angles). Inside the anode chamber there are also two diaphragms (*9* and *10*). Diaphragm *9* is a cylinder with rectangular slits in the ends, and diaphragm *10* is a disk with a rectangular slit. These elements of the electron-optical system can be oriented arbitrarily in space, as is allowed by the program for modeling electron-optical systems. In particular, an electron-optical system with the photocathode inclined to and displaced relative to the axis of the device (design specifications) was simulated numerically, the effect of displacements of the deflection plates *7* and *8* relative to the device axis, separately and together, was studied in order to determine the assembly tolerances, etc. The numerical experiments, and comparisons of these with data from real experiments, showed that good accuracy in achieved in the calculations when the order of discretization in this model for the electron-optical system is 5000 or more, i.e., the original problem is reduced to a system of linear algebraic equations of dimensions $[5000\times 5000]$ or higher. The device can be modeled with specified potentials on its electrodes or by calculating the influence functions based on the principle of superposition with subsequent choice of a specific potential distribution that satisfies given specifications. The electrostatic field of the device was also calculated in two ways. If it is required to determine a small number of electron trajectories (on the order of a few dozen), then it is more

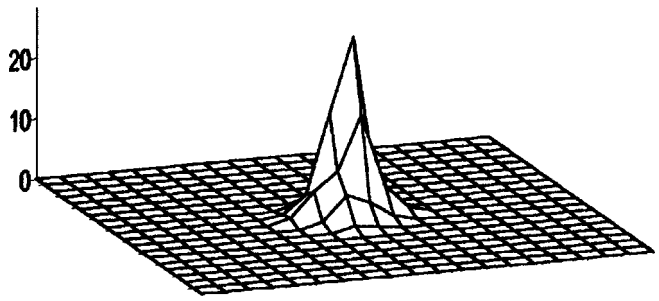


FIG. 4.

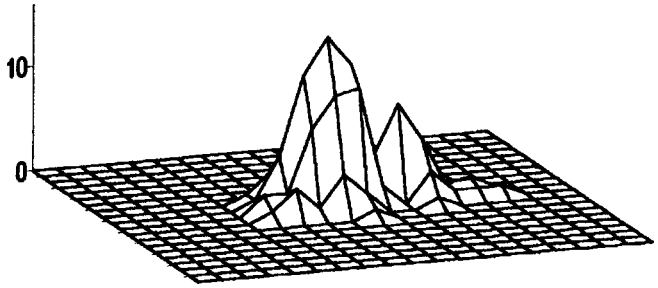


FIG. 5.

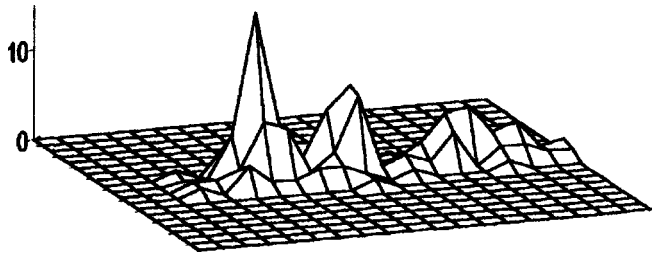


FIG. 6.

economical to calculate the field in the device only along the electron trajectories. In the other case, a spatial grid is constructed and the unknown values of the potential are calculated at its nodes.[1] With a suitable approximation, the latter method makes it possible to determine the device field at any point of the working region economically. With this approach, the time spent calculating an individual trajectory is minimal, so, if needed, it is possible to determine hundreds or thousands of electron trajectories.

This mathematical model of an electron-optical system and its computer implementation can be used to determine various physical characteristics of a device with fair accuracy. For example, Fig. 2 shows the (magnified) scattering ellipses of electron beams from point sources positioned along a single line. With different potentials on the deflection plates, this kind of numerical calculation makes it possible to determine the magnification, as well as such model characteristics as spatial resolution, distortion, chromatic aberration, and, if necessary, the influence of asymmetric components.

As an example, Fig. 3 shows the scattering function from a point source (the impulse characteristic of the system, i.e., the response of the system to an impulse function) with coordinates $Y=5$ mm and $X=0$ when the potential on the deflection system (7) equals the potential $u_0=15$ kV on the anode chamber (6). For the same values of the potential, Fig. 4 shows the scattering function from a point source at $Y=0$ and $X=0$. Figures 5 and 6, in turn, show the scattering function from the same point source, but with potentials on the plates of the deflection system (7) of, respectively, $u=u_0\pm 50$ V and $u=u_0\pm 100$ V, where u_0 is the potential on the anode chamber (6).

[1] S. K. Demin, S. I. Safronov, and R. P. Tarasov, Zh. Tekh. Fiz. **68**(2), 97 (1998) [Tech. Phys. **43**, 222 (1998)].

Translated by D. H. McNeill

Calculating the thermodynamic parameters of dense gases and weakly ionized plasmas including three-body interactions

A. I. Vovchenko, I. A. Mulenko, and V. B. Soloveĭ

Institute of Pulsed Processes and Technology, National Academy of Sciences of Ukraine, 327018 Nikolaev, Ukraine

(Submitted September 7, 1997)
Zh. Tekh. Fiz. **68**, 130–132 (July 1998)

A model is developed for the equation of state of dense gases and nonideal plasmas in the approximation of three-body interactions. The reduced third virial coefficient is calculated for a number of spherically symmetric pairwise additive interaction potentials of neutral and charged particles. Its temperature dependence is constructed for various potentials. The density and composition of plasmas in a number of pure substances are calculated numerically.
 [S1063-7842(98)02307-1]

Increasing the accuracy of calculations of the thermodynamic parameters of multicomponent nonideal gases and high-density multicomponent plasmas usually requires that higher orders of the virial expansion of the thermodynamic functions in powers of the density be taken into account. The main problem then reduces to calculating the higher virial coefficients. Calculating the thermodynamic parameters of gases and plasmas in the approximation of the second, third, etc., virial coefficients involves including particle interactions of the corresponding orders.[1] Calculations of this sort have been done by Semenov[2] and by Kriger *et al.*[3] These theoretical calculations are in fully satisfactory agreement with experimental data for the alkali metal vapors.

In computing the thermodynamic parameters of partially ionized plasmas containing a large number of particle species, one encounters the problem of calculating the virial coefficients for different potentials, which describe the interactions between neutral particles of the same or different species, as well as the interaction between neutral and charged particles. Corrections for the interaction of free charges are then conveniently taken into account by the method of correlation functions.

The purpose of this paper is to calculate the thermodynamic functions and composition of dense gases and weakly ionized atom–molecule plasmas of several pure substances in the approximation of three-body interactions among all the species of neutral and charged particles.

The higher virial coefficients are usually calculated in terms of the group integral method of Mayer.[4] Thus, the third virial coefficient has the form[4]

$$C(T) = -\frac{2}{3}\beta_2 N^2, \tag{1}$$

where N is the number of particles in the gas, β_2, known as the second irreducible group integral, is defined by

$$\beta_2 = \frac{1}{2}\int_V\int f(r_{12})f(r_{13})f(r_{23})d\mathbf{r}_1 d\mathbf{r}_2, \tag{2}$$

where $f(r) = \exp(-U(r)/T) - 1$ is the Mayer function, $U(r)$ is the potential energy of interaction of two particles separated by a distance r, and T is the gas (plasma) temperature in energy units.

Integrals of the form (2) are not calculated analytically, even in the simplest case of pairwise additive spherically symmetric potentials. Direct numerical calculation according to Eq. (2) requires large amounts of computer time or is done with some loss of accuracy by choosing one or another way of limiting the domain of integration.[2,3]

The problem of calculating the third virial coefficient is greatly simplified by using the Fourier transform[5]

$$C(T) = -\frac{4\pi}{3}(2\pi)^{3/2}N^2\int_0^\infty dk k^2[\tilde{f}(k)]^3. \tag{3}$$

The Fourier transform of the Mayer function, $\tilde{f}(k)$, is calculated using the formula

$$\tilde{f}(k) = \frac{4\pi}{(2\pi)^{3/2}k}\int_0^\infty r\,\sin(kr)f(r)dr. \tag{4}$$

As a result of using the Fourier transform procedure, the sixfold integral (2) is reduced to a double integral, which leads to considerable savings of computer time. The third virial coefficient was calculated using Eqs. (3) and (4) for Lennard-Jones potentials.[6] We have used the 12–6 and 12–4 potentials characteristic of the interaction between neutral particles and that of neutral particles with charged particles, respectively. The temperature dependences of the reduced third virial coefficient,

$$\tilde{C}(T^*) = C(T^*) \Big/ \left(\frac{2}{3}\pi N_A r_m^3\right)^2$$

for different interparticle interaction potentials are shown in Fig. 1. ($T^* = T/\varepsilon$ and N_A is Avogadro's number.) Besides the Lennard-Jones potentials, the figure shows the exp–6 potential[5] with different values of the fitting parameter α, for which there are published data.[5,7,8] The calculations were done for values of the reduced temperature $T^* = 0.2 - 10^3$. Extending the range of integration in Ref. 3 near the singular

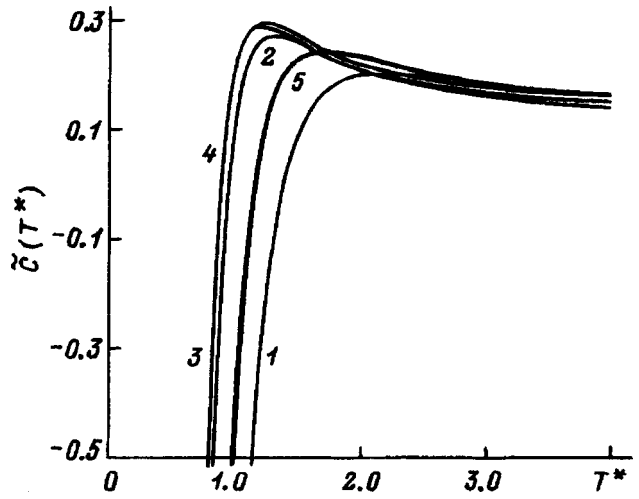
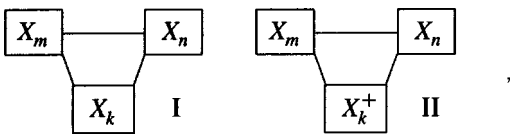


FIG. 1. The temperature dependence of the reduced third virial coefficient for different interparticle interaction potentials: *1,2*—Lennard-Jones 12–4 and 12–6; *3,4*—the potential of Ref. 5 for $\alpha = 13.5$ and 12; *5*—the Lennard-Jones potential for configuration II.

point $r=0$ made it possible to increase the accuracy of the calculation of $C(T)$ compared to the data of Ref. 5 at low temperatures and to extend the temperature interval to $T^*=0.2$.

We now consider a multicomponent plasma containing M particle species. We take the interactions among the charged particles into account in the Debye approximation, assuming that the ions produced by ionization of the many-electron atoms and molecules have nonzero intrinsic linear dimensions (i.e., ion core). We take the interactions involving neutral particles into account in the approximation of the third virial coefficient. In order to retain the interpolation properties of the equation of state in the near-critical region for the second virial coefficient of any pair of interacting neutral, as well as a neutral and charged, particles we shall use the van der Waals approximation. We shall assume that the electron component is weakly degenerate.

In calculating the contributions of three-particle interactions we shall consider the following configurations:



where X is the symbol of a chemical element and $n,m,k=1,2$.

The correction to the free energy owing to three-body interactions in a multicomponent plasma will have the form

$$F_3 = -TV\left\{ -\frac{1}{2}\sum_{i=1}^{M} n_i^3 C_{iii} - \frac{3}{2}\sum_{\substack{i=1\\(j\neq i)}}^{M}\sum_{j=1}^{M} n_i n_j C_{ijj} - 3\sum_{i=1}^{M-2}\sum_{j=i+1}^{M-1}\sum_{k=j+1}^{M} n_i n_j n_k C_{ijk}\right\}. \quad (5)$$

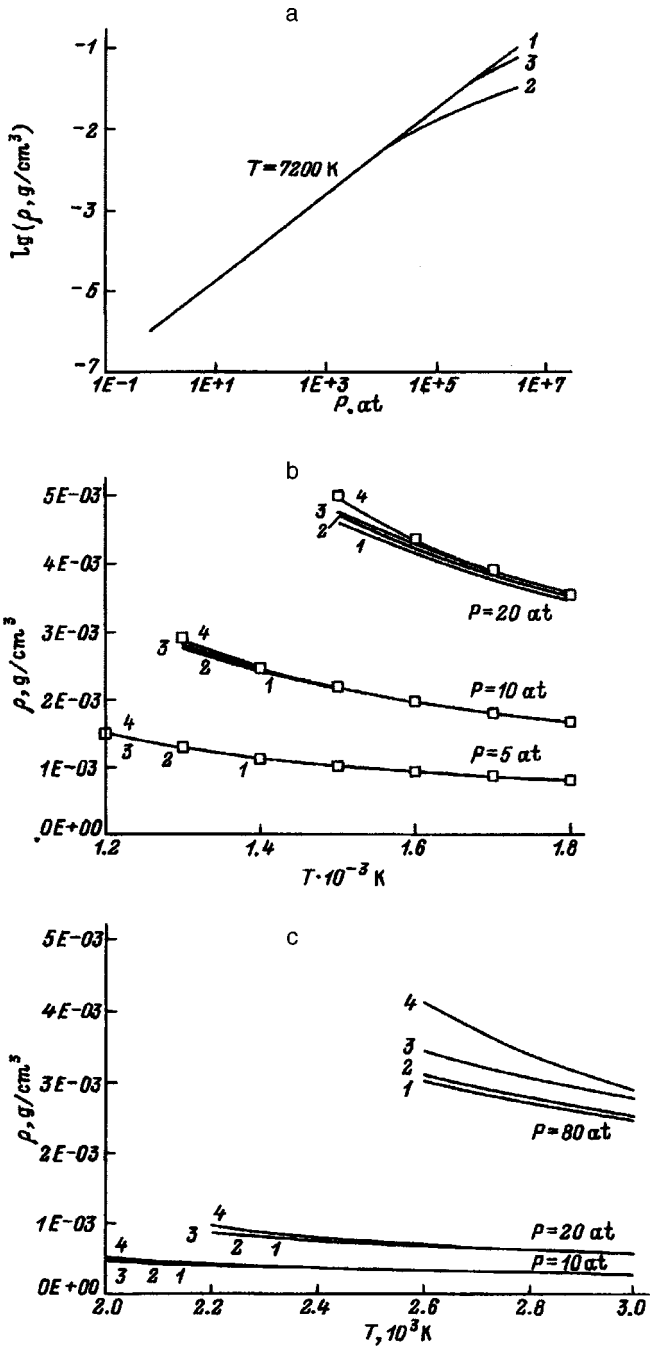


FIG. 2. The density of plasmas in several substances as a function of pressure and temperature for different models. (a) Hydrogen: *1*—ideal gas model, *2*—Ref. 11, *3*—this paper; (b) sodium: *1*—ideal gas model, *2*—two-body interaction approximation,[9] *3*—this paper, *4*—Ref. 2, (box) experiment;[10] (c) lithium: *1*—ideal gas model, *2*—two-body interaction approximation,[9] *3*—this paper, *4*—Ref. 3.

Using the expressions for the free energy of a multicomponent plasma calculated previously[9,10] in the approximation of pairwise interactions, we finally obtain

$$F = -TV\left\{ \sum_{i=1}^{M} n_i \ln\frac{eZ_i}{n_i}\left[1-\sum_{j=1}^{M} n_j b_{ij}\right] + \frac{1}{T}\sum_{i=1}^{M} n_i \sum_{j=1}^{M} n_j a_{ij} - \frac{n_e^2\lambda_e^3}{2^{7/2}} + \frac{e^2}{3T}\sqrt{\frac{4\pi e^2}{T}}\left[\sum_{i=1}^{M} n_i z_i^2\right]\right.$$

$$1-\frac{3}{4}\sqrt{\frac{4\pi e^2}{T}}\frac{\Sigma_{i=1}^{M} n_i z_i^2 \Sigma_{j=1}^{M} n_j z_j^2 r_{ij}}{[\Sigma_{i=1}^{M} n_i z_i^2]^{3/2}}\Bigg\}\Bigg\}+F_3. \qquad (6)$$

If the free energy of the plasma is known, then it is easy to obtain expressions for all its remaining thermodynamic functions, as well as an equation of state and a system of equations for the ionization and chemical equilibrium; in this way a closed thermodynamic model has been constructed.

In the framework of a developed model for the equation of state of a nonideal plasma, we have made a massive calculation of the thermodynamic functions and composition of plasmas of a number of pure substances: hydrogen, oxygen, nitrogen, alkali metal vapors, water vapor, etc. As an example, Fig. 2a–2c shows the calculations for the density of hydrogen, lithium, and sodium plasmas. A comparison is given there with theoretical calculations[2,3,10] based on different model assumptions, as well as with experimental data.[11] At high pressures there is a substantial deviation from the ideal gas model. Along with thermal ionization and dissociation, there is a substantial rise in the contribution of pressure-induced ionization and dissociation.

[1] L. D. Landau and E. M. Lifshitz, *Statistical Physics* [Pergamon Press, Oxford, 1980; Nauka, Moscow, 1976].
[2] A. M. Semenov, DAN SSSR **278**, 866 (1984) [Sov. Phys. Dokl. **29**, 824 (1984)].
[3] V. A. Kriger, V. Yu. Lunin, and A. M. Semenov, Teplofiz. Vys. Temp. **24**, 482 (1986).
[4] M. P. Vukalovich and I. I. Novikov, *Equations of State of Real Gases* [in Russian], GÉI (1948).
[5] M. B. Bekker and L. S. Turchinskaya, Teplofiz. Vys. Temp. **25**, 1012 (1987).
[6] E. A. Mason and T. H. Spurling, *The Virial Equation of State* [Pergamon Press, Oxford, 1969; Mir, Moscow (1972), 280 pp.].
[7] R. Bergeon, Res. CNRS **9**, 171 (1958).
[8] R. Bergeon, F. Ceshino, L. Hennebutte, and B. Vobar, C. R. Acad. Sci. **240**, 1320 (1955).
[9] Yu. A. Bykov, V. M. Grodnov, I. A. Mulenko, and A. L. Khomkin, Teplofiz. Vys. Temp. **34**, 853 (1996).
[10] D. Saumon and G. Chabrier, Phys. Rev. A **46**, 2084 (1992).
[11] J. R. Stone, C. T. Ewing, J. R. Span *et al.*, J. Chem. Eng. Data **11**, 309 (1966).

Translated by D. H. McNeill

Growth of negative filamentary crystals during spark cutting of silicon carbide

V. A. Karachinov

Yaroslav the Wise Novgorod State University, 173003 Novgorod, Russia
(Submitted September 2, 1997)
Zh. Tekh. Fiz. **68**, 133–135 (July 1998)

Experiments are described which show that under pulsed thermal loading conditions, a damaged layer is formed in SiC which inherits the typical erosion defects (craters, chips, microcracks).
 [S1063-7842(98)02407-6]

1. A number of technological applications of silicon carbide require both bulk profiled single crystals and filamentary (whisker) crystals, which have a unique shape and specific properties.[1–4] Besides the purposeful, oriented production of filamentary crystals, under an electric field, for example, they can appear as an accompanying growth in the form of defects during crystallization processes on nuclei from the vapor phase or involving metal solvents.[5,6] Here two forms of filamentary crystals occur: growth structures and evaporation structures (negative crystals). As structural imperfections, they change the morphology, optical transparency, and other properties of grown crystals.[4,7]

It is known that the methods of erosion technology ensure the uniform profiling of slab and bulk crystals of silicon carbide.[8] The processes and phenomena which accompany these methods, however, have not yet been studied adequately. In this paper we present the results of some experimental studies of erosion tracks formed in silicon carbide crystals acted on by pulsed electrical discharges.

2. Spark cutting of silicon carbide was performed both on industrial and laboratory equipment by the principle of hole punching, with three-dimensional copying of the electrode shape.[8,9] The cutting was carried out in transformer oil. Profiling electrodes with simple and complex (ribbed, hemispherical) shapes made of copper, aluminum, brass, nickel, steel, silicon, graphite, silicon carbide, and tin alloys were used. The initial material for these studies was in the form of single crystal slabs of silicon carbide of polytype 6H, with $N_s - N_a \approx 5\times10^{17} - 3\times10^{18}$ cm^{-3} and thickness $d \approx 450 - 500$ μm.

The experiments, during which holes (passing the crystal) and cavities were made and problems of engraving were studied, showed that during pulsed thermal loading a damaged layer is formed on SiC which inherits the typical erosion defects (craters, shear fracture, microcracks).

It is known that the elementary shaping cell in erosion profiling processes in metals is the crater.[9] Optical studies of erosion tracks on the polar faces of SiC crystals (0001)C and (0001)Si (metallographic microscopy) have shown that an individual erosion crater can have a complex structure. It should be noted that carbon self-decoration of the erosion front of craters when a crystal is optically transparent created the conditions for a detailed analysis of their structure and made reliable measurements possible. As a rule, craters with a flat bottom were observed, as well as craters within which there were groups of craters or hexagonal etch pits. Each crater contained erosion products: carbon, slag, melt. With increasing energy of the electrical pulse, the fraction of carbon and slag increased. Although the chemical composition of the melt was not analyzed in the present experiments, we may assume that it is based on silicon as a dissociation product of SiC, along with the electrode metal (or silicide). This is in good agreement with the results of an experiment in which SiC was used as a profiling electrode and the melt was observed in an erosion crater. The appearance of the melt during erosion leads to etching of the crystal surface and the appearance of (0001)Si faceting of the craters. The geometric dimensions of the craters depended on the pulse energy and the crystallographic direction (Fig. 1). However, the growth in the crater diameter with rising pulse energy rapidly reached saturation, and single craters with diameters in excess of 300 μm were not observed.

3. In our studies of primary erosion tracks we have observed a new type of erosion defect in silicon carbide (Fig. 2). Because of the outward appearance of the defect, we have decided to call it an EFT[1] (erosion flat thorn). An EFT is a defect localized on the surface of a crystal. Its structure is based on an erosion crater, surrounded by extended voids in the shape of spines which extend under the crystal surface. The crystallographic character of the defect is manifest in the predominant growth of the spines in the $[11\bar{2}0]$ direction. As can be seen from Fig. 2, the crater is partially filled with frozen melt, traces of which surround the wall of that part of

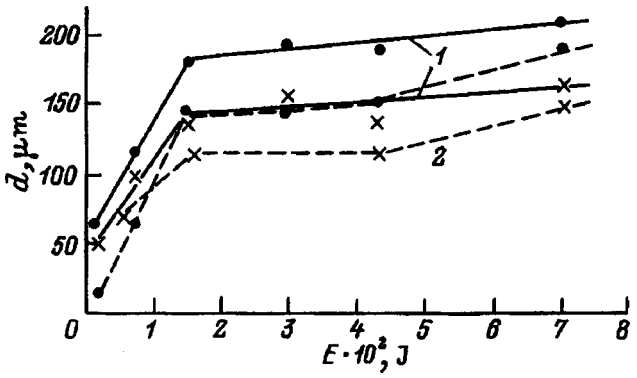


FIG. 1. The experimental dependence of the average crater diameter on the pulse energy; 6H-SiC; transformer oil. ●—(0001)C, ×—(0001)Si; *1*—copper electrode (smooth curves), *2*—SiC electrode (dashed curves).

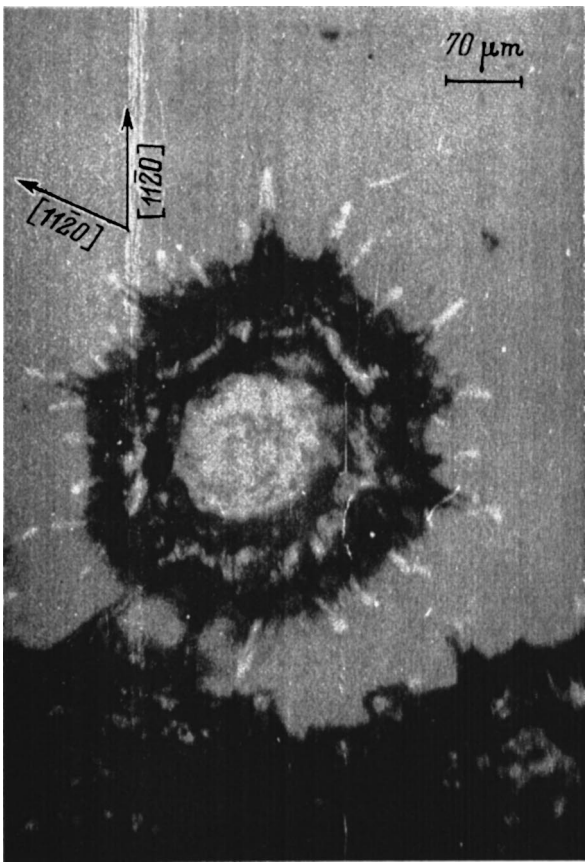


FIG. 2. A complete EFT erosion defect. 6H-SiC, (0001)Si face, copper electrode, $E_p = 1.48 \times 10^{-2}$ J.

the spines which lie on the crystal surface. In the experiments we found the following forms of EFT: complete and fragmentary. The complete defects are isolated and appear during the initial stages of erosion. They can be formed on either silicon or carbon (0001) faces of the crystal. On the (0001)C faces, however, there were defects with short, blunt spines. A similar spine structure was observed on the (0001)Si facets at high pulse energies.

The fragmentary EFTs contained a limited number of spines of different lengths and part of a crater. They most often develop from the edge of a hole (cavity) during erosion of the crystal.

The most structurally active elements of an EFT defect are the spines. The spines can undergo branching and bending, but isolated defects can also interact through them. As the spines develop, they also manifest a sensitivity to the inherited growth defects of the SiC crystal.

An analysis of the experimental facts described above showed that the observed EFT defects are formed by a system of negative filamentary crystals.[4] Their nucleation and growth are directly related to the evolution of the liquid phase formed within the local erosion volume of the SiC. In fact, negative filamentary crystals grow by a crystal–liquid–vapor mechanism.[10] The complex relief of the EFT defect (Fig. 2) is caused by the fact that evaporation of SiC involving the liquid phase takes place far from equilibrium and is explosive in character. For example, the experimentally determined linear growth rate of a spine in the $[11\bar{2}0]$ direction was ~ 0.7 m/s for a pulse duration $\tau_p = 100$ μs.

4. Let us conclude by summarizing the main results of this work: (a) the geometric dimensions of the erosion craters and the degree to which they are filled inside depend on the energy of the electrical pulse and the crystallographic direction; (b) melt formation during the erosion process causes faceting of the craters and growth of filamentary negative crystals; (c) EFT defects have a pronounced crystallographic character; and, (d) the spines of an EFT defect are structurally active elements.

[1] V. N. Levin, Yu. M. Tairov, M. G. Travadzhyan, and V. F. Tsvetkov, Izv. Akad. Nauk SSSR, Neorg. Mater. **14**, 1062 (1978).
[2] V. A. Karachinov, ''Dislocation structure of profiled single crystals of 6H polytype silicon carbide,'' Intercollegiate collection, NRI, Novgorod (1989), pp. 81–86.
[3] A. Yu. Maksimov, A. A. Mal'tsev, N. K. Yushin, and I. S. Barash, Pis'ma Zh. Tekh. Fiz. **21**(10), 20 (1995) [Tech. Phys. Lett. **21**(5), 360 (1995)].
[4] E. I. Givargizov, *Growth of Filamentary and Slab Crystals from Vapor* [in Russian] (Nauka, Moscow, 1977), 304 pp.
[5] V. I. Muratova, in *Filamentary Crystals for New Technologies* [in Russian] (Voronezh, 1979), pp. 45–49.
[6] G. K. Safaraliev, B. A. Bilalov, and A. V. Éfendiev, Zh. Tekh. Fiz. **54**, 2016 (1984) [Sov. Phys. Tech. Phys. **29**, 1181 (1984)].
[7] S. K. Lilov, Yu. M. Tairov, V. F. Tsvetkov, and M. A. Chernov, in *Filamentary Crystals* [in Russian] (Voronezh, 1975), p. 106.
[8] O. G. Bazhenov and V. A. Karachinov, Pis'ma Zh. Tekh. Fiz. **22**(21), 26 (1996) [Tech. Phys. Lett. **22**(11), 875 (1996)].
[9] A. L. Livshits, A. G. Kravets, N. S. Rogachev, and A. B. Sosenko, *Electropulse Metal Working* [in Russian] (Mashinostroenie, Moscow, 1967), 295 pp.
[10] R. S. Wagner, J. Cryst. Growth **3**(4), 159 (1968).

Translated by D. H. McNeill

Enhancing the intensity of x-ray reflection from surfaces by depositing diamondlike carbon films on them

A. M. Baranov

State Scientific-Research Institute for Vacuum Techniques, 113105 Moscow, Russia
(Submitted September 2, 1997)
Zh. Tekh. Fiz. **68**, 136–138 (July 1998)

The possibility of fabricating reflecting surfaces with a roughness low enough for x-ray optics is demonstrated experimentally. [S1063-7842(98)02507-0]

In recent years, there has been an ever wider range of scientific and applied problems in which x rays and neutron fluxes are used. Examples include x-ray and neutron spectroscopy and microscopy, x-ray lithography, etc.[1] These radiations are steered by grazing-incidence mirrors and multilayer x-ray mirrors. The intensity of x-ray reflection from mirrors depends strongly on their surface roughness.[2] For this reason, substrates with ultrasmooth surfaces are used in x-ray optics to reduce scattering losses. A large number of different types of polishing have been developed in order to obtain ultrasmooth surfaces.[3,4] However, since the commonly used surface finishing techniques were developed for visible-range optics and microelectronics, they do not always yield the surface quality required for x-ray optics.

For this reason, surfaces are prepolished by directed beams of inert and chemically active gas ions: argon, oxygen, etc. Ion-plasma polishing of materials, however, does have its limitations. The problem is that the materials to be treated often have their own surface microdefects (dislocations, inclusions, inhomogeneities in chemical composition, etc.). This means that various parts of a material surface are etched at different rates. Because of this, depending on the initial quality of the surface to be treated, there is always a residual surface roughness (root-mean-square roughness height $\sigma \approx 3-10$ Å), which is hard to eliminate. Wave scattering on this sort of roughness is one of the most important factors influencing the quality of x-ray and neutron optics. One way of reducing the roughness may involve depositing a thin film of another material (isotropic and uniform in structure and properties) on the original substrate and then etching this film. Then one may hope to obtain reduced roughness in the substrate–film system.

In this paper we study diamondlike carbon films which can have a surface roughness $\sigma_f < 10$ Å under certain conditions.[5] In addition, diamondlike carbon films are easily etched in oxygen plasmas. Thus, one can choose the conditions for etching of the diamondlike carbon film to be such that there will also be a simultaneous reduction of the surface roughness during the etching process. Diamondlike carbon films were deposited from the gaseous phase in an rf discharge and by magnetron sputtering of a graphite target.[6,7] Ar and cyclohexane C_6H_{12} were used as working gases. An *in situ* x-ray monitor of the thickness and roughness of the surface layers was installed in the vacuum chamber to monitor the deposition of the layers and their subsequent polishing by etching.[6–8]

Monitoring was based on the change in the intensity of x rays ($\lambda = 1.54$ Å) specularly reflected from the film–substrate system during film deposition or etching. The angle of incidence was $\theta = 1°$. As the thickness of the layer changes, oscillations appear in the curve owing to interference of x rays reflected by the upper and lower boundaries. Polished wafers of silicon with an initial rms roughness of $\sigma_s = 8$ Å were used.

Whether the surface roughness is increasing or decreasing can be judged from changes in the intensity of the x-ray flux reflected from the film–substrate system. In fact, the reflection coefficient of a smooth surface will be maximal, since all the incident flux is reflected at an angle θ equal to the angle of incidence. When the surface has a relief, this situation changes. Part of the radiation will be scattered at arbitrary angles and will not be picked up by the detector. Thus, it is possible to evaluate the changes in surface roughness from the changes in intensity of the specularly reflected radiation.

The intensity J of the x radiation reflected from the film–substrate system varies cosinusoidally with the film thickness d.[6] The shape of the cosinusoid depends both on the substrate parameters, which are known, and on the film parameters. A change in the film surface roughness σ_f will lead to a change in the average intensity of the reflected radiation, $J_{\text{mid}} = (J_{\max} + J_{\min})/2$. ($J_{\max}$ and $J_{\min}$ are the maximum and minimum intensities in the $J = f(d)$ curve). If σ_f increases compared to σ_s during deposition or etching of the film coating, then J_{mid} will decrease.[7] If σ_f decreases, then J_{mid} will increase.

A typical experimental plot of the reflected radiation intensity as a function of the duration of the deposition process for a carbon film is shown in Fig. 1 (curve *1*). The diamondlike carbon film was obtained from C_6H_{12} in an rf discharge ($f = 13.56$ MHz). Electrical power was delivered to the electrode on which the silicon wafer was placed. It is clear from Fig. 1 that the $J_0(t)$ curve oscillates as the carbon film grows. The arrows in Fig. 1 indicate the start (s) and finish (f) of the deposition and etching processes. The intensity of the reflected light at the maxima does not exceed the reflected intensity from the clean substrate, $J_s = 1$. After a

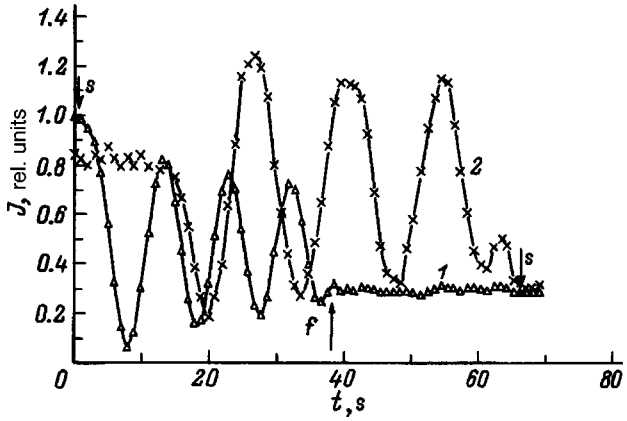


FIG. 1. The intensity of reflected light as a function of time during deposition (*1*) and etching (*2*) of a diamondlike carbon film.

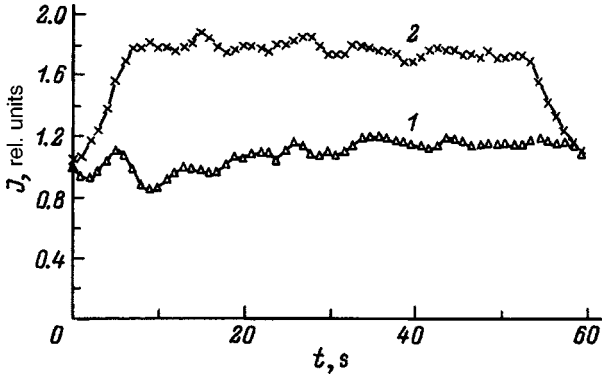


FIG. 2. $J_0(t)$ for a film obtained by magnetron sputtering in a C_6H_{12}/Ar=1/2 medium (*1*) and $J_T(t)$ (*2*).

thickness of 160 Å was reached, the film growth process was halted.

The resulting carbon film was etched in an rf discharge in oxygen. The resulting $J_T(t)$ curve is shown in Fig. 1 (curve *2*). It is seen that $J_T(t)$ is a good repetition of $J_0(t)$ in the opposite direction. However, $J_T(t)$ reaches higher absolute values than $J_0(t)$. Furthermore, $J_{\max} > J_s$. This is a consequence of the fact that the surface roughness of the film during etching is not only smaller than when it was being deposited, but is smaller than that of the original substrate. A calculation[7] showed that σ_f during etching was ≈ 6.5 Å. After the film was removed from the Si surface, the reflection coefficient ceased to oscillate.

Figure 2 shows plots of $J_0(t)$ and $J_T(t)$ of a film obtained by magnetron sputtering in a C_6H_{12}/Ar=1/1 mixture. It is clear that during film growth on a silicon substrate, $J_0(t)$ is essentially invariant and remains roughly equal to J_s. The curve can have this form if the optical constants of the film and substrate are the same and the surface roughness of the film is the same as that of the silicon surface.

No oscillations were observed during etching, either. However, $J_T(t)$ goes substantially higher than $J_0(t)$. Thus, even in this case, during etching the surface roughness σ_f of the film is smaller than that of the silicon, by ≈ 2.5 Å. After removal of the film from the silicon surface, the intensity of the reflected radiation fell to a level corresponding to the reflection from the clean substrate before the deposition process was begun.

We may therefore conclude that during deposition from the gaseous phase, the diamondlike films reproduce the relief of the substrate surface. Etching them in an oxygen plasma, however, leads to a reduction in the surface roughness of the film. This is evidently a result of the combined action of physical and chemical etching mechanisms. As a result, it is possible to grow a diamondlike carbon film of a certain thickness and then partially etch it away. Ultimately, we obtain a substrate with a film on it whose surface roughness is smaller than that of the substrate itself. It is important to note that the film thickness can be several tens of angstroms.

This method may be useful when surface polishing by other means is difficult (e.g., if the original substrate is curved).

[1] N. N. Salashchenko, Yu. Ya. Platonov, and S. Yu. Zuev, Poverkhnost', No. 10, 5 (1995).
[2] *X-ray Mirror Optics* [in Russian] (Mashinostroenie, Leningrad, 1989), 463 pp.
[3] I. Ali, S. R. Roy, and G. Shinn, Solid State Technol. No. 10, 63 (1994).
[4] *Processing of Semiconductor Materials* [in Russian] (Naukova Dumka, Kiev, 1982), 256 pp.
[5] R. Prioli, S. I. Zanette *et al.*, J. Vac. Sci. Technol. A **14**, 2351 (1996).
[6] I. F. Mikhailov, V. I. Pinegin, V. V. Sleptsov, and A. M. Baranov, Cryst. Res. Technol. **30**, 643 (1995).
[7] A. Baranov, S. Tereshin, and I. Mikhailov, Proc. Soc. Photo-Opt. Instrum. Eng. (SPIE), **2863**, 359 (1996).
[8] A. M. Baranov and I. F. Mikhaĭlov, Pis'ma Zh. Tekh. Fiz. **22**(23), 60 (1996) [Tech. Phys. Lett. **22**(12), 977 (1996)].

Translated by D. H. McNeill

Annular laser speckle patterns

Yu. V. Vasil'ev, A. V. Kozar', E. F. Kuritsyna, and A. E. Luk'yanov

M. V. Lomonosov Moscow State University, 199899 Moscow, Russia
(Submitted March 19, 1997)
Zh. Tekh. Fiz. **68**, 139–140 (July 1998)

[S1063-7842(98)02607-5]

The diffraction of a Gaussian laser beam at the straight edge of a semi-infinite metal screen is characterized by a number of features owing to the edge nature of the diffracted radiation. Although the main processes leading to transformation of a wave front take place in the immediate neighborhood of the edge, it is customary to evaluate them in terms of various optical effects which are observed at a rather large distance. For example, aligning the edge of the screen with a diameter of a beam incident perpendicular to the plane of the screen leads to a maximum in the diffractive scattering of the laser beam, as a result of which the observer's eye sees distinctly a bright, ''self-luminous'' straight line segment at the edge of the screen. Objective detection of the scattered light with a sheet of photographic paper, etc., confirms that the diffracted radiation diverges from the ''self-luminous'' segment in different spatial directions with unequal intensities. Along predominantly forward directions, it has a complicated structure, the fine structure being due to the development of anisomeric laser speckle. The structuring is caused by large-scale and small-scale spatial modulation of the wave front resulting from unavoidable local deviations in the profile of the real edge of the screen from strict straightness.[1]

In this paper we report the possible existence of a more complicated, ultrafine-scale modulation in the wave front. It was encountered during a study of the structure of the annular laser speckle we have observed and detected objectively in diffraction experiments using a type UIG-22M holographic measurement system.

An LGN-503 argon laser generates cw monochromatic radiation in the form of a Gaussian beam (TEM_{00} spatial mode, effective diameter 2.5 mm, wavelength 488 nm). The beam is incident normally on the plane of a diffracting screen consisting of a flat, 80-μm-thick steel safety razor blade. Diffraction takes place on the vertically oriented straight sharpened edge of the blade, which is aligned with a diameter of the beam such that the oscillations of the electric field of the incident wave are parallel to the edge. The sharpened edge, of length 37 mm, was given a two-sided symmetric wedge shape, with a characteristic size on each side on the order of 1 mm.

As detailed studies of the blade with a JSM-U3 scanning electron microscope (maximum on-screen magnification 10 000×, resolution 20 nm) show, the top of the assumed wedge is by no means a geometrically ideal, infinitely thin ridge. In fact, the actual edge ends in a very narrow area—the cutting edge, whose average width is 300 nm. This fact is illustrated by Fig. 1, which shows a typical fragment of the sharpened blade edge (side view) viewing an object area of 33×33 μm^2 in the frame. This micrograph was obtained using the photographic attachment of the microscope at a magnification of 3000×.

In Fig. 1 the cutting area of the blade appears in the form of a continuous dark strip crossing the center of the frame from left to right at an angle to the horizontal. Above and below it, one can see both sides of the wedge edge, with their individual surface microdefects. The two edges, along which the sharp boundary of the cutting area passes on the corresponding side, change their local spatial directions independently of one another. This indicates that the local thickness of the end of the edge along the blade fluctuates chaotically. Measurements with the microscope show that the variations in the thickness of the edge of the edge mainly lie with in a range from 100 to 500 nm.

This circumstance evidently affects optical phenomena at large diffraction angles, when the effect of the large-scale modulation in the wave front has become insignificant and, in first order, all the perturbations in the laser light on the order of a wavelength come to the fore. Thus, in the experiment the photographic frame of the UIG-22M system was placed in a horizontal position below the edge at a distance L along the vertical from the center of the illuminated part of the sharpened edge. The distribution of the illumination in the frame window was recorded objectively on a flat sheet of photographic paper (at different exposures and the values of L of interest). For example, Fig. 2 shows one of the optical illumination patterns obtained for $L=5$ cm (negative image).

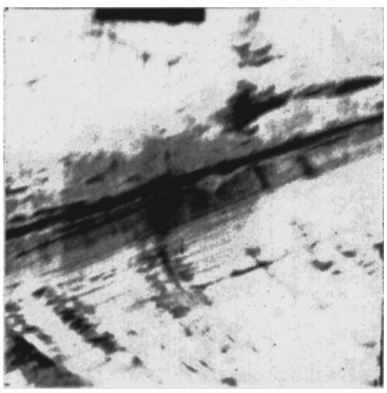


FIG. 1. The microscopic structure at the end of the edge of a wedge shaped sharpened blade (side view).

Here there are two regions of detected laser speckle. In the lower part of the picture (from the side of the plane of the blade illuminated by the laser beam) and in the upper (on the shadow side). The lower part, which is almost completely exposed at the chosen exposure, mainly carries information on the roughness of the metal surface, which scatters the coherent laser light incoherently into the half space.[2] The upper part of the picture illustrates the system of annular laser speckle created by all sorts of chaotic microscopic irregularities at the end of the sharpened blade edge during scattering of the laser light. The edge nature of the annular speckle is indicated, first of all, by its concentric disposition relative to the wake of the imaginary extension of the straight blade edge to the plane of the paper and, second, by the observed darkening of this system of speckle by the edge of the inclined side, far away in relation to the laser source. It shows up in the form of a triangular discontinuity between the surface and edge speckle on the left of Fig. 2. It is clear that this discontinuity expands linearly upward, beginning from the center of the annular speckle. On the right-hand side of the figure, on the other hand, the two speckle systems overlap.

The most surprising experimental observation was the existence of peaks in the integrated intensity of the system of annular speckle precisely in the direction along the boundary of the region where the two speckle systems overlap on the right-hand side of the figure, while along the direction upward from the center of the annular speckle, i.e., along the propagation path of the transmitted part of the beam beyond the screen, their integral intensity is lower, and it becomes very small along the directions in the upper left part of the figure, beyond the shadow side of the blade. This type of scattering (and not Rayleigh and not quasispecular scattering), which shows up when an annular speckle system develops, suggests an important role for the impedance contribution to the scattering mechanism resulting from the finite conductivity of the screen material, whose thickness close to the end of the edge is locally several times smaller than the wavelength of the incident coherent light.

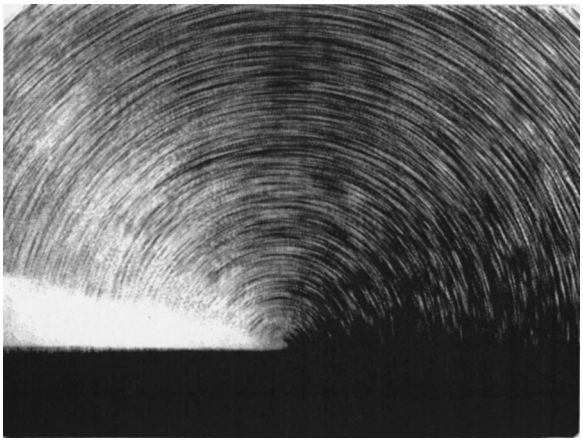


FIG. 2. Laser speckle systems of surface and edge origin in a negative image. The plane of the sheet of photographic paper is perpendicular to the straight sharpened edge of the blade, and the distance from the center of the illuminated segment of the edge to the paper is 5 cm.

[1] Yu. V. Vasil'ev, A. V. Kozar', E. F. Kuritsyna, and A. E. Luk'yanov, Pis'ma Zh. Tekh. Fiz. **19** (14), 29 (1993) [Tech. Phys. Lett., **19**, 445 (1993)].
[2] M. Francon, *Laser Speckle and Applications in Optics* [Academic Press, New York (1979); Mir, Moscow (1980), 171 pp.].

Translated by D. H. McNeill

Influence of deformation on the force characteristics of shape memory effects in alloys based on titanium nickelide

T. Yu. Maletkina and V. É. Gyunter

Medical-Engineering Center for Implants with Shape Memory, 634034 Tomsk, Russia
(Submitted September 2, 1997)
Zh. Tekh. Fiz. **68**, 141–142 (July 1998)

Studies of the influence of plastic deformation and the straining temperature on the temperature dependence of the stresses developed in titanium nickelide shape memory alloys are described. © *1998 American Institute of Physics.* [S1063-7842(98)02707-X]

Alloys based on titanium nickelide have unique shape memory properties, superelasticity, and plasticity. The medical application of titanium nickelide devices and structures as implants is often based on the material's overcoming an external resistance in the course of recovering its shape during heating after prestraining. The external resistance limits the recovery of shape and causes internal stresses to develop in the material. In the limiting situation of a rigidly confined sample, the recovery of shape is completely prevented, and the internal stress reaches its highest level, which is close to the critical stress for the rapid development of the martensitic deformation at temperatures below M_d and close to the yield point above M_d (Fig. 1b, curve *1*).[1] For a 2–6% prestrain the temperature dependence of the developed stresses, $\sigma-T$, below the temperature M_d coincides completely with the temperature dependence of the critical stress for the martensitic strain.[2]

Since force and strain effects in titanium nickelide are closely coupled, information obtained from a plot of the temperature dependence of the developed stresses can be used to make a qualitative, and in some cases, even quantitative, estimate of its shape memory and superelasticity characteristics.

The purpose of this paper is to study the influence of plastic deformation and the straining temperature on the temperature dependence of the developed stresses in titanium nickelide alloys using TN-10 alloy, one of the most widely used in medicine, as an example. TN-10 alloy is characterized by superior physical and mechanical properties and has a convenient temperature interval for shape recovery.

Both in the original undeformed state and after deformation, there are two martensitic transformations in the chosen alloy: at a temperature $T_R=20$ °C from the high-temperature $B2$ phase to the K phase, and then at a temperature $M_s=-5$ °C to the $B19'$ phase.

The samples were deformed by 1.5, 4, 8, 12, and 16%, each of which corresponds to different stages in the development of the martensitic and plastic deformation components, on a type UPR universal testing machine operating under tensile straining conditions at temperatures of -196, -150, -50, 0, 25, 50, and 100 °C. The magnitude of the deformation was determined from the $\sigma-\varepsilon$ diagram taken at room temperature before fracture. In order to study the characteristics of the developed stresses, samples confined in the apparatus, which had been strained and not unloaded, were heated by an electric oven to a temperature of 200–250 °C with simultaneous recording of the $\sigma-T$ diagram. The stress relaxation after straining of the samples at temperatures of 0, 50, 100, 150, and 200 °C, was studied by cooling the confined material and simultaneously recording the $\sigma-T$ diagram.

Figure 1b shows the temperature dependence of the internal stresses during cooling (curve *1*) and subsequent heating (curve *2*) after deformation by 4% in the martensitic state. In the course of continuous heating through the interval of the martensitic transformations in the confined alloy, stresses are observed to relax and be generated. The minimum value of the martensitic shear stress $\sigma_{\min}$ in the $\sigma-T$ diagram (Fig. 1b, curve *1*) corresponds roughly to the onset temperature M_s for formation of thermal martensite $B19'$. At the temperature M_d, the stresses reach the yield point and have their maximum value, $\sigma_{\max}$. Subsequent heating leads to relaxation of the stresses by plastic shear, the resistance to which corresponds to the yield point and is found on the $\sigma-T$ curve by some extrapolation curve from high to low temperatures (Fig. 1b, dashed line). The yield points for all temperatures below M_d are close to the maximum value of the martensitic shear stress $\sigma_{\max}$ at the temperature M_d.

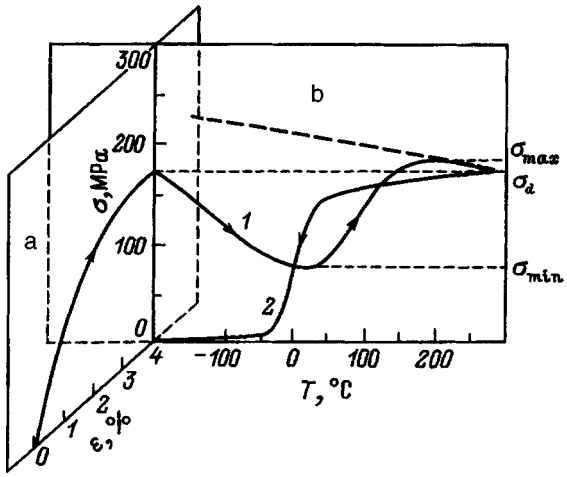


FIG. 1. Loading (a) and subsequent change in the external stress (b) during continuous heating (*1*) and cooling (*2*) of TN-10 alloy.

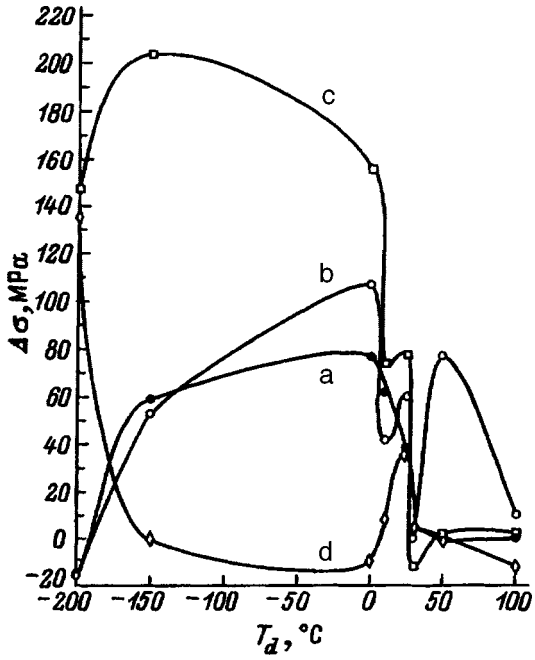


FIG. 2. $\Delta\sigma=|\sigma_{\max}-\sigma_{\min}|$ as a function of straining temperature for different amounts of prestrain of the material. ε (%): 1.5 (a), 4 (b), 8 (c), 12 (d).

Thus the difference between the maximum and minimum internal stresses, $\Delta\sigma=|\sigma_{\max}-\sigma_{\min}|$, for a given temperature characterizes the susceptibility of the material to plastic deformation and, as a consequence, the contribution of the plastic component of the strain to recovery of shape in the shape memory effects. The higher the yield point and the lower the martensitic shear stress, the lower the contribution of plastic deformation to the strain of recovery and the higher the degree of shape recovery.

Figure 2 shows plots of $\Delta\sigma$ as a function of the deformation temperature T_d for different amounts of initial prestrain. The highest values of $\Delta\sigma$ correspond to an 8% prestrain in the temperature region in which martensite $B19'$ exists (Fig. 2d); this is evidence of the possibility of superelastic behavior in this alloy and of its high reserve of recoverable strain within this temperature interval. Evidently, this amount of strain leads to the development of the most favorable internal conditions for realization of the martensitic transformations and to an influence of the R phase's on the stabilization of the $B19'$ martensite. Increasing the strain suppresses shape memory and superelasticity effects because of the increased contribution of the plastic component of the strain, which inhibits movement of the interphase boundaries during martensitic transformations (Fig. 2d).

Cooling through the interval of the martensitic transformations is accompanied by relaxation of the accumulated stresses, independently of the straining temperature (Fig. 1b, curves *2* and *3*). It is clear from a comparison of curves *1* and *2* of Fig. 1 that a rapid reduction of internal stresses during cooling is observed in the range of temperatures of the martensitic transformations, which corresponds to the minimum values of the martensitic shear stresses. During deformation in the high-temperature state, the level of the maximum stresses $\sigma_{\max}$ depends on the deformation temperature. After straining at 200 °C, the $\sigma-T$ curve has two stages. The first stage of stress relaxation occurs because, during straining near M_d, the plastic component of the strain leads to the creation of stress concentrators, which stimulate the nucleation and growth of martensite crystals as the temperature is lowered. The major stress reduction takes place in the range of the martensitic transformations and is described by the Clausius–Clapeyron equation.[3] Cooling to the temperature of liquid nitrogen leads to a rise in σ, i.e., stress again develops in the medium. This fact has been observed for the first time and is attributed to a reorientation and twinning of the cooled martensite. The results obtained here can serve as a guide in choosing the conditions for and changing the parameters of the shape memory in a way appropriate to the problems being addressed.

[1] V. É. Gyunter, V. I. Itin, L. A. Monasevich *et al.*, *Shape Memory Effects and their Applications in Medicine* [in Russian] (Nauka, Novosibirsk, 1992), 742 pp.
[2] V. É. Gyunter, T. Yu. Serikova, L. A. Monasevich, and Yu. A. Paskal', Inventor's Certificate No. 1698688 (USSR), MKI Go1 No. 128; Byull. Izobret., No. 46 (1991).
[3] C. Rodrigez and L. C. Brown, in *Shape Memory Effects in Alloys* (Metallurgiya, Moscow, 1979, pp. 35–39).

Translated by D. H. McNeill